SPRINGER REFERENCE

Stefano Battilossi
Youssef Cassis
Kazuhiko Yago
Editors

Handbook of the History of Money and Currency

Springer

Handbook of the History of Money and Currency

Stefano Battilossi • Youssef Cassis
Kazuhiko Yago
Editors

Handbook of the History of Money and Currency

With 101 Figures and 56 Tables

Springer

Editors
Stefano Battilossi
Department of Social Sciences
Universidad Carlos III de Madrid
Madrid, Spain

Youssef Cassis
European University Institute
Florence, Italy

Kazuhiko Yago
School of Commerce
Waseda University
Tokyo, Japan

ISBN 978-981-13-0595-5 ISBN 978-981-13-0596-2 (eBook)
ISBN 978-981-13-0597-9 (print and electronic bundle)
https://doi.org/10.1007/978-981-13-0596-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Financial history as an academic discipline enjoyed tremendous growth in the 1980s and 1990s. Economic historians' renewed interest in a hitherto neglected field reflected the growing shift from manufacturing industry to services in the Western world and the rise of a global economy fueled by the resumption of international capital flows. Interest in financial history has continued unabated but has taken a different turn in the wake of the Global Financial Crisis of 2008. No longer confined to academic circles, the search for the meaning of past experiences has extended to policy makers and even to banking practitioners trying to make sense of the enormity of the debacle that had shaken the financial world. Professional economic historians, including economists engaged with the past, have had to bear a new responsibility: to extend the depth and scope of their investigations, share their results with a broader audience, and maintain exacting academic standards.

This task can be fulfilled by exploiting the extremely rich vein of financial history. Finance is a highly technical specialism, in both theory and practice, but it also touches upon all strands of economic, social, political, and cultural life. Financial history is concerned with the contribution of banks and financial markets to economic development; with the impact of monetary policy on economic stability and economic growth; with capital exports, foreign investment, and their effects on both creditor and debtor nations; and with the management and governance of financial institutions. It is also interested in the people involved in financial transactions – from the wealth, status, and power of financial elites to the financial behavior of small savers. Financial history's remit also includes the relationships between finance and politics, whether at national or international level – state intervention in financial affairs, the political influence of financial interests, and the interactions between finance and international relations. The list is not exhaustive.

The increasing level of attention recently paid to the history of financial crises represents, as it were, a synthesis of these various approaches. For it is in time of panic and crisis that the interconnection between the business, economic, social, political, and international sides of financial activity is revealed in its most glaring way – as was vividly exposed by the Global Financial Crisis of 2008, with bank failures and near failures, systemic risks, bankers' responsibility, state intervention, the Great Recession, unorthodox monetary policies, international cooperation, and so on.

Monetary history has been one of the fastest-growing subfields of financial history – an interest spurred by the inflationary experiences of the late twentieth century and their effects on economy, society, and politics; the end of Bretton Woods and the advent of flexible exchange rates; the search for monetary stability; the growing importance of central banks and their conduct of monetary policy; and of course Europe's monetary unification.

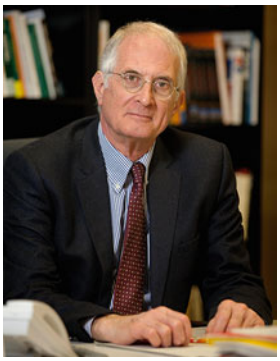
The publication of the *Handbook of the History of Money and Currency* represents the culmination of nearly 50 years of research in these areas and beyond. It is an impressive collective effort, whether viewed thematically (its 40 chapters cover all relevant issues), chronologically (from the ancient world to the present), geographically (with its global view), and methodologically (multidisciplinary approach, theoretical insights, state of the art). It provides a long-term historical perspective to current issues and integrates monetary history into the broader spectrum of financial history. The book is a fascinating journey into the multifaceted world of money and currency. It will also be an essential tool for social scientists and a handy companion in the hands for decision makers.

Florence, Italy Youssef Cassis
January 2020

Contents

About the Editors

Stefano Battilossi is Associate Professor of Economic History in the Department of Social Sciences, Universidad Carlos III de Madrid. His research interests include international banking, financial regulation, macroeconomic policies, and stock markets in historical perspective, with a special focus on Western Europe in the nineteenth and twentieth centuries. He has published articles in *The Economic History Review*, the *European Review of Economic History*, and *Cliometrica* and contributed chapters to *The Cambridge Economic History of Modern Europe* (Cambridge University Press, 2010, with James Foreman-Peck), *The Oxford Handbook of the Italian Economy Since Unification* (Oxford University Press, 2013), and *The Oxford Handbook of Banking and Financial History* (Oxford University Press, 2016). He has edited *European Banks and the American Challenge: Competition and Cooperation in International Banking Under Bretton Woods* (Oxford University Press, 2002, with Youssef Cassis), and *State and Financial Systems in Europe and the USA: Historical Perspectives on Regulation and Supervision in the Nineteenth and Twentieth Centuries* (Ashgate, 2010, with Jaime Reis). He has been an editor of the *Financial History Review* (Cambridge Journals) since 2010. He served as a member of the Board of Trustees of the European Historical Economics Society (2006–2014) and sits in the Academic Advisory Council of the European Association for Banking and Financial History (EBHA) (Frankfurt a.M.).



Youssef Cassis is Professor Emeritus of Economic History at the European University Institute, Florence. His work mainly focuses on banking and financial history, as well as business history more generally. His most recent publications include *Capitals of Capital: A History of International Financial Centres, 1780–2005* (Cambridge University Press, 2006, 2nd revised edition, 2009), *Crises and Opportunities: The Shaping of Modern Finance* (Oxford University Press, 2011), and *Private Banking in Europe: Rise, Retreat, and Resurgence* (Oxford University Press, 2015, with Philip Cottrell). He has also recently co-edited *The Oxford Handbook of Banking and Financial History* (Oxford University Press, 2016, with Richard Grossman and Catherine Schenk), *International Financial Centres After the Global Financial Crisis and Brexit* (Oxford University Press, 2018, with Dariusz Wojcik), and *Financial Elites and European Banking: Historical Perspectives* (Oxford University Press, 2018, with Giuseppe Telesca). He was the co-founder/editor, in 1994, of *Financial History Review* (Cambridge University Press). He was a long-serving member of the Academic Advisory Council of the European Association for Banking and Financial History (EBHA) and past President (2005–2007) of the European Business History Association.

Kazuhiko Yago is Professor of Economic History at the School of Commerce, Waseda University, Tokyo. He obtained a docteur en histoire de l'Université Paris X-Nanterre in 1996. His recent publications include "A Crisis Manager for the International Monetary and Financial System? The Rise and Fall of the OECD Working Party 3, 1961–1980" in *The OECD and the International Political Economy Since 1948* (Palgrave, 2017; Matthieu Leimgruber and Matthias Schmelzer, eds.); as an editor, with Hubert Bonin and Nuno Valério, of *Asian Imperial Banking History* (Taylor and Francis, 2015); and as the author of *The Financial History of the Bank for International Settlements* (Routledge, 2012).

Contributors

Olivier Accominotti London School of Economics and Political Science, London, UK

Centre for Economic Policy Research, London, UK

Massimo Amato Department of Social and Political Sciences, Università Bocconi, Milan, Italy

Stefano Battilossi Department of Social Sciences, Universidad Carlos III de Madrid, Madrid, Spain

Martin T. Bohl School of Business and Economics, University of Münster, Münster, Germany

Richard C. K. Burdekin Robert Day School of Economics and Finance, Claremont McKenna College, Claremont, CA, USA

Georges Depeyrot Centre National de la Recherche Scientifiquen, École Normale Supérieure, Cahors, France

Barry Eichengreen George C. Pardee and Helen N. Pardee Professor of Economics and Political Science, University of California, Berkeley, Berkeley, CA, USA

Colin P. Elliott Department of History, Indiana University, Bloomington, IN, USA

Michael Ellman Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

Cambridge Journal of Economics, Cambridge, UK

Luca Fantacci Department of Social and Political Sciences, Università Bocconi, Milan, Italy

David Fox University of Edinburgh, Edinburgh, Scotland, UK

Atish Ghosh International Monetary Fund, Washington, DC, USA

Farley Grubb Alfred Lerner College of Business and Economics, University of Delaware and NBER, Newark, DE, USA

Anne-Marie Gulde International Monetary Fund, Washington, DC, USA

Joshua R. Hendrickson Department of Economics, University of Mississippi, Oxford, MS, USA

Robert L. Hetzel Federal Reserve Bank of Richmond, Richmond, VA, USA

Niv Horesh Durham University, Durham, UK

Western Sydney University, Sydney, NSW, Australia

Michael Hudson University of Missouri-Kansas City, Kansas City, MO, USA

Levy Economics Institute of Bard College, New York, NY, USA

Institute for Long-term Economic Trends, Peking University, Beijing, China

Alejandra Irigoin Economic History, London School of Economics and Political Sciences, London, UK

John A. James University of Virginia, Charlottesville, VA, USA

Matthew Jaremski Colgate University, New York, NY, USA

National Bureau of Economic Research (NBER), Cambridge, MA, USA

Yohei Kakinuma Teikyo University, Tokyo, Japan

Seung Woo Kim Graduate Institute of International and Development Studies, Geneva, Switzerland

Meir Kohn Department of Economics, Dartmouth College, Hanover, NH, USA

Peter Kugler University of Basel, Basel, Switzerland

Bill Maurer Department of Anthropology, School of Social Sciences, University of California, Irvine, Irvine, CA, USA

Kenneth Mouré History and Classics, University of Alberta, Edmonton, AB, Canada

Emmanuel Mourlon-Druol Adam Smith Business School, University of Glasgow, Glasgow, UK

Duncan J. Needham University of Cambridge, Cambridge, UK

Pilar Nogues-Marco University of Geneva and CEPR, Geneva, Switzerland

Lawrence H. Officer Department of Economics, College of Liberal Arts and Sciences, University of Illinois at Chicago, Chicago, IL, USA

Anders Ögren Department of Economic History, Lund University School of Economics and Management, Lund, Sweden

---

Robert L. Hetzel has retired.

John A. James: deceased.

Nuno Palma Department of Economics, School of Social Sciences, University of Manchester, Manchester, UK

Instituto de Ciências Sociais, Universidade de Lisboa, Lisboa, Portugal

CEPR, London, UK

Esteban Pérez Caldentey Economic Commission for Latin America and the Caribbean ECLAC, Santiago, Chile

Catherine Schenk Faculty of History, University of Oxford, Oxford, UK

Michael Schiltz Department of International History, Graduate Institute, Geneva, Geneva, Switzerland

George Selgin Center for Monetary and Financial Alternatives, The Cato Institute, Washington, DC, USA

Masato Shizume Waseda University, Tokyo, Japan

Pierre L. Siklos Wilfrid Laurier University and Balsillie School of International Affairs, Waterloo, ON, Canada

Tobias Straumann University of Zurich, Zürich, Switzerland

Hisashi Takagi Faculty of Letters, Yasuda Women's University, Hiroshima, Japan

Stefano Ugolini Sciences Po Toulouse and LEREPS, University of Toulouse, Toulouse, France

François R. Velde Federal Reserve Bank of Chicago, Chicago, IL, USA

Matías Vernengo Bucknell University, Lewisburg, PA, USA

Oliver Volckart Economic History Department, London School of Economics and Political Science, London, UK

David F. Weiman Barnard College-Columbia University, New York, NY, USA

Holger Wolf School of Foreign Service, Georgetown University, Washington, DC, USA

Kazuhiko Yago School of Commerce, Waseda University, Tokyo, Japan

Introduction: New Research in Monetary History – A Map

1

Stefano Battilossi and Kazuhiko Yago

Contents

Abstract

This handbook aims to provide a comprehensive (though obviously not exhaustive) picture of state-of-the-art international scholarship on the history of money and currency. The chapters of this handbook cover a wide selection of research topics. They span chronologically from antiquity to nowadays and are geographically stretched from Latin America to Asia, although most of them focus on Western Europe and the USA, as a large part of the existing research does. The authors of these chapters constitute, we hope, a balanced sample of various generations of scholars who contributed to what Barry Eichengreen defined as "the new monetary and financial history" – an approach that combines the analysis of monetary aggregates and policies with the structure and dynamics

S. Battilossi (✉)
Department of Social Sciences, Universidad Carlos III de Madrid, Madrid, Spain
e-mail: battilos@clio.uc3m.es

K. Yago
School of Commerce, Waseda University, Tokyo, Japan
e-mail: yago@waseda.jp

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_55

of the banking sector and financial markets. We have structured this handbook in ten broad thematic parts: the historical origins of money; money, coinage, and the state; trade, money markets, and international currencies; money and metals; monetary experiments; Asian monetary systems; exchange rate regimes; monetary integration; central banking and monetary policy; and aggregate price shocks. In this introduction, we offer for each part some historical context, a few key insights from the literature, and a brief analytical summary of each chapter. Our aim is to draw a map that hopefully will help readers to organize their journey through this very wide and diverse research area.

Keywords
Monetary and financial history · Monetary regimes · Monetary theories · Monetary policies and institutions

This handbook aims to provide a comprehensive (though obviously not exhaustive) picture of state-of-the-art international scholarship on the history of money and currency. Its chapters cover a wide selection of research topics. They span chronologically from antiquity to nowadays and are geographically stretched from Latin America to Asia, although most of them focus on Western Europe and the USA, as a large part of the existing research does. Their authors constitute, we hope, a balanced sample of various generations of scholars who contributed to what Barry Eichengreen defined as "the new monetary and financial history" (Eichengreen 2011) – an approach that combines the analysis of monetary aggregates and policies with the structure and dynamics of the banking sector and financial markets.

The synthesis of money and finance is a development firmly rooted in history. On one hand, "outside money" (issued by monetary authorities) was increasingly complemented by "inside money" (private liabilities such as bills of exchange, banknotes, and bank deposits) in achieving payment finality. On the other hand, monetary policies were always intimately connected to government finance, especially in wartime, from debasements in the Middle Ages to massive debt monetization in the world wars of the twentieth century. In modern economies, money markets based either on short-term private or public debt evolved into a pillar of liquidity management by commercial banks and a privileged channel through which central banks influenced monetary conditions. As a consequence, the regulation of governments' access to central bank money and the stability of banks and money markets became essential tasks of monetary authorities in order to preserve the value of money and the functioning of the payment system. These historical developments explain why money, banking, and government finance evolved as overlapping and deeply integrated fields both in theoretical and empirical research under what is known today as monetary economics (Champ et al. 2018; Walsh 2010). In historical research, as emphasized by Eichengreen, topics such as the interaction between monetary policy and banking failures in the Great Depression, the impact of Gold Standard adherence on sovereign spreads, or the role of credit boom-bust cycles in financial crises are prominent examples of this synthesis.

The "new monetary and financial history" accounts for a substantial portion of research in economic history. In the *Journal of Economic History*, one of the leading publications in the field, articles dealing with monetary and financial topics showed a permanent upward shift in the 1980s compared to the previous decades. Jointly, papers in the subfields "money," "banking and credit," "public finance," and "business cycles and depressions" accounted for 14.7% of pages in the 1980s and 18.6% in the 1990s (Whaples 2002). Over the entire period (1941–2016), they represent 8.9% of published papers – a share quite far from "economic growth" (16.6%) but substantially larger than any other topic (Wehrheim 2019). Research in monetary and financial history also critically contributed to the professional integration of economic history into economics, as reflected by the rising trend of economic history papers published in economics journals since the 1990s (Abramitzky 2015; Margo 2018). Macroeconomic, monetary, and financial history accounts for 25% of the 82 economic history papers published in the top five economics journals between 2013 and 2018 (Jaremski 2019). A quick look at the bibliographical references at the end of this introduction confirms the remarkable capacity of research in monetary history to reach out to academic audiences beyond professional economic historians.

This handbook reflects the key characteristics of historical research on money in the last 30 years. Its analytical framework is strongly influenced not only by modern monetary and financial economics but also by other theoretical strands such as network and information economics. At the same time, it draws increasingly on new institutional economics to shed light on the historical development of legal and institutional factors that underpin money creation and management. It also makes regular and extensive use of empirical methods to analyze original historical datasets and test hypothesis. Readers interested in this specific aspect can refer to the chapters on financial markets, payments systems, financial panics and crashes, financial systems, and the Great Depression in the recent *Handbook of Cliometrics* (Diebolt and Haupert 2016).

We have structured this handbook in ten broad thematic parts: the historical origins of money; money, coinage, and the state; trade, money markets, and international currencies; money and metals; monetary experiments; Asian monetary systems; exchange rate regimes; monetary integration; central banking and monetary policy; and aggregate price shocks. In the following pages, we offer for each part some historical context, a few key insights from the literature, and a brief analytical summary of each chapter. Our aim is not to provide a survey of the existing body of knowledge (a daunting task beyond the scope of this introduction) but rather to draw a map that hopefully will help readers to organize their journey through this very wide and diverse research area.

Historical Origins of Money

Since the nineteenth century, the debate on the origins of money and the sources of its value is dominated by a controversy. One school of thought regards money as a discovery of rational private agents bargaining in incipient market economies: it was

invented in order to reduce the high transaction costs implied by the use of barter in the exchange of goods and services. While in principle any commodity could be used for that purposes, the advantages of precious metals (durable, divisible, portable, saleable) led to the emergence of coins as the standard form of currency. Its value (i.e., its purchasing power) was intrinsically determined by its metallic content. The transaction cost theory of the origins of money, which can be easily found in standard economics textbooks, is a hypothetical and formalist narrative (Wray 2012) – a creation story (Desan 2014) constructed around the function of money as a medium of exchange. This story is both logically and historically inconsistent (Goodhart 1989, 1998). It neglects the substantial costs of verifying the quality (fineness) and quantity (weight) of metal, as well as the historical evidence that those costs were reduced by stamping on coins a quality guarantee in a minting process regulated, supervised, and often directly operated by the state. In turn, credible minting required the use of legitimate violence against the theft of metal inventories, as well as operators with a sufficiently long time horizon to address the time inconsistency problem of quality preservation against the incentives to debase the currency for private enrichment. These requirements could be met only by settled and strong governments.

This view connects to the basic tenets of a second school of thought, known as chartalism, which is the foundation for the so-called Modern Money Theory (Wray 2014). This intellectual tradition claims that the power of the state played a central role in the evolution and use of money. In fact the role of the state goes beyond the certification of metal quality. The creation of money increases the fiscal capacity of a sovereign ruler: it makes easier to levy direct and indirect taxes and provides additional fiscal revenues through seigniorage (a tax on minting services) (Goodhart 1998; Le Rider 2001). This perspective is much more consistent with archeological and numismatic evidence. The origins of money – though "lost in the mists of time" (Wray 2012) – are related to the emergence of complex political economies with centralized power, social hierarchies, and distributional economies (Earle 1991, 2002). Written documentation reveals the use of grain and precious metals as weight standards and value equivalents under public oversight in the accounting systems of ancient Mesopotamian cities, with a highly centralized bureaucratic management and the extensive use of credit to finance long-distance trade (Van de Mieroop 2004). This was related to the development of mathematical thought that allowed small literate elites to use numeration and metrological systems to record, manage, and predict the value of private and public assets, including income rights (Robson 1999, 2007).

Since its remote origins, therefore, money emerged as a "constitutional project," to use Christine Desan's definition – a mechanism of governance through which stakeholders standing at the hub of a community (i.e., rulers) mobilized resources to produce "collective endeavors" (infrastructures, defense) based on labor services and in-kind contributions. In return, they released tokens or receipts marked in standard units with real fiscal value, as they were based on obligations to contribute to the collectivity through tribute, rents, fees, tithes, or penalties. By issuing liabilities to buy goods and services and by credibly committing to redeem them in

discharge of obligations, ancient rulers developed the power to spend and tax. At the same time, as the value of those liabilities was recognized by everybody owing regular tribute to the same ruler (their common creditor), they became transferable, provided valuable cash services, and allowed stranger parties to complete transactions. Money was basically a product of fiscal engineering (Desan 2014).

► Chapter 2, "Origins of Money and Interest: Palatial Credit, Not Barter" by **Michael Hudson** illustrates the administrative and fiscal purposes of a grain-silver bimonetary standard in palatial societies of the Neolithic and Bronze Ages, such as Mesopotamia and Egypt. Grain was used as unit of account to calculate values, measure labor time and land yield, and as means of payment in agricultural and handicraft activities. Silver was used as means of payments for taxes and fees and for long-distance trade. The stability of the grain/silver ratio was legislated by the king and guaranteed by temples. Transactions involved debt relationship related to the cyclical patterns of harvests and trade, and interest rates on silver-denominated official credit were regulated administratively. Commensurability, Hudson argues, represented the key innovation that marked the departure of palatial money from "primitive," special-purpose money used for wergild (the compensation paid by offenders to injured parties), dowries (the price of a bride), or donations, which included also animals, textiles, and jewels. As public oversight extended to silver quality standards and minting to guarantee specified degrees of purity, coinage became widespread in the classical antiquity. Through trade, Near Eastern weights, measures, and practices – including interest-bearing debt and the use of coinage to pay taxes and fees – penetrated in the Mediterranean world.

The latter is the focus of **Colin Elliot**'s (► Chap. 3, "The Role of Money in the Economies of Ancient Greece and Rome"). In the Mediterranean area, the coinage of different metals (gold, silver, copper, bronze), supplemented by in-kind moneys (grains and other commodities), emerged in the political context of the Greek city-states. Hellenization promoted the development of market exchange and financial institutions. On the contrary, war-related fiscal shocks seem to have been critical for the monetization of Rome, which adopted a coherent Greek-style coinage system only in response to mobilization in the Second Punic War. Later one, Roman Emperors established their rights to coin money and fix the coins' legal value and legislated to protect their exclusive prerogatives against criminal offences. The territorial expansion of the Empire challenged monetary unification, leading initially to "a hodgepodge of monetary systems and coinages," possibly to enhance the payment of taxes and tributes in conquered territories. Only later, coinage was centralized under strict legal tender laws, but the monetary system was exposed to frequent shortages, debasements, recoinages, and reforms. In the third century AD, laws commanded the acceptance of official coins and prohibited their trade and assay in black markets by independent money changers, which can be interpreted as a signal of weak trust in the quality of coinage by the public. The progressive demonetization that was generally associated with the territorial contraction of the Empire confirms the intimate connection between money and the quality of political institutions in the ancient world. After the retreat of the Empire from Britain, for instance, the circulation of Roman currency broke down, market exchange

collapsed, and no alternative medium of exchange filled the void – a dramatic "fall in the sophistication of material culture" that provides an enlightening natural experiment on the origins of money (Desan 2014).

The use of money as "value equivalent" in the measurement of economic obligations between private agents or between them and public authorities is so deeply embedded in the Western tradition since ancient times that it is difficult to imagine for it having a different role. When colonial conquest brought colonizers in contact with so-called simple (i.e., small-scale, stateless) societies, in which "odd and curious" objects (feather, beads, cowries) seemed to serve some or all functions of money, their "coin consciousness" shaped by Western historical experience failed to understand the role of this "primitive money" in local social and political relationships. This issue lies at the heart of ▶ Chap. 4, "Primitive and Nonmetallic Money" by **Bill Maurer**, which discusses critically the contribution of anthropologists and substantivist economists to a deeper understanding of the social roots of money. In this perspective, practices such as bridewealth and wergild can be thought of as "social currencies" that built or repaired social relations. Their role was less about setting value standards and more about playing differentiated social and political games. For instance, in a system of mutual obligations (i.e., equal bridewealth), the gift was not exchanged for a wife (i.e., did not set a person's value according to a standard), but substituted for her in the system of relations; in a hierarchical system (i.e., unequal bridewealth), the gift was used to assert social ranks. As Maurer explains, the process of colonization can be seen as the interaction of competing political hierarchies that affected also monetary standards. When European merchants engaging in slave and palm oil trade with West Africa flooded the region with cowrie shells – the local "currency" – from Zanzibar, the burst of the cowrie shell bubble led this "primitive money" to lose its value. However, this cowrie inflation was not driven by excess supply, but by falling demand, as colonizers imposed tax obligations and forced local subjects to use the Western monetary standard to meet them. In this sense, money measured the interaction between ascending and declining state power, which can be understood also as a "contest over standard setting."

Money, Coinage, and the State

Medieval Europe inherited from the Greek-Roman world the key notions that a coin's value rests on its content of precious metals and that the exclusive power of giving money its value depends on its legal enactment under the authority of a sovereign body (*ius cudendae monetae*). By proclamation, rule, and law, a new coinage system emerged in the Middle Ages: the so-called free minting system in English parlance (Allen 2012; Desan 2014). Since the monetary reform of Charlemagne in 794 AD until the early modern era, European coinages evolved from a monometallic system with virtually one single denomination (the silver penny) to pluri-metallic systems based on gold (for high-value coins to be used in trade), silver, and copper alloys (such as billon and bronze, to provide small change for daily transactions) and an increasing range of denominations (Volckart 2018).

Political fragmentation led to monetary fragmentation, as local rulers successfully claimed coinage rights. The exercise of monetary sovereignty included establishing a territorial unit of account and its fractions (in which prices were expressed), setting the terms for the coinage of domestic coins (the price of bullion at the mint and the number, finesse, weight, and denominations of coins produced out of it), and levying taxes on minting as a source of fiscal revenues (brassage and seigniorage). Rulers also established the conditions under which foreign coins could be accepted as means of payment in domestic transactions. Coins carried no marks of value and could circulate either "by tale" (i.e., on the base of their legal or "extrinsic" value) or alternatively "by weight" (i.e., an "intrinsic" value determined by private agents on the base of their metal content). The medieval and early modern history of money was characterized by governments' attempts to limit the scope of this privately managed circulation (Boyer-Xambeau et al. 1994). This also generated intense legal controversies, in which a nominalist approach eventually prevailed (Fox and Ernst 2016). While in the early Middle Age the legal tradition widely accepted circulation "by weight," by the seventeenth century, the legal foundations of circulation "by tale" were well established (Sargent and Velde 2003).

Georges Depeyrot (▶ Chap. 5, "Monetary System of the "Ancient Régime" (Third to Eighteenth Centuries)") reconstructs chronologically the historical evolution of monetary systems in Europe and the Mediterranean basin from the late Roman Empire to the early modern period. They were characterized by the coexistence of a "Roman system," in which coins' legal value was fixed in terms of units of account, and an "Alexandrian system" in which coins were basically regarded as fractions of bullion and their value "floated" with the relative price of the two metals (the bimetallic ratio) as determined in the market. He also surveys the main technological innovations (such as the screw press and the rolling mill) that allowed the coinage of heavier coins and reduced scope for counterfeiting and clipping. **David Fox** (▶ Chap. 6, "Money, Law, and Institutions") explains how medieval legal systems in Britain and continental Europe elaborated on a common body of principles that had been first clearly articulated in Roman monetary laws and converted them into a practical system of rules especially suitable for the commercial and monetary conditions of the time. Whether coins should be valued "by weight" or "by tale" for the purpose of discharging debts was a legal controversy that jurists and courts had to deal with for a long time. English common law quickly embraced a nominalist approach, under which debt was denominated in terms of a generic unit of account; this prevented any judicial inquiry about changes in the intrinsic value of the coinage between the stipulation and the finalization of contracts. On the contrary, in continental traditions the notion that debt should be repaid in coins with the same intrinsic value prevailed until the late seventeenth and early eighteenth centuries. This implied that coins were legally interpreted as a special kind of bullion, whose quality (weight and finesse) was certified by the sovereign. Legal differences may reflect the relative stability of the English coinage until the mid-sixteenth century (before the Great Debasement of 1544–51), compared with the recurrent debasements that characterized France and other continental polities.

In fact, the sovereign's exclusive right to fix the valuation of the coins implied a prerogative to change the monetary standard, i.e., to debase the currency. Debasements implied an increase in the ratio between the nominal and intrinsic value of money, as a consequence of a reduction in bullion content (finesse), a reduction in coins' weight, or an increase in their official value. Historians' traditional view of debasements indicted them as examples of disrupting monetary policies driven by war-related fiscal motivations – a form of hidden and regressive taxation that harmed earners of nominal rents and forced the extraction of seigniorage through recoinage (Spufford 1991). Others pointed to the "monetary" nature of some episode of debasements, as an attempt to stabilize money supply during the "bullion famines" of the late fourteenth and early fifteenth centuries (North 1994) or to address recurrent shortages of small-denomination coins ("the big problem of small change") providing liquidity services that large-denomination coins failed to perform (Redish 2000; Sargent and Velde 1999, 2003). Recent research, in turn, explains debasements as policies that aimed to address temporary misalignments of local gold-silver ratios (which triggered bullion flows) or to reduce monetary uncertainty created by the circulation of "bad" money – i.e., coins with lower intrinsic value – by periodically "cleaning" money markets, thus reducing transaction costs. The new evidence also suggests that, while gold debasements were essentially driven by "monetary" motivations, silver debasements were indeed related to warfare and used much more extensively by princely polities than by city-states or monetary unions formed by different cities (such as the Wendish Monetary League led by Hamburg) (Chilosi and Volckart 2010, 2011). However, recent comparative research over the very long run, from the Middle Age to World War I, supports the political, seigniorage-driven origins of monetary instability: depreciations of the domestic monetary unit were associated to war, states with intermediate fiscal capacity, and the lack of institutional constraints on rulers that prevented predatory monetary policies (Karaman et al. forthcoming).

Debasements followed a regular pattern. They coincided with unusually large minting volumes and rising seigniorage rates, which allowed rulers to extract huge fiscal revenues. After debasement, old and new coins circulated jointly and were valued "by weight" (i.e., according to their intrinsic content of precious metal). This poses a challenge to monetary theory, as if coins circulated by weight, the opportunity to change heavy coins into light coins at a higher cost did not provide well-informed agents with any additional incentive to bring bullion to the mint (Rolnick et al. 1996). Alternative explanations of this "debasement puzzle" – such as circulation of coins "by tale" (according to their legal tender value) or sluggish price and wage adjustment (Bordo 1986; Sussman 1993; Sussman and Zeira 2003) – do not seem sufficiently supported by empirical evidence. **Oliver Volckart** (▶ Chap. 7, "Premodern Debasement: A Messy Affair") suggests that bullion merchants kept supplying gold and silver to the mint as only small groups of experts in the economy became immediately aware of the alteration. The wider public of illiterate and disenfranchised consumers continued to use coins as if they were unadulterated or simply had no power to refuse to accept them in payment for wages and goods. High information asymmetries also prevented rulers from exerting effective control of

mint officials (who could exploit their autonomy in order to debase coins on their own account). These costs also made it impossible to monitor markets closely enough to enforce regulations, while weak border controls reduced their ability to prevent the import of foreign coins with similar nominal value but lower intrinsic value. In turn, the resulting (and highly profitable) "trade in coinage" created incentives for governments to issue inferior copies of their neighbors' coins – a practice that had the same consequences as a debasement – and forced the affected governments to follow suit by debasing their own coinage in order to reduce the disruption of market transactions.

The joint circulation of "heavy" and "light" coins after debasement challenges a well-established notion in monetary economics: that bad (overvalued or underweight) money tends to drive good (undervalued or full-weight) money out of circulation. This notion is widely known as Gresham's Law, from the name of Sir Thomas Gresham. A financier and monetary advisor to Tudor's monarchs, in 1558 Gresham drew the attention of young Queen Elizabeth I on the consequences of the Great Debasement of English coinage carried out under her father Henry VIII and half brother Edward VI. Observing that "good and bad coin cannot circulate together," he suggested that the deliberate reduction in the metallic content of silver coins had resulted in a massive export of gold coins. Since the mid-nineteenth century, Gresham's Law has remained a subject of controversy among economist, as the historical evidence provides a wide selection of counterexamples both in bimetallic and monometallic regimes. A qualified version suggests that it holds only if the exchange rate between bad and good money (their par price) is fixed (an argument proposed by Milton Friedman and Anna Schwartz in their classical monetary history of the USA). Another restatement – which tends to reject its general validity as a fallacy – is that if both good and bad money must be accepted at par value under a legal tender law, prices are set in terms of the overvalued currency, but only small denominations of the undervalued currency (circulating at a premium) disappear as agents economize in transaction costs when using large denominations as medium of exchange (Rolnick and Weber 1986). A third perspective – which reinstates the validity of Gresham's Law – contends that legal tender legislation tends to favor "bad" over "good" money by increasing the risk and costs for agents to reveal their monetary preferences, i.e., by sanctioning attempts to place a discount on bad money or to refuse it altogether (Selgin 1996). The issue is somehow theoretically elusive, as both cash-in-advance models (Sargent and Smith 1997) and search-based models (Velde et al. 1999) of commodity money systems predict multiple equilibria. **George Selgin** (▶ Chap. 8, "Gresham's Law") explains monetary selection as the outcome of a coordination game between sellers and buyers in imperfect currency markets. As legal tender laws made it costly to assign distinct nominal values to coins with different intrinsic values, the exchange game between sellers moved the economy into a "bad" money equilibrium in which only adulterated coins or irredeemable paper notes are circulated. But even in the absence of strict legal tender laws, the concentration of private information about intrinsic value in the hands of money changers and merchants could lead to similar results. As Selgin explains, the same logic holds in bimetallic regimes in which the

ratio of gold and silver prices in bullion markets deviated from the ratio of their official values at the mint, which explains the inadvertent switch of England to a de facto Gold Standard in the early eighteenth century and the swings to de facto silver standard and then to a de facto Gold Standard of the USA in the first half of the nineteenth century.

Trade, Money Markets, and International Currencies

In the Middle Age, the wide diversity in climate and resource endowments across European regions, the geographical configuration of the continent with long coastlines and plenty of navigable rivers, and the legal protection of commercial activities guaranteed by rulers enhanced the development of large-scale, long-distance trade organized on a dense network of commercial centers. In the absence of an efficient system of international payments and fund transfers, trade would have been seriously constrained by market frictions. Trading with distant centers implied strong information asymmetries about parties' creditworthiness. Political fragmentation might have favored economic development through flexibility and institutional competition, but it also posed challenges to the deepening of a market economy. Each polity had its coinage and unit of account, and mints often competed for seigniorage within its borders, which affected coinage uniformity. The bewildering multiplicity of small and large coins in circulation with different weight and fineness created serious market frictions. Shipping coins was costly and risky, and trade coins had to be assayed and protected. Money changers had large opportunities for profits from arbitrage and strong incentives to cull out "good" coins, which affected the volume and quality of money supply. Epstein (2001) contended that the reduction in monetary fragmentation allowed by currency unions – both between sovereign polities and between cities adopting a common currency supplied by the same territorial ruler – played a critical role in market integration and growth in the late Middle Age. The reconstruction of local gold-silver ratios (an indicator of prices paid in local money markets) in commercial centers along the transcontinental trade route that linked Northern Italy and South East England and the empirical analysis of their spreads confirm that monetary diversity had adverse effects on bulk trade. The creation of currency unions favored the integration of money markets; at the same time, unions were to a certain extent endogenous – that is, tended to emerge when the member centers were already linked by strong trade connections (Boerner and Volckart 2011).

Costs and frictions generated incentives to monetary innovations: the use of privileged types of private debt (IOUs) with special characteristics that conferred upon them the ability to achieve finality in settling transactions (Kahn and Roberds 2007). These can be considered as new forms of "inside money" as opposed to metallic "outside money." **Meir Kohn** (▶ Chap. 9, "Money, Trade, and Payments in Preindustrial Europe") analyzes how a European system of international payments and remittances based on inside money developed since the Middle Age. Initially, banks in major commercial centers emerged as trusted third parties whose liabilities

(deposits) could be used to settle payments between strangers. Later on, to fix the problems created by deposit banks' instability, original money-like instruments (promissory notes, letters obligatory, bills of exchange) were developed to transfer funds across banking places connected by the extensive branch networks of international trading companies and merchant banks. In major banking places, such as Antwerp in the sixteenth century, organized foreign exchange markets emerged, and new payment systems based on the assignment of private third-party debt were developed, thanks to legal innovations that allowed its transferability and negotiability under the joint contingent liability of all parties involved in the transaction, which generated strong incentives to screen the quality of the circulating debt. A successful alternative was a payment system based on the exchange of negotiable instruments with public banks that offered transferable accounts redeemable in hard money – an innovation pioneered by Venice and quickly imitated by Amsterdam and other North European commercial cities. Both practices facilitated multilateral trade, allowed commercial specialization, and provided international liquidity facilities that critically contributed to economic development. **Pilar Nogues-Marco** (► Chap. 10, "Money Markets and Exchange Rates in Preindustrial Europe") dwells deeper into the functioning of international money markets based on bills of exchange. An instrument of remittance and credit at the same time, bills were used both in commercial and purely financial transactions between merchants as a practical means of currency exchange and circumvention of the Church's prohibition of usury. Nogues-Marco explains how exchange rates fluctuated between a floor and a ceiling defined by the "specie points" – the levels beyond which the international transfer of money by shipping bullion became profitable. Arbitrage was enhanced by the generalized liberalization of bullion movements at the end of the seventeenth century, which favored the efficiency of the specie-point mechanism and the integration of money markets, as in the case of London and Amsterdam. The network of intercity monetary linkages, reconstructed on the base of price cross-quotations (a proxy for market liquidity), draws a "geography of money" in mid-eighteenth-century Europe strongly dominated by monetary agglomeration in Amsterdam, London, and Paris. This triangle provided the lynchpin of a multilateral settlement system that reached out to minor hubs in the north (Hamburg) and the south (Genoa). Given the over-the-counter nature of markets based on bills of exchange and the fact that interest was embedded in the exchange rate at maturity, money market interest rates were not quoted and must be inferred indirectly from the price ratio of bills maturing at sight and bills maturing at a certain future date. Estimated "shadow" interest rates suggest a high and increasing level of money market integration between London, Amsterdam, and Paris, the most liquid markets of the period.

In the second half of the nineteenth century, the London money market, increasingly based on bankers' acceptances (bills accepted by merchant banks on behalf of foreign clients and discounted by specialized intermediaries known as discount houses), reached a truly global dimension. In the heydays of the classical Gold Standard and the First Globalization before 1914, the sterling-denominated "bill on London," quoted and traded in financial centers worldwide, became the key instrument to finance international trade and capital movements, turning the City of

London into "the center of world liquidity" (Flandreau and Jobst 2005; Flandreau and Ugolini 2013). In the interwar period, monetary and financial instability, the contraction of trade, and the widespread introduction of exchange controls and barriers to capital mobility brought this experience to an end. An international money market with global reach re-emerged in the 1960s once the world returned to a situation of generalized external convertibility under the Bretton Woods system. The market, centered again on London but connecting financial centers worldwide, was now based on different money-like instruments (short-term bank deposits) mainly (although not exclusively) denominated in US dollars – the so-called Eurodollar market. **Stefano Battilossi** (▶ Chap. 11, "International Money Markets: Eurocurrencies") analyzes the economic and regulatory factors that drove its secular expansion and its key role in enhancing the mobility of short-term capital as well as in the creation and distribution of international liquidity through global interbank connections. Its offshore nature and its implications for monetary policies and financial stability were a recurrent source of concerns for central bankers and policymakers, who tried with little success to impose multilateral controls to limit its scale and scope. As Battilossi shows, since the late 1960s, large jumps in the spread between Eurodollar and US Treasury bill rates in coincidence with unanticipated financial crises signaled conditions of funding and market illiquidity similar to those observed in the Great Financial Crisis of 2007–2008. ▶ Chapter 12, "The Asian Dollar Market" by **Seung Woo Kim** focuses on the historical development of the Eurodollar market in Asia and the deliberate policies that promoted the emergence of Singapore as the dominant offshore financial center in the Asia-Pacific region.

The dominance of the British pound and the US dollar in the international money markets of the nineteenth and twentieth centuries reflects their role of key international currencies. Before 1914, other currencies (the French franc, the German mark) enjoyed an international status, reflected in the development of sizeable money markets and their use as reserves by other central banks. The relative decline of the British pound against the rise of the US dollar was a secular process during which both currencies continued to be extensively used in commercial and financial transactions, as well as central banks' reserves worldwide. More recently, the Japanese yen, the Euro, and the Chinese renminbi began to rival with the US dollar. In sum, history suggests that the international monetary and financial system had always a degree of multipolarity. This evidence challenges conventional economic models and historical narratives based on network, lock-in, and inertia effects, according to which the advantages of incumbency leave room only for one dominant currency (Chitu et al. 2014; Eichengreen 2010; Eichengreen and Flandreau 2009, 2010; Eichengreen et al. 2018; Flandreau and Jobst 2009). **Barry Eichengreen** (▶ Chap. 13, "International Currencies in the Lens of History") explores the structural factors that explain the international status achieved by different currencies in different historical periods. These include the size of the issuing country's economy, the network externalities generated by the volume of its international transactions, the stability of its value over time, the size and liquidity of its financial markets, and the capacity to project military and diplomatic power. At the same time, history

suggests that the status of international currency is not a natural monopoly: portfolio diversification explains why different international currencies can coexist and transitions across dominant units can be prolonged. This analytical framework can be successfully applied not only to the British pound in the nineteenth century and the US dollar in the twentieth century. It can explain also the dominance of the silver drachma in the Athenian Empire, of the gold aureus and the silver denarius in the Roman Empire, and of the solidus in the Byzantine Empire; the Genoese genoin, the Florentine florin, and the Venetian ducat in the Middle Age; and the Dutch guilder in the seventeenth and eighteenth centuries. The lens of history, Eichengreen suggests, shed also light on the future: as the deepening of modern financial markets and advances in information technologies reduce the costs of currency diversification, competition between the incumbent dollar and its prospective rivals is likely to increase.

Money and Metals

In a hard money economy based almost exclusively on coins, the expansion of money supply was strongly constrained by the available stock of metals. In medieval Europe, coinage was heavily dependent on silver and copper from mines in South Germany and Central Europe. When mining production declined (i.e., between the 1320s and the 1470s), economies experienced periodic "bullion famines," that is, periods of extreme scarcity of circulating coins relative to the transaction demand for money (although this notion is not entirely uncontroversial; Sussman 1998). These were occasionally exacerbated by hoarding and constraints on bullion trade and led to protracted periods of price deflation (Miskimin 1984; Spufford 1991; Munro 1992, 2016). In turn, the silver-mining boom of the late fifteenth century significantly expanded the overall stock of bullion in Europe and facilitated the expansion of the European economy and the recovery of prices (Munro 1991, 1992, 2003). In a similar fashion, the mining boom of silver in Spanish America and the huge inflow of silver coins into Spain are traditionally associated with a sustained increase in price level across Europe – the so-called price revolution of the sixteenth century.

Since the pioneering studies by Earl J Hamilton in the 1930s, monetary interpretations of price dynamics in pre-industrial Europe used the analytical framework of the quantity theory of money (MV=PY in its Fisherian formulation) (Munro 2008). The traditional view hinged on the logic of the price-specie flow model, in which Spain's inflation and trade deficit generated synchronized upward shifts in the monetary base and prices of its trading partners. Challenged by internal inconsistencies (the lack of synchronization between specie inflows and inflationary episodes, the evidence of massive exports of American silver to Asia) and alternative explanations (emphasizing especially urbanization and demographic growth), monetary interpretations were reformulated with a focus on estimated changes in the velocity of circulation. However, the claim that velocity increased as a consequence of specialization and urbanization (Goldstone 1984; Lindert 1985) – an example of Smithian growth – runs counter to the notion that in a modernizing economy with a

rising level of monetization, velocity should fall as a consequence of increased demand for money (Mayhew 1995, 2013). The lack of reliable data on money stocks (M) and the level of economic activity (Y) preclude robust estimates of V, which is obtained as a residual (Nicolini and Ramos 2010).

In the case of pre-industrial England, for instance, a positive impact of American metals on the actual level of monetization of the economy seems incompatible with qualitative evidence that emphasizes a chronic scarcity of low-denomination coins (especially for wage payments) well into the eighteenth century (Muldrew 2001, 2007; Muldrew and King 2003; Selgin 2008). **Nuno Palma** (▶ Chap. 14, "American Precious Metals and Their Consequences for Early Modern Europe") reviews the findings of the most recent research (including his own: Palma 2018) to show that in fact American silver and gold did contribute critically to relieve constraints on monetary expansion in early modern Europe. In the case of England, this claim is supported by a range of quantitative evidence, from the reconstruction of minting volumes and coin stocks, to the presence of Spanish-American silver in English coinage detected by chemical analysis, to the increased availability of small-denomination coins suggested by random coin finds. However, Nuno argues it is reasonable to assume similar developments in other receiving countries such as France and Holland. In turn, higher levels of monetization and liquidity, by reducing transaction costs, enhanced market deepening, agglomeration economies, and fiscal capacity, providing a ground for sustained growth. In contrast, in first-order receivers, such as Spain, the continuous arrival of American metals affected negatively both inflation (an example of Dutch disease) and the quality of political institutions, thus acting as a curse in the long run (Prados de la Escosura and Álvarez Nogal 2013).

Among the positive effect of American metals for Europe, Palma also highlights their critical role in enhancing trade between Europe and Asia, which was conducive to deep changes in households' demand for consumption and labor supply in exchange for monetary wages (the so-called industrious revolution) in some regions of Northern Europe (De Vries 2006). This issue connects with the stream of research on the role of international silver flows in the emergence of a deeply integrated global economy in the sixteenth century (Flynn 2013). In this literature, the rising demand for silver (coming either from Europe or Japan) by a fast-expanding Chinese economy and the critical role of foreign currency supply in its monetization play a central role (Flynn 2013; Flynn and Giraldez 1995a, b, 2002; Flynn et al. 2003; von Glahn 1996, 2016). This global perspective on monetary history connects with the comparative analysis of economic development in Europe and Asia – the so-called Great Divergence debate (Allen et al. 2011; Broadberry and Gupta 2006; Pomerantz 2000). This is the focus of **Alejandra Irigoin**'s ▶ Chap. 15, "Rise and Demise of the Global Silver Standard," which provides an exhaustive summary of state-of-art knowledge about mining production, minting policies, and metal exports from Spanish America. She also outlines patterns of global trade in silver coins and bullion and their impact on the fluctuations of gold-silver ratios. Against traditional interpretations of silver as a pure commodity in long-distance trade, Irigoin especially emphasizes the monetary role of Spanish dollars as the main international

means of payment in the global silver standard of the seventeenth and eighteenth centuries. This also suggests that imported currency played a critical role in the monetization and specialization of Asian economies and gave a critical contribution to global Smithian growth.

Monetary Experiments

As we have seen in Part "Trade, Money Markets, and International Currencies," the advantages of new forms of "inside money" over hard money in terms of costs of production and flexibility to accommodate the needs of a deepening market economy were widely recognized since the Middle Age. However, payment based on the transfer of endorsed money-like private debt was limited to commercial elites of sufficient reputation. Similar limitations applied to local payments by checks or drafts, non-negotiable but payable on demand between businessmen who held account balances with deposit banks. The late seventeenth century saw the emergence of a monetary innovation that facilitated transactions between agents of minor social and economic standing: the issue of liabilities by deposit banks, in the form of large-denomination promissory notes payable to the bearer, against the purchase of short-term assets (e.g., bills of exchange). These would be universally accepted in payments, provided that banks had verifiable wealth and enjoyed sufficient economies of scale in collecting information about the creditworthiness of their counterparties in the asset market. The system formed by goldsmith bankers in London, based on co-monitoring and interbank clearing of checks and notes, is an example of a successful "inside money" experiment (Quinn 1997). However, banknotes' role as payment instruments became established only once their volume reached sufficient scale to generate wide network externalities and increase their liquidity. This was achieved by the massive issue of bearer notes by the Bank of England (some of them with sufficiently low denominations) to acquire government debt. This converted banknotes into claims backed by safe assets and readily redeemable in coins, although they would acquire the status of legal tender only in the 1830s (Quinn and Roberds 2003, 2008). In retrospect, the establishment in the 1690s of the Bank of England with the primary function of a bank of issue, and the circulation of its banknotes (paper promises of value borrowed by the government from the Bank) as currency, was a successful monetary experiment that paved the ground to the invention of modern money (Desan 2014; Kleer 2008, 2015 and 2017).

Monetary and financial experiments flourished in the laboratory of the seventeenth and eighteenth centuries (Neal 2000). A stable system of fiat money (paper money not convertible into coins) was introduced by the Bank of Amsterdam, a public bank with the near-monopoly of domestic large-scale payments, in the late 1680s. On one hand, the Bank issued negotiable receipts against the deposit of large-value trade coins while at the same time removing the traditional right of withdrawal of coins on existing deposit accounts. This reduced the Bank's vulnerability to runs, preserved the stability of the value of the unit of account (the florin), and enhanced

its role as the dominant international currency of the time (Quinn and Roberds 2014). Other experiments – such as the attempt by John Law to implement a full-scale fiduciary currency system to stimulate the French economy with abundant credit and efficient payments while financing large government deficits in the 1710s (Neal 2012; Velde 2007, 2008) – were ill-fated. As **Francois Velde** illustrates in his ▶ Chap. 16, "Experiments with Paper Money," paper money (an innovation widely used by Chinese imperial dynasties from 960 AD to the 1430s) suffers from a specific weakness that makes it difficult to manage: a permanent threat of overissue, depreciation, and inconvertibility, exacerbated by the pressing needs of government finance, especially in wartime. Two models of paper money emerged from the experiments of the eighteenth century: one based on the issue of large denominations by private banks for wholesale payments and the other one based on issue of small denominations by the state for retail payments. The nineteenth century demonstrated the superiority of the former.

British North American colonies were especially affected by a chronic scarcity of coin circulation, and original media-of-exchange methods were developed by using commodities with different degrees of moneyness (Sylla 1982). In his ▶ Chap. 17, "Money and Prices in Colonial America" by **Farley Grubb** shows how, under the pressure of emergency wartime expenditures, colonial legislatures ultimately resorted to an original form of paper money: the emission of bills of credit (either directly or through land banks), in most cases bearing no interest, to be redeemed at a future date (maturity) against sinking funds, tax revenues, or land-backed loans. Structured like bearer zero-coupon bonds, colonial paper money traded below par values. Elaborating on insights from his own research (Celia and Grubb 2016; Grubb 2016a, b), Grubb suggests that, contrary to previous interpretations based on the quantity theory of money (McCallum 1992; Rousseau 2007) or the "backing" theory (Smith 1985, 1988; Michener 1987, 1988, 2015; Michener and Wright 2005), the discount on par value did not reflect currency depreciation but rational bond pricing, which implied time discounting according to their maturity structure.

The US War of Independence was largely financed by inconvertible and rapidly depreciating paper money ("continental dollars"). After independence, the US Constitution prohibited states to issue money, and the Congress regarded the issue of federal paper currency with great suspicious. As a consequence, banknotes issued by private banks against the collateral of state or federal debt played a critical role as media of exchange in an undermonetized economy, rivalling in volume with federal currency during the whole antebellum period. Their circulation beyond local economies (usually below par value) was facilitated by special institutional arrangements (clearinghouse associations) that enhanced the clearing of banknotes and issued special forms of emergency paper money (clearinghouse loan certificates and notes, a joint liability of the member banks) to stabilize liquidity during financial panics. As **Matthew Jaremski** shows in his ▶ Chap. 18, "Privately Issued Money in the United States," recent empirical research suggests that the liquidity benefits of banknotes most likely outweighed the market frictions created by their wide variety and that banknotes' discount efficiently reflected redemption costs and the default risk of the issuing banks. Research also identifies an apparent paradox: banknotes

were highly profitable, but on the aggregate their volume remained below the maximum level allowed by outstanding stock of collateralizable bonds. Recent studies suggest that this "note issue puzzle" may reflect cross-regional differences in opportunity costs of investing in bonds.

▶ Chapter 19, "Money, Prices, and Payments in Planned Economies" by **Michael Ellman** brings us into a different kind of monetary experiment: money in the Soviet Union. Although the survival of money was an aberration in light of the Marxist-Leninist theory, Soviet rulers justified it on practical grounds, first as a necessary evil in the transition from socialism to communism and later on as a necessity for planning. Soviet fiat money performed (at least in part) the functions of unit of account (e.g., production targets were expressed in monetary rather than in physical units) and medium of exchange. At the same time, monetary institutions reflected the unique characteristics of an administrative-command economy. The system was based on a central bank (Gosbank) operating on the base of planned cash emissions. Circulation was split into two separated circuits: noncash ("passive") money for payments between state-owned enterprises (in practice a state giro system, also used to control firms' adherence to state plans) and cash ("active") money used by firms to pay wages and by workers in state retail trade and in the free market for food. After WWII, overissue of cash and noncash money, mainly to finance large budget deficits, led to high inflation in the free market and acute shortages in the state retail market (where prices were fixed) – a chronic situation of "shortageflation," which the reforms of the late 1980s exacerbated, leading to the collapse of the system.

Massimo Amato and **Luca Fantacci** (▶ Chap. 20, "Complementary Currencies") deal with past and present experiments in monetary plurality: the development of alternative monetary instruments (designated as complementary, social, parallel, or community currencies) that challenge the domestic monopoly of the currency issued by national monetary authorities. Monetary plurality was pervasive in medieval and early modern periods, in which (as we have seen in previous chapters) means of exchange were separated from units of accounts, different monies and monetary circuits performed different functions, and foreign currencies circulated side by side with domestic currencies (Kuroda 2008a). On the contrary, the overlapping of political and monetary spaces that led to the dominance of territorial currencies is a relatively recent phenomenon dating back to the nineteenth century (Helleiner 2003). Before World War II, monetary "supplements" typically emerged in response to economic crises. The most relevant examples were local currencies and stamp scrips issued by local authorities in Germany, Switzerland, and the USA during the Great Depression. This practice resembled the issue in early modern era of token coins, scrips, and "necessity money" by local churches, cities, merchants, and entrepreneurs, especially in periods of war and economic hardship. The recent proliferation of heterodox monetary circuits (some backed by official currencies, other based on local clearing systems) since 2007 confirms that depressions are especially conducive to the emergence of complementary currencies, whose main purpose is to enhance the velocity of circulation of money. This clearly differentiates them from recent cryptocurrencies (i.e., bitcoins), which presume to offer a radical

alternative to official currency but are more akin to commodity mainly held for speculative purposes.

Asian Monetary Systems

For a long time, the Western perception of monetary institutions in pre-modern China was influenced by the view of late nineteenth-century British observers, who could discern only a system of "chaotic eccentricities." In fact, the late imperial monetary system was the shadow of the flexible silver-based standard established under the Ming dynasty's "Silver Century" (1550–1650) – a monetary policy choice that met the requirements of an economy increasingly integrated in global trade and made possible by massive purchases of Spanish-American silver (von Glahn 1996; 2013). By the time the British navy forced the opening of Chinese ports in the 1840s, the unit of account (*tael*) had become territorially dishomogeneous, as it was expressed in terms of unminted silver ingots of similar weights but different shapes, purity, and finesses. At the same time, low-value copper cash coins, a form of countable currency used for small and local transactions, were strung together into another unit of account (*tiao*), which also varied by regions and trades. Domestic coinage was supplemented by Spanish silver dollars, trading at a premium on their silver value; it was used for large transactions and long-distance trade, but its penetration was limited to southern and coastal regions. In order to provide an anchor for such an amalgam of currencies, local imaginary units of accounts, usually linked to tax collection, were introduced. The result was, in Debin Ma's words, "a nearly infinite set of cross exchange rates among the imaginary units, between the imaginary units and actual currencies, and among the actual currencies" – a reflection of "both the arbitrary nature and limited reach of an absolutist political regime" that "imposed high transaction costs on market exchanges across a vast empire" (Ma 2013). Ma's remark warns against the temptation to limit the analysis to superficial similarities between Asian monetary systems and the experiences of medieval and early modern Europe. In fact recent research on China and East Asia exploits a comparative perspective with Europe to highlight the economic and institutional peculiarities that shaped Asian monetary systems. While the "big problem of small change" (Sargent and Velde 2003) was common to both areas, the facts that seasonal shortages caused low-denomination coins to appreciate in terms of silver taels (for instance, after the harvest season) and low-denomination currencies showed a secular appreciating trend with respect to silver in the eighteenth century may reflect the higher degree of monetization of Asian rural markets and the weak integration of money markets for low and high denominations, which created a structural complementarity between "concurrent but not-integrable currency circuits" (Kuroda 2008b). Currency circuits also reflected differences in market structures and cultures, which led to cross-regional variations in monetary practices and limited the geographical penetration of the silver dollar as unified fiduciary monetary standard (von Glahn 2007).

These traits seem to be deeply rooted in Chinese history. **Yohei Kakinuma** (▶ Chap. 21, "Monetary System in Ancient China") provides a critical discussion of Asian scholarship on the development of a monetary economy in feudal China from the Shang dynasty (around 1600 BC) to the end of the Southern dynasties around 600 AD. The traditional view (the "Rise and Fall Theory") identifies a peak in the use of metallic forms of money during the Western Han dynasty (206 BC–24 AD) – a period of political unification and commercial expansion – followed by a progressive decline characterized by chronic shortages. More recent perspectives emphasize how the monetary functions performed by gold, bronze coins, and standardized commodities (bolts of textiles, grain) generated a system of multiple (concurrent) currencies with different economic and social functions: for instance, under the Jin dynasty (265 AD–420 AD), textiles were used in transactions with the government (taxes and tributes), while market transactions were mainly settled in coins. **Niv Horesh** (▶ Chap. 22, "The Monetary System of China Under the Qing Dynasty") illustrates not only the key differences between the bimetallic regime of the Qing era (1644–1911 AD) and European bimetallism but also the governance principles that informed monetary management in China, i.e., the regulation of coin production across such a vast empire as a means of stabilizing grain prices and the political rationale for a limited use of paper money. He also highlights how research on Chinese monetary history contributes to wider historiographical debates, such as the "Great Divergence" between Western Europe and Asia, or the role of debasement and seigniorage as sources of revenues in the territorial expansion of imperial rule (which connects with the Chinese revisionist literature known as the "New Qing History").

Finally, **Hisashi Takagi** (▶ Chap. 23, "The Monetary System of Japan in the Tokugawa Period") discusses the results of Japanese research on the "triple standard system" (gold coins, silver currency by weight, and copper coins, each one circulating in different spheres and on the base of its own denominational units) that prevailed under the last military regime (shogunate) of feudal Japan (1603–1867) (Ohkura and Shimbo 1978). The shogunate aimed to affirm its exclusive coinage prerogatives and establish a homogenous national system based on official exchange rates. However, currencies issued by the central authority were exchanged on the base of floating market rates and were supplemented by local currencies (including paper notes) issued by feudal lords as well as private paper notes issued by merchants. The system experienced frequent recoinages and alternated periodic inflationary debasements and deflationary increases in metal content. By the early nineteenth century, a large share of the circulation was in token silver coins denominated in gold units (*ryō*). However, in the absence of a free bullion market, the domestic gold-silver ratio deviated significantly from the international ratio. When the Harris Treaty opened Japanese ports to Western trade in 1858, arbitrage led to massive outflows of gold and inflows of silver which forced a debasement of gold coinage to align it to the international ratio. Thus, the monetary development of the late Tokugawa period paved the way to Japan's transition to a Gold Standard under the Meiji period at the end of the nineteenth century.

Exchange Rate Regimes

An international monetary order is defined by the set of rules and practices by which national governments and financial intermediaries provide international money to settle cross-country trade and financial transactions (McKinnon 1993). In this perspective, the choice of the exchange rate regime is a key policy decision that contributes to shape a country's monetary regime. The chapters included in this part survey and discuss research on the historical evolution of exchange rate regimes in the nineteenth and twentieth centuries (Eichengreen 1996 and 2009; Bordo and Eichengreen 1993; Bordo and Schwartz 1984). An influential analytical framework is the so-called trilemma or impossible trinity (a policy trade-off between exchange rate stability, open capital markets, and monetary autonomy) based on the Mundell-Fleming model (Obstfeld et al. 2005). The classification of exchange rate regimes on the base of "de jure" or "de facto" criteria (Ghosh et al. 2003; Reinhart and Rogoff 2004; Rose 2011) and "trilemma configurations" that quantify governments' achievements along the three dimensions (Aizenman et al. 2010, 2013) are related issues. A second perspective is the "nominal anchor" or "credibility" approach – a trade-off between the monetary stability (and easier access to international capital markets) generated by pre-committing to policy rules, such as under a Gold Standard, and the flexibility allowed by discretional policies – inspired by the Barro-Gordon model with rational expectations and time inconsistency (Giovannini 1992; Bordo 2003; Bordo and Kydland 1995; Bordo and Rockoff 1996). Models of currency crises ("speculative attacks") of first (inconsistent fundamentals), second (self-fulfilling expectations), or third generation (sudden stops, balance-sheet effects) and their interaction with banking and sovereign debt crises ("twin" and "triple" crises) are a third strand of literature with a strong impact on historical studies. A comparative historical analysis of the incidence of currency and "multiple" crises highlights a contrast between relative stability under the Gold Standard and Bretton Woods and the pronounced instability of the 1920s to 1930s and the post-1973 period (Bordo and Schwartz 2000; Bordo et al. 2001). Emerging economies were especially vulnerable to currency crises in the past (as they were at the end of the twentieth century) due to high foreign currency debt (a consequence of "original sin," i.e., the inability to issue debt in their domestic currency), poor policy credibility, and sudden reversals in international capital flows (Bordo and Flandreau 2003; Bordo and Meissner 2007; Bordo et al. 2010; Catao 2007). The interaction between a run of the currency and a run on banks in the context of a sudden stop of capital inflows was also at the heart of the German crisis of 1931, whose spillovers contributed to the sterling crisis and brought an end to the interwar Gold Exchange Standard (Accominotti 2012; Accominotti and Eichengreen 2016; Ferguson and Temin 2004; Schnabel 2004a, b, 2009).

Lawrence Officer (▶ Chap. 24, "International Monetary Regimes: The Gold Standard") reviews critically the results of the last generation of research on the Gold Standard. This was an international monetary regime under which countries with domestic circulation mainly based on gold coins and gold-backed fiduciary money allowed free trade of gold and capital flows. Exchange rates were fixed at the mint

parity and could fluctuate within a narrow band of "gold points," determined by the cost of shipping gold. Officer explains the criteria used to identify countries on an "effective" (i.e., operational) Gold Standard and to locate them in a center-core-periphery international hierarchy which reflects each country's relative importance in the establishment and maintenance of the standard. Empirical studies suggest that the post-1870 "scramble to gold" (the widespread adoption of Gold Standard by an increasing number of countries) was driven by network externalities in international trade and a search for exchange rate stability and easy access to international capital markets. Recent research also suggests that adjustment of external disequilibria under the Gold Standard was driven by global monetary mechanisms quite different from the classical price specie-flow model. In a similar fashion, empirical research challenges the traditional view of central banks "playing by rules" that strongly constrained their monetary independence, showing how discount rates and central bank assets were managed to stabilize domestic monetary conditions. The Gold Standard's stability rested on the absolute credibility of private sector commitment to fixed gold parities and an ideology of monetary and fiscal orthodoxy shared by policymakers and a wide coalition of economic interests. However, the notion that the Gold Standard championed monetary stability compared to later regimes holds for the mean of inflation, but not for its volatility; and the presence of strong deflationary forces suggests that there might have been a trade-off between stability and growth.

Olivier Accominotti (▶ Chap. 25, "International Monetary Regimes: The Interwar Gold Exchange Standard") shows how the macroeconomic and political circumstances of the pre-1914 period changed dramatically after World War I. The interwar Gold Exchange Standard (in which central banks held most of their reserves in short-term assets denominated in gold-convertible international currencies) was intrinsically vulnerable. Any sizable liquidation of international reserves (mainly British pounds and US dollars) into gold tended to weaken key international currencies and contracted liquidity in the London and New York money markets, which in turn increased pressure on the international monetary system. Recent research inspired by currency crisis models explored this mechanism in the "sudden stop" of 1929–1931 that followed the foreign capital boom-bust cycle of 1924–1928. The interwar regime also lacked the credibility of its predecessor, since the postwar mass mobilization exposed monetary and fiscal policies to unprecedented political pressure. In this low-credibility environment, investors were extremely sensitive to policies that might signal a weakening of governments' commitment to the recently re-established gold parities. As a response, policymakers became reluctant to engage in expansionary policies in response to economic recessions with rising unemployment. This political economy dynamic imparted on the monetary regime a strong "deflationary bias," which reached its peak in the perverse policy choices that led to the Great Depression. In this perspective, however, recent research emphasizes that, beyond economic policy mistakes, the Depression was the result of a complex interaction between governments, central bankers, and investors' expectations, in an international environment not conducive to cooperative behavior.

Peter Kugler and **Tobias Straumann** (▶ Chap. 26, "International Monetary Regimes: The Bretton Woods System") illustrate how the new rules of the game designed at the Bretton Woods conference envisaged a system (a pegged-but-adjustable exchange rate supported by the liquidity provided by the IMF) that would generate exchange rate stability without the destructive deflationary adjustment of the interwar period. On retrospective analysis, however, the system operated on the base of three structural and interrelated weaknesses: a heavy reliance on the US dollar and the Federal Reserve as sources of international liquidity, which loosened the external constraints on the US balance of payments; governments' reluctance to use demand-management and devaluation to reduce external imbalances, which delayed adjustment and exacerbated pressure on international liquidity; and the escalation of US external liabilities compared to its gold reserves, which undermined the stability of the anchor currency. Kugler and Strauman suggest that the evolution of the system into a pure dollar standard resembled the dynamics of a bimetallic system in which large and persistent deviations of the relative market price from the legal ratio undermine the monetary use of the legally undervalued metal – a version of Gresham's Law. As the real dollar price of gold in the gold market declined in the inflationary 1960s, dollars replaced gold in central banks' international reserves. Empirical research based on time-series methods supports this interpretation.

Finally **Atish Ghosh**, **Anne-Marie Gulde,** and **Holger Wolf** (▶ Chap. 27, "Currency Boards") analyze the historical record of currency boards, a fixed exchange rate regime widely used under different historical circumstance to achieve monetary stability. Ghosh, Gulde, and Wolf analyze the characteristics and implications of soft and hard variants of currency boards and discuss recent research on three cases: colonial boards in the British Empire of the late nineteenth and early twentieth centuries, Argentina dollar board, and the Euro boards adopted by Baltic and Eastern European countries in the 1990s. Studies on the colonial experiences agree that currency boards achieved their primary goals of monetary stability and trade facilitation but disagree on their possible costs in terms of sterilization of resources and constraints on financial development, reserve diversification, seigniorage revenues, monetary flexibility, and growth. More recent cases in emerging economies suggest that currency boards' credibility as an anti-inflationary arrangement depends on the fact that it makes external adjustment practically impossible and an exit strategy extremely costly both politically and economically. The lesson is that boards are sustainable only in political economies that generate a wide societal consensus on redistributional policies constrained by fiscal discipline.

Monetary Integration

Monetary integration can be measured in terms of convergence of short-term nominal interest rates in money markets across regions or countries. In the "trilemma," nominal convergence is achieved by combining fixed exchange rates with free capital movements, which reduce the scope for national monetary authorities to determine domestic interest rates and inflation autonomously. An extension is the

optimum currency area (OCA) theory, an analytical framework originally built on the base of contributions by Robert Mundel, Peter Kenen, and Ronald McKinnon in the 1960s and subsequently revised and expanded. The theory is based on the key notion that countries with high level of economic integration – measured by bilateral trade and business cycle synchronization – can reap large gains in monetary efficiency (i.e., reduction in transaction costs) by forming a monetary union with a single currency and a single monetary policy. For sufficiently integrated countries, these benefits largely outweigh the costs implied by the loss of the exchange rate as an instrument of macroeconomic stabilization. Fiscal transfers that accommodate the impact of asymmetric shocks improve the sustainability of a monetary union in imperfect OCAs. A variant of the theory is its "credibility" version, in which the inclusion in a monetary union of countries with a credible record of good macroeconomic management and high-quality institutions provides especially high benefits to member countries with endemic macroeconomic instability. An extension, proposed by Jeffrey Frankel and Andres Rose, suggests that the very creation of a monetary union may foster trade, financial integration, and symmetric shocks – i.e., optimum currency areas are endogenous (Tavlas 1993; Dellas and Tavlas 2009). Since the 1990s the process of monetary integration that eventually led to the establishment of the Economic and Monetary Union in Europe led to the emergence of a large body of historical research that uses, challenges, and enriches the OCA perspective. While studies of national monetary unifications in Europe (Italy, Germany) are limited, international currency unions of the nineteenth century (the Latin Monetary Union, the Scandinavian Monetary Union) drew considerable attention. In the Italian case, research suggests that monetary unification failed to overcome domestic financial market segmentation (Toniolo et al. 2003) and to produce a reconfiguration of trade flows and economic structure in line with OCAs criteria (Foreman-Peck 2005). Several studies explored the political and economic factors that shaped the experience of the Latin Monetary Union in the context of transition from bimetallism to the Gold Standard (Einaudi 1997, 2000a, b, 2001; Flandreau 1995a, 2000, 2003). Although the Union qualified as an OCA in terms of trade and financial integration and had a high degree of convergence in coinage standards (Flandreau 1995b; Redish 2000), its functioning was rapidly disrupted by arbitrage forces triggered by the instability of the bimetallic standard and undisciplined coinage and overissue by free-riding members (especially Italy). Empirical research finds that the Union lacked credibility and did not enhance nominal convergence among member countries (Bae and Bailey 2011), while its positive impact on trade was marginal and limited both in scope and time (Flandreau and Maurel 2005; Timini 2018). Scholars agree that the most important lesson from past experiences of monetary unions is the critical importance of fiscal integration for their sustainability (Bordo and Jonung 1999; Bordo et al. 2013; De Cecco 1996; James 1997). In this perspective, the USA provides an especially interesting historical example. Early experiments with a common currency circulating side by side with state currencies failed, and the US federal political economy complicated the transition to a stable monetary union with a uniform currency and robust safeguards against financial instability during the nineteenth century. Ultimately, the structural problem of

asymmetric regional shocks was gradually addressed by institutional reforms, such as the establishment of the Federal Reserve in 1914, the expansion of federal fiscal transfers, and the establishment of bank deposit insurance in the 1930s (Bordo 2004; Grubb 2003; 2006; Rockoff 2000; Rousseau 2006; 2015; Michener and Wright 2006; Selgin 2007; Sylla 2006, 2014).

David Weiman and **John James** (▶ Chap. 28, "The Evolution of the Modern US Monetary and Payments System") illustrate the key elements of the US national monetary-payments union from the 1790s to the 1920s: a common unit of account (the dollar), a common currency (national and then Federal Reserve banknotes), and an integrated banking system with a common national bank deposit money. The US monetary-payments system experienced periods during which a central federal authority realized the criteria of a "more perfect" monetary union (until 1836 with the Bank of the United States, after 1914 with the Federal Reserve) and periods in which states' prerogatives on banking and money prevailed. Weiman and James show how a continuously expanding hierarchical network of interconnected banks created positive externalities that produced an efficient payment system. The logic of network economics was explicitly recognized in the structural design of the National Banking and the Federal Reserve System. Although a central monetary authority was not a necessary precondition for the development of an efficient monetary-payment system based on bank liabilities, it became a necessary institutional change to guarantee the value of bank deposits and the execution of payments during banking panics.

Anders Ögren (▶ Chap. 29, "Currency Unions") critically discusses the basic tenets and inconsistencies of the optimum currency area theory and analyzes the most important historical cases of national monetary unifications (the USA, Germany) and international monetary unions (the Latin Monetary Union and the Scandinavian Monetary Union). Research unanimously suggests that successful monetary integration requires a high degree of political and fiscal integration which go hand in hand. In fact national monetary systems work as politically integrated currency unions supplemented by fiscal transfers from surplus to deficit regions, both in centralized and federal polities. Transfers are especially necessary in shocks that hit asymmetrically sectors and regions. This lesson holds also for "multinational" currency unions, EMU included, if they aim at being sustainable. Examples of less politically integrated unions can be found in history, but they were more akin to fixed exchange rate regimes managed by independent central banks.

Emmanuel Mourlon-Druol (▶ Chap. 32, "European Monetary Integration") deals directly with the EMU. His chapter analyzes the key controversies that articulated the long economic and political debate over plans for European monetary integration since the 1960s. One is the confrontation between "monetarists" (according to whom monetary integration would lead to macroeconomic convergence) and "economists" (monetary integration should "coronate" a process of convergence). Another thread deals with different approaches to monetary integration. These included a European Unit of Account (a basket of national currencies); a common (parallel) currency circulating alongside national currencies; the exchange rate mechanisms under the "Snake" (1973–1979) and the European Monetary

System (1979–1997); and the two single-currency grand schemes devised in the Werner Report (1970) and the Delors Report (1989). As Mourlon-Druol emphasizes, the post-2008 eurozone crisis shed light on some of the weaknesses correctly identified but left unaddressed in the architecture of the EMU as finally materialized in the Maastricht Treaty.

► Chapters 31, "Currency Blocs: The Yen" by **Michael Schiltz** and ► 30, "The Sterling Area 1945–1972" by **Catherine Schenk** deal with another type of monetary integration schemes: currency blocs, i.e., international arrangements that promote the exclusive international status of a currency of reference within group of countries tied to a core country by strong political and economic connections. Schiltz characterizes the yen bloc as a projection of imperial aspirations in a geopolitical environment (East Asia) which was fundamentally adverse to Japan's political and economic influence. The case of Korea (under Japanese rule since 1905) is illustrative of a pattern of currency imperialism which, through monetary reforms and exchange rate choices, was explicitly designed to promote the adoption by colonies of a de facto yen standard in support of Japan's financial domination. This had been worked out in Taiwan and would be replicated with minor variants in Manchuria and other occupied territories in the 1930s. After 1941, the "Greater East-Asian Financial Sphere" turned into a hyperinflationary experiment that flooded captive economies with inconvertible yens in support of the Japanese war machine. Although the sterling area found its origins in World War II, it was a loose international arrangement under which former and actual British colonies pegged their exchange rate to the sterling, used almost exclusively sterlings as central bank's reserves, and adopted common exchange and capital controls to limit the conversion of sterling assets into dollars. In return, they enjoyed preferential treatment when trading with Britain and issuing debt in the London capital market. The scholarly debate revolves around various burdens that the sterling area allegedly imposed on the British economy: external fragility, excess capital exports, high interest rates, and a misguided initial hesitancy to join in the process of the European economic integration. From a different angle, the sterling area was regarded as a constraint on monetary management and economic diversification of peripheral member countries. As Schenks's critical review of the literature shows, research suggests that most of these claims were fallacious. Overall, sterling area's trade and monetary arrangements contributed to macroeconomic stability of member countries while allowing them to pursue national economic interests, including central bank reserves' diversification. By the late 1960s, a declining sterling area was reduced to a set of bilateral agreements to manage in a coordinated way central bank reserves' diversification, in the wider context of multilateral cooperative efforts to prop up exchange rate pegs in the crisis of the Bretton Woods system.

Central Banking and Monetary Policy

The history of central banking and monetary policy is a research area in which the integration of money and finance is stronger. This reflects principally the fact that in the historical evolution of central banks, the pursuits of monetary and financial

stability – that is, the protection of the payments system and the role of lender of last resort to the banking system in periods of acute stress (later on complemented with regulation and supervision) – were twin responsibilities that evolved side by side and often posed complex trade-offs to central bankers (Bordo and Siklos 2018; Capie 1995; Capie et al. 1994; Goodhart 1988, 2011; Toniolo 1988; Toniolo and White 2015). Macroeconomic models that emphasize the key role of the banking sector as a channel of transmission of monetary policy through lending and balance-sheet effects and the impact of monetary policy on asset prices also contributed to move historical research in a similar direction. Recent macrofinancial studies explored the role of monetary policy in driving credit and asset boom-bust cycles and their ability to predict financial crises over the long run (Schularick and Taylor 2012; Jorda et al. 2011, 2013, 2015; Meissner 2013). Some historical episodes have emerged as privileged ground for empirical analysis, for instance, the channel of transmission of monetary shocks and the role of non-monetary shocks in the Great Depression (Bernanke 1983; Hamilton 1987, 1992; Cecchetti 1992; Romer and Romer 2013) or the interaction between monetary shocks, interest rates, and stock prices in nineteenth-century banking panics (Canova 1991; Rousseau 2002; Davis et al. 2009; Hanes and Rhode 2013). In line with the "nominal anchor" or "credibility" approach discussed in Part "Exchange Rate Regimes," research on monetary policy has increasingly focused on "monetary policy regimes," which jointly with the exchange rate determine a country's monetary regime. Monetary policy regimes can be defined as sets of rules for the formulation of monetary policy (usually differentiated in terms of "targets": gold price, the exchange rate, monetary aggregates, inflation) that generate stable expectations between policymakers and the public (Eichengreen 1991; Mishkin 1999). In international comparative perspective, scholars have explored the performance of alternative monetary policy regimes in terms of macroeconomic stability (Bordo and Schwartz 1999; Rolnick and Weber 1997) and the long-run evolution of central banks' credibility and reputation in delivering price stability (Bordo and Siklos 2014, 2015). International comparison of monetary policy regimes has been enhanced by the increasing availability of historical monographs with detailed insights of monetary institutions and policymaking in the USA (Meltzer 2003, 2010; Hetzel 2008), the UK (Capie 2010; Needham 2014), France (Patat and Lutfalla 1990; Monnet 2018), Germany (Bundesbank 1999), and Italy (Fratianni and Spinelli 1997). In addition, long-run perspectives on the evolution of national central banks are available for most Western European countries, the USA, Japan, and China (Edvinsson et al. 2018). Research on central banks' cooperation from the Gold Standard to Bretton Woods also produced a vast literature, often based on primary archival sources of national central banks and multilateral organizations such as the IMF and the BIS (Bordo and Schenk 2016; Cooper 2006; Eichengreen 1984; Flandreau 1997; James 1996; Mouré 2002; Toniolo 2005).

Many of these issues are reflected in the chapters of this part. **Stefano Ugolini** (▶ Chap. 33, "The Historical Evolution of Central Banking") provides a secular perspective that covers a variety of polities with different institutional and political economy characteristics, from medieval city-states and centralized states of the modern period to decentralized territorial polities, such as the USA and the

European Union. For that purpose, Ugolini differentiates conceptually between central banks, which are a relatively recent phenomenon, and central banking as a set of public policies aiming at enhancing monetary and financial stability. These include the management of the payment system (a natural monopoly), the establishment of prudential lending standards, the provision of lending of last resort, the monetization of public debt, and the management of money creation. These functions, which are the prerogatives of modern central banks, were performed by a variety of institutions over the course of history, such as public banks and banks of issue. Ugolini's analysis suggests that the historical development of central banking did not follow an evolutionary process driven by the principle of "survival of the fittest" or by the adoption of superior organizational solutions. It was rather the result of continuous experimentation and adaptation in response to domestic political economies.

The other chapters in this part deal with the historical evolution of monetary policy in the nineteenth and twentieth centuries – the periods that saw the emergence of modern central banks – under different exchange rate regimes, from the Gold Standard to the recent Great Recession. They reflect how different theoretical frameworks (Fisher-Friedman vs Keynes), by embedding different ideas about the characteristics of the economy (inherently stable or unstable), define the policy goals that central banks are expected to achieve and can be held responsible for – a non-activist regime based on an optimal monetary rule vs a more activist regime trying to exploit the trade-off of the Phillips curve. They also illustrate how the modern conception of monetary policy – the setting of a policy instrument to influence the behavior of the economy – only emerged after World War II together with modern macroeconomics.

Duncan Needham (▶ Chap. 34, "The Evolution of Monetary Policy (Goals and Targets) in Western Europe") provides a critical analysis of the very large literature on the three largest European economies: the UK, France, and Germany. Beyond similarities and differences in institutional frameworks, monetary policy design, and execution, he emphasizes the discontinuity represented by World War I and the long-term influence of each country's interwar experience (hyperinflation for Germany, unemployment for Britain) on the postwar preferences (price stability vs employment) of their central banks. **Robert Hetzel** (▶ Chap. 35, "The Evolution of US Monetary Policy") offers an analytical narrative based on the continuous interaction between the monetary rules followed by Fed policymakers, the nature and properties of the monetary regime in force, and the consequences of Fed's policy. The 1920s provide an example. The establishment of the Federal Reserve in 1914 aimed at fixing the problem of financial instability (speculative credit booms and asset bubbles followed by panics and deflation) that had shaken periodically the US economy in the nineteenth century. However, Hetzel argues the real bills doctrine (a variant of the quantity theory) that inspired its policy rule and the peculiar characteristics of the US Gold Standard (which operated somehow as a de facto fiat standard) prevented the Fed from fully understanding the impact of its monetary regime on prices and, as a consequence, misguided it into the contractionary policy that exacerbated financial instability in the first phase of the Great Depression. This

was just the first episode of a never-ending debate over the evolution of the monetary regime and the optimal monetary policy rule that would deliver economic stability, which Hetzel reconstructs until the recent Great Recession.

Masato Shizume (▶ Chap. 36, "The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy") outlines the long-run fluctuations in the Bank of Japan's institutional framework, policy objectives, and operative instruments from the late nineteenth century to the recent period. Established in the 1880s with a mandate to mobilize resources in support of Japan's industrialization, in peacetime the size of its balance sheet with respect to GDP remained stable. This fact, Shizume suggests, reflects the stability of demand for central bank money and the accommodative stance of the Bank's policy. Massive debt monetization to support war finance during World Word II and the adoption of unconventional quantitative easing in 2001 (in response to the deflationary pressures that prevailed after the banking crisis of the 1990s) represent the only significant exceptions to this long-run equilibrium.

Finally **Esteban Pérez Caldentey** and **Matias Vernengo** (▶ Chap. 37, "The Historical Evolution of Monetary Policy in Latin America") outline the main phases of monetary policymaking in the recent history of Latin America. This started in the 1920s when Latin American countries joined the Gold Exchange Standard and the first central banks were established under the guidance and supervision of foreign "money doctors" (US and European experts of central banking). In the 1930s and 1940s, the adoption by central banks of the goals of the state-led inward-looking model of economic development (Taylor 1998) marked a major departure from the original design, whose impact lasted until the 1970s. The economic and financial reforms of the 1980s were a third turning point. Since then the mandates and practices of Latin American central banks have converged toward those of industrial countries, by setting price stability as their overarching objective and later on by adopting monetary policy rules based on inflation targeting and complemented by fiscal rules. Pérez Caldentey and Vernengo discuss whether these rules are optimal for a region with peculiar economic characteristics (i.e., its strong dependence on international trade and capital flows and its exposure to boom-bust cycles) and to what extent the subordination of fiscal policy and its reduction to social welfare spending reduces the ability of Latin American policymakers to achieve macroeconomic stabilization.

Aggregate Price Shocks

In the secular swing between success and failure that characterized the quest for monetary stability and credibility since the late nineteenth century, World War I and II represented extreme circumstances that turned central banks into "engines of inflation" (Bordo 2018). The wartime inflation and postwar hyperinflations of the twentieth century just confirmed the association between inflationary finance (i.e., the systematic monetization of large budget deficits), an acceleration of money

growth, and extreme price dynamics already observed in the American revolutionary war (Baak 2006; Grubb 2008, 2012, 2013), the French revolution (Sargent and Velde 1995; White 1995), and the American civil war (Calomiris 1988; Burdekin and Weidenmier 2001). All those episodes occurred – in fact, they became conceivable only under a fiat money monetary regime, in which governments, by suspending the convertibility of paper money into specie, removed any physical constraint on money issue (Capie 1986, 1991). The global transition to a fiat money regime in the 1970s also witnessed an acceleration of money growth that produced the largest and longest peacetime departure from price stability of the twentieth century (Friedman 1986). Overall the historical record suggests that the quantity theory of money can adequately explain price dynamics under purely fiduciary money. Empirical analysis of long time series confirms that the growth rates of monetary aggregates exhibited stronger cross-correlation and correlation with inflation under fiat money after World War II than under the Gold Standard; this correlation was especially strong in coincidence with the inflationary outbursts of WWI and the Great Inflation (less so during WWII, possibly due to price controls) (Rolnick and Weber 1997; Fischer et al. 2002; Benati 2009). In this sense, historical research vindicates Milton Friedman's claim that "inflation is always and everywhere a monetary phenomenon" (especially high inflation). At the same time, cross-country data spanning the nineteenth and twentieth centuries reveal a long-run equilibrium between monetary aggregates, GDP, and short-term interest rates. Historical analysis therefore suggests that the long-run money demand remained remarkably stable over time. The implication is that monetary aggregates might have been dismissed too early both as components of macroeconomic models and targets in monetary policymaking (Benati et al. 2016).

Recent research on hyperinflationary episodes shows a wide consensus on Sargent's approach focused on fiscal regimes and expectations (Sargent 1982). The dynamics of public debt (and, in some cases, also private debt and nominal wages) and fiscal news were the key drivers (either directly or indirectly) of rising inflationary expectations and monetary expansion. Expectations reversed and inflationary pressures quickly subsided as soon as governments committed credibly to a drastic change in fiscal and monetary policy regime, supported by legal and institutional reforms that limited governments' access to central bank borrowing (Siklos 1990; Burdekin and Burkett 1992; Ferguson 1996). A recent qualification suggests that high economic uncertainty, by preventing governments from committing to credible fiscal policies, might have contributed to hyperinflationary dynamics (Lopez and Mitchener 2018).

On the contrary, the scholarly debate on the Great Inflation of the 1970s is largely unsettled. Beyond the old (but still lively) controversy between monetarists (blaming overly expansionary monetary policies) and supporters of a "supply shock" interpretation, the monetarist camp has divided on why monetary authorities lost control of price dynamics (Bordo and Orphanides 2013). While some blame central banks' attempt to ride the Phillips curve to reach full employment, others contend that they accommodated private expectations of rising inflation because pursuing disinflation would have been too costly in terms of output and employment losses. Another

hypothesis is that central banks avoided monetary policy actions because they regarded inflation as a cost-push phenomenon, while others suggest that they were misled by an excessive focus on output gaps and unemployment, or by misperceptions of the natural rate of unemployment, or by poor real-time data. Whatever its determinants, the monetary policy failure of the 1970s translated into a highly persistent inflation, which some economists suggest, should be "hardwired" as a structural element into macroeconomic models. Historical research, on the contrary, demonstrates that inflation persistence greatly varied across monetary regimes and was almost absent in stable regimes with clearly defined anchors, such as the Gold Standard and, more recently, inflation targeting (Benati 2005, 2008).

The termination of high and extreme inflations and macroeconomic stabilization inevitably implies negative money supply shocks that generated disinflation or outright deflation. While the latter are usually associated with recessions and unemployment, historical research shows that their costs are affected by policy credibility. The historical experience of the USA suggests that a gradual approach to disinflation (such as in the period after the civil war) causes less harm to the real economy than a sharp monetary contraction (e.g., after World War I) if the monetary regime has a credible long-run anchor and the policy is predictable, as under a Gold Standard. In contrast, only an aggressive policy with high costs (as in the disinflation of the early 1980s) is likely to achieve its objectives in a monetary regime with poor credibility and predictability (Bordo et al. 2007). In general, negative money shocks which are non-neutral for a significant period tend to produce "bad" deflations with significant contractions in aggregate demand. However, this effect can be offset by positive supply shocks that reflects increases in productivity. Research shows that this was actually the case during the pre-1914 period, during which the secular decline in prices reflected mainly "good" deflation (Bordo et al. 2004; Bordo and Redish 2004). In turn, the Great Depression was characterized by the international transmission of deflationary pressures that led to several cases of "ugly" deflation with very large output and employment losses. In the case of the USA, deflationary shocks were also regularly associated with banking and financial crises until the 1930s (Bordo et al. 2002). However, the Great Depression is a very special case not representative of the overall historical experience, in which an association between asset price deflations and positive growth seems to prevail (Bordo and Filardo 2005a and 2005b; Borio and Filardo 2004; Borio et al. 2015).

The chapters of this part echo many of these issues. The institution of banknotes' convertibility into high-value coins at a predetermined fixed rate aimed at managing the overissue problem by constraining the supply of paper money. Under exceptional circumstances, however, such as a state of war, governments could temporarily suspend convertibility; this was the case in Sweden in the early 1740s and in England and Ireland in the late 1790s. Contemporaries observed that, in a monetary standard based on inconvertible paper, the price of specie tended to increase in terms of the domestic currency leading to a depreciation of the exchange rate. A controversy followed between those who interpreted this as evidence of price inflation driven by excess circulation of banknotes ("bullionist") and those who pointed to real factors as an alternative explanation. **Joshua Hendrickson**

(▶ Chap. 38, "Bullionism") explains how in modern historical research the bullionist controversy led to a debate about the determinants of the price level under different monetary regimes, raising the critical questions of whether the quantity theory is the appropriate analytical framework for the relationship between money and prices only under conditions of inconvertibility. Empirical tests based on time-series methods have provided mixed results, but the most recent studies point to a strong inflationary impact of shocks in the supply of banknotes in the cases of England and Sweden, which supports the bullionist view.

In ▶ Chap. 39, "Money in Wars" by **Kenneth Moure** deals with the monetary dimensions of the two "total wars" of the twentieth century. His comparative analysis emphasizes the importance of state tax and borrowing capacity – a long-run process dating back to the early modern period – to achieve a successful economic mobilization and reduce the use of the money press for war financing, thus limiting postwar monetary overhang and inflation. However, the unprecedented scale of the resources required and the characteristics of the wars (for instance, military occupations) disrupted monetary regimes and paved the ground for significant monetary changes. During World War I, the suspension of gold convertibility and the withdrawal (or hoarding) of metal coins installed a fiat money regime, with a circulation largely based on banknotes of small denomination; acute money shortages were relaxed by the issue of "necessity money," currency bonds, and tokens by local authorities and private organizations. In World War II, the failed policies of the previous war provided governments with a blueprint for a more successful management of war finance. In the USA and the UK, governments elaborated rudimentary national accounts and macroeconomic models to keep inflation under control. On the contrary, war finance and monetary management in totalitarian countries resorted less to planning and more to coercion. In occupied territories, Germany introduced an extractive system of forced credits, clearing accounts and administered exchange rates to pay for "occupation costs" which fed the German war machine at the cost of local treasuries, decreased inflationary pressure in Germany, and left occupied countries sitting on an inflationary time bomb. On the contrary, allied occupation authorities managed monetary emissions in liberated territories in coordination with local governments. In many countries, overcoming the legacy of wartime disruption required comprehensive monetary reforms to restore circulation and eradicate inflationary pressures.

Pierre Siklos and **Martin Bohl** (▶ Chap. 40, "The Anatomy of Inflation: An Economic History Perspective") review the theoretical and empirical literature on the determinants of inflation, its relationship with relative price variability, and its welfare costs. They also compare the record of price inflation under different monetary regimes and discuss the role of government credibility in anchoring inflationary expectations. An important conclusion is that inflation, if fully anticipated, produces modest social costs. The consensus view is that low and stable inflation is desirable as it comes closest to reducing the loss of purchasing power of money and (at least in theory) is easier to forecast, thus reducing the probability of large and persistent forecast errors. Economists however have failed to identify an "optimal" inflation rate and to provide a conclusive account of the dynamics of

inflation. The last part of their chapter is devoted to extraordinary spells of inflation – i.e., hyperinflations. As Siklos and Bohl show, empirical research confirms Cagan's hypothesis of a stable money demand function even under extreme monetary stress; in turn, the presence of self-generating inflation based on rational bubbles is excluded. As a consequence, they can be stopped by credible stabilization programs that affect fundamentals.

In the last chapter, **Richard Burdekin** (▶ Chap. 41, "Deflations in History") emphasizes the profound differences between historical and recent episodes of deflations. In the past, deflations were frequent in the second half of the nineteenth century and widespread during the Great Depression. These episodes were regularly associated to monetary contractions, mostly associated to the operation of the Gold Standard, although "good deflation" (falling prices driven by positive supply-side shocks) might have played a role. On the contrary, recent deflationary pressures, which emerged in Japan and China after the turn of the twenty-first century and in the USA and the Euro Area after the 2007–2008 crisis, occurred in a context of expansionary policies and a pronounced fall in the money multiplier – an apparent paradox which raised a lively debate. Burdekin also reviews critically the main negative effects that deflations can have on real economic activity through a reduction in aggregate demand (Mundell-Tobin effect), debt deflation (Fisher), or balance-sheet effects (Bernanke). Finally he discusses price-level targeting (an approach followed by the Swedish central bank in the 1930s; Straumann and Woitek 2009) as a possible monetary policy rule (alternative to inflation targeting) to anchor expectations and achieve price stability when conventional anti-deflationary policies fail.

References

Abramitzky R (2015) Economics and the modern economic historian. J Econ Hist 75(4):1240–1251

Accominotti O (2012) London merchant banks, the central European panic, and the sterling crisis of 1931. J Econ Hist 72(1):1–43

Accominotti O, Eichengreen B (2016) The mother of all sudden stops: capital flows and reversals in Europe, 1919-32. Econ Hist Rev 69(2):469–492

Aizenman J, Chinn MD, Ito H (2010) The emerging global financial architecture: tracing and evaluating the new patterns of the trilemma's configurations. J Int Money Financ 29(4):615–641

Aizenman J, Chinn MD, Ito H (2013) The 'impossible trinity' hypothesis in an era of global imbalances: measurement and testing. Rev Int Econ 21(3):447–458

Allen R, Bassino J-P, Ma D, Moll-Durata C, Van Zanden L (2011) Wages, prices and living standards in China 1738–1925: in comparison with Europe. Jpn India Econ Hist Rev 64:8–38

Allen M (2012) Mints and money in medieval England. Cambridge University Press, Cambridge

Baak B (2006) America's first monetary policy: inflation and seigniorage during the revolutionary war. Financ Hist Rev 15(2):107–121

Bae K-H, Bailey W (2011) The Latin monetary union: some evidence on Europe's failed common currency. Rev Dev Financ 1:131–149

Benati L (2005) The inflation-targeting framework from an historical perspective. Bank England Q Bull 45(2):160–168

Benati L (2008) Investigating inflation persistence across monetary regimes. Q J Econ 123(3):1005–1060

Benati L (2009) Long run evidence on money growth and inflation. European Central Bank Working Paper n. 1027, Frankfurt A.M.
Benati L, Lucas R, Nicolini JP, Weber W (2016) International evidence on long run money demand. NBER Working Paper n. 22475
Bernanke BS (1983) Nonmonetary effects of the financial crisis in the propagation of the great depression. Am Econ Rev 73(3):257–276
Boerner L, Volckart O (2011) The utility of a common coinage: currency unions and the integration of money markets in late medieval Central Europe. Explor Econ Hist 48:53–65
Bordo MD (1986) Money deflation and seigniorage in the 15th century. J Monet Econ 18(11):337–346
Bordo MD (2003) Exchange rate regime choice in historical perspective. IMF Working Papers, n. 160, Washington DC
Bordo MD (2004) The United States as a monetary union and the euro: a historical perspective. Cato J 24(1/2):163–170
Bordo MD (2018) A historical perspective on the quest for financial stability and the monetary policy regime. J Econ Hist 78(2):319–357
Bordo MD, Dueker MJ, Wheelock DC (2002) Aggregate price shocks and financial instability: a historical analysis. Econ Inq 40(4):521–538
Bordo MD, Eichengreen B (eds) (1993) A retrospective on the Bretton woods system. University of Chicago Press, Chicago/London
Bordo MD, Eichengreen B, Klingebiel D, Martinez-Peria MS, Rose AK (2001) Is the crisis problem growing more severe? Econ Policy 16(32):51–82
Bordo MD, Erceg C, Levin C, Michaels R (2007) Three great American disinflations. NBER Working Paper n. 12982
Bordo MD, Filardo A (2005a) Deflation in a historical perspective. BIS Working Paper n. 186, Basel
Bordo M, Filardo A (2005b) Deflation and monetary policy in a historical perspective: remembering the past or being condemned to repeat it? Econ Policy 20(44):799–844
Bordo MD, Flandreau M (2003) Core, periphery, exchange rate regimes, and globalization. In: Bordo MD, Taylor AM, Williamson JG (eds) Globalization in historical perspective. University of Chicago Press, Chicago/London, pp 417–472
Bordo MD, Jonung L (1999) The future of EMU: what does the history of monetary unions tell us? NBER Working Paper n. 7365
Bordo MD, Jonung L, Markievicz A (2013) A fiscal union for the euro: some lessons from history. CESifo Econ Stud 59(3):449–488
Bordo MD, Kydland FE (1995) The gold standard as a rule. Explor Econ Hist 32:423–464
Bordo MD, Landon Lane J, Redish A (2004) Good versus bad deflation: lessons from the Gold Standard era. NBER Working Paper n. 10329
Bordo MD, Meissner C (2007) Financial crises, 1880–1913: the role of foreign currency debt. In: Edwards S, Esquivel G, Márquez G (eds) The decline of Latin American economies: growth, institutions, and crises. University of Chicago Press, Chicago, pp 139–194
Bordo MD, Meissner CM, Stuckler D (2010) Foreign currency debt, financial crises and economic growth: a long run view. J Int Money Financ 29(4):642–665
Bordo MD, Orphanides A (2013) Introduction. In: Bordo MD, Orphanides A (eds) The great inflation. The rebirth of modern central banking. The University of Chicago Press, Chicago/London, pp 1–22
Bordo M, Redish A (2004) Is deflation depressing: evidence from the classical gold standard. In: Burdekin R, Siklos P (eds) Deflation: current and historical perspectives. Cambridge University Press, Cambridge
Bordo MD, Rockoff H (1996) The gold standard as a "good housekeeping seal of approval". J Econ Hist 56(2):389–428
Bordo MD, Schenk CR (2016) Monetary policy cooperation and coordination: an historical perspective on the importance of rules. In: Bordo MD, Taylor J (eds) Rules for international stability: past, present and future. Hoover Institution Press, Stanford, pp 205–262

Bordo MD, Schwartz A (eds) (1984) A retrospective on the classical gold standard 1821–1931. University of Chicago Press, Chicago/London
Bordo MD, Schwartz A (1999) Monetary policy regimes and economic performance: the historical record. In: Taylor JB, Woodford M (eds) Handbook of macroeconomics. North-Holland, Amsterdam, pp 149–234
Bordo MD, Schwartz A (2000) Measuring real economic effects of bailouts: historical perspectives on how countries in financial distress have fared with and without bailouts. Carn-Roch Conf Ser Public Policy 53(1):81–167
Bordo MD, Siklos P (2014) Central bank credibility, reputation and inflation targeting in historical perspective. NBER Working Paper n. 20693
Bordo MD, Siklos P (2015) Central bank credibility: an historical and quantitative exploration. NBER Working Paper n. 20824
Bordo MD, Siklos P (2018) Central banks: evolution and innovation in historical perspective. In: Edvinsson R, Jacobson T, Waldenström D (eds) Sveriges Riksbank and the history of central banking. Cambridge University Press, Cambridge, pp 26–89
Borio C, Filardo A (2004) Looking back at the international deflation record. N Am J Econ Financ 15:287–311
Borio C, Erdem M, Filardo A, Hofmann B (2015) The costs of deflations: a historical perspective. BIS Q Rev 15(3):31–54
Boyer-Xambeau MT, Gillard L, Deleplace G (1994) Private money and public currencies: the sixteenth century challenge. Taylor and Francis, London
Broadberry S, Gupta B (2006) The early modern great divergence: wages, prices and economic development in Europe and Asia, 1500–1800. Econ Hist Rev 59(1):2–31
Bundesbank (1999) Fifty years of the deutsche mark. Central bank and the currency in Germany since. Oxford University Press, Oxford, p 1948
Burdekin RCK, Burkett P (1992) Money, credit and wages in hyperinflation: post-world war I Germany. Econ Inq 30(3):479–495
Burdekin RCK, Weidenmier MD (2001) Inflation is always and everywhere a monetary phenomenon: Richmond vs. Houston in 1864. Am Econ Rev 91(5):1621–1630
Calomiris CW (1988) Price and exchange rate determination during the greenback suspension. Oxf Econ Pap 40(4):719–750
Canova F (1991) The sources of financial crises: pre- and post-fed evidence. Int Econ Rev 32(3):689–713
Capie F (1986) Conditions in which very rapid inflation has appeared. Carn-Roch Conf Ser Public Policy 24:115–168
Capie F (1991) Introduction. In: Capie F (ed) Major inflations in history. Edward Elgar, Cheltenham, pp 1–10
Capie F (1995) The evolution of central banking. World Bank Policy Research Working Paper n. 1534, Washington DC
Capie F (2010) The Bank of England 1950s to 1979. Cambridge University Press, Cambridge
Capie F, Goodhart CAE, Schnadt N (1994) The development of central banking. In: Capie F, CAE G, Schnadt N, Fischer S (eds) The future of central banking. Cambridge University Press, Cambridge, pp 1–261
Catao L (2007) Sudden stops and currency drops: a historical look. In: Edwards S, Esquivel G, Márquez G (eds) The decline of Latin American economies: growth, institutions, and crises. University of Chicago Press, Chicago, pp 243–289
Cecchetti SG (1992) Prices during the great depression: was the deflation of 1930–1932 really unanticipated? Am Econ Rev 82(1):141–156
Celia J, Grubb F (2016) Non-legal-tender paper money: the structure and performance of Maryland's bills of credit, 1767–75. Economic History Review 69(4):1132–1156
Champ B, Freeman S, Haslag J (2018) Modelling monetary economies. Cambridge University Press, Cambridge
Chilosi D, Volckart O (2010) Good or bad money? Debasement, society, and the state in the late Middle Ages. Working Paper 140/10, London School of Economic and Political Science, Economic History Working Paper n. 140, London

Chilosi D, Volckart O (2011) Money, states, and empire: financial integration and institutional change in Central Europe, 1400–1520. J Econ Hist 71(3):762–791
Chitu L, Eichengreen B, Mehl A (2014) When did the dollar overtake sterling as the leading international currency? Evidence from the bond markets. J Dev Econ 111(3):225–245
Cooper RN (2006) Almost a century of central bank cooperation. BIS Working Paper n. 198, Basle
Davis JH, Hanes C, Rhode PW (2009) Harvests and business cycles in nineteenth-century America. Q J Econ 124(4):1675–1727
De Cecco M (1996) The European monetary union: lessons of historical experience. PSL Q Rev 46:55–68
Dellas H, Tavlas GS (2009) An optimum-currency-area odyssey. J Int Money Financ 28(7):1117–1137
Desan C (2014) Making money: coin, currency, and the coming of capitalism. Oxford University Press, Oxford
de Vries J (2006) The industrious revolution: consumer behavior and the household economy, 1650 to present. Cambridge University Press, Cambridge
Diebolt C, Haupert M (2016) Handbook of cliometrics. Springer, Berlin/Heidelberg
Earle T (1991) Chiefdoms: power, economy and ideology. Cambridge University Press, Cambridge
Earle T (2002) Bronze age economics. Westwiev Press, Boulder
Edvinsson R, Jacobson T, Waldenström D (eds) (2018) Sveriges Riksbank and the history of central banking. Cambridge University Press, Cambridge
Eichengreen B (1984) Central bank cooperation under the interwar gold standard. Explor Econ Hist 21(1):64–87
Eichengreen B (1991) Introduction. In: Eichengreen B (ed) Monetary regime transformation. Edward Elgar, Cheltenham
Eichengreen B (1996) Golden fetters: the gold standard and the great depression 1919–1939. Oxford University Press, Oxford
Eichengreen B (2009) Globalizing capital. A history of the international monetary system. Princeton University Press, Princeton
Eichengreen B (2010) Exorbitant privilege. The rise and fall of the dollar and the future of the international monetary system. Oxford University Press, Oxford
Eichengreen B (2011) The new monetary and financial history. In: Wood G et al (eds) Monetary and banking history. Essays in honour of Forrest Capie. Routledge, London, pp 27–48
Eichengreen B, Flandreau M (2009) The rise and fall of the dollar (or when did the dollar replace sterling as the leading reserve currency?). Eur Rev Econ Hist 13(3):377–411
Eichengreen B, Flandreau M (2010) The Federal Reserve, the Bank of England and the rise of the dollar as an international currency 1914–39. BIS Working Papers, No. 328, Basel
Eichengreen B, Mehl A, Chitu L (2018) How global currencies work: past, present, and future. Princeton University Press, Princeton
Einaudi L (1997) Monetary unions and free riders, the case of the Latin monetary union (1865-78). Rivista di Storia Economica 3:327–362
Einaudi L (2000a) From the franc to the 'Europe': the attempted transformation of the Latin monetary union into a European monetary union, 1865–1873. Econ Hist Rev 53(2):284–308
Einaudi L (2000b) 'The generous utopia of yesterday can become the practical achievement of tomorrow': 1000 years of monetary union in Europe. Natl Inst Econ Rev 172:90–104
Einaudi L (2001) Money and politics: European monetary unification and the international gold standard (1865–1873). Oxford University Press, Oxford
Epstein S (2001) The late medieval crisis as an 'integration crisis'. In: Prak M (ed) Early modern capitalism: economic and social change in Europe 1400–1800. Routledge, London, pp 25–50
Ferguson N (1996) Constraints and room for manoeuvre in the German inflation of the early 1920. Econ Hist Rev 49(4):635–666
Ferguson T, Temin P (2004) Comment on 'the German twin crisis of 1931'. J Econ Hist 64(3):872–876
Fischer S, Sahay R, Vegh C (2002) Modern hyper- and high inflations. J Econ Lit 40(3):837–880
Flandreau M (1995a) L'or du monde. la France et la stabilité du système monétaire international 1848–1873. L'Harmattan, Paris

Flandreau M (1995b) Was the Latin monetary union a franc zone? In: Reis J (ed) International systems in historical perspective. McMillan, London, pp 71–89
Flandreau M (1997) Central bank cooperation in historical perspective: a sceptical view. Econ Hist Rev 50(4):735–763
Flandreau M (2000) The economics and politics of monetary unions: a reassessment of the Latin monetary union, 1865–71. Financ Hist Rev 7:25–43
Flandreau M (2003) The glitter of gold. France, bimetallism and the emergence of the International Gold Standard 1848–1873. Oxford University Press, Oxford
Flandreau M, Jobst C (2005) The ties that divide: a network analysis of the international monetary system, 1890–1910. J Econ Hist 65(4):977–1007
Flandreau M, Jobst C (2009) The empirics of international currencies: network externalities, history and persistence. Econ J 119(537):643–664
Flandreau M, Maurel M (2005) Monetary union, trade integration, and business cycles in 19th century Europe. Open Econ Rev 16:135–152
Flandreau M, Ugolini S (2013) Where it all began: lending of last resort and Bank of England monitoring during the Overend-gurney panic of 1866. In: Bordo MD, Roberds W (eds) The origins, history and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, New York, pp 113–161
Flynn D (2013) Precious metals and moneys. In: Caprio G et al (eds) Handbook of key global financial markets, institutions, and infrastructure. Elsevier, London/San Francisco, pp 1200–1800
Flynn DO, Giraldez A (1995a) Born with a 'silver spoon'. The origin of world trade in 1571. J World Hist 6(2):201–221
Flynn DO, Giraldez A (1995b) Arbitrage, China and world trade in the early modern period. J Econ Soc Hist Orient 38(4):429–448
Flynn DO, Giraldez A (2002) Cycles of silver: global economic unity through the mid-eighteenth century. J World Hist 13(2):391–427
Flynn DO, Giraldez A, von Glahn R (2003) (eds) Global connections and monetary history 1470–1800. Ashgate, Aldershot
Foreman-Peck J (2005) Lessons from Italian monetary unification. Cardiff Economics Working Papers n. E2005/4
Fox D, Ernst W (2016) Money in the Western legal tradition: middle ages to Bretton woods. Oxford University Press, Oxford
Fratianni M, Spinelli F (1997) A monetary history of Italy. Cambridge University Press, Cambridge
Friedman M (1986) Monetary policy in a fiat world. Contemp Policy Issues 4(1):1–9
Giovannini A (1992) Bretton woods and its precursors: rules versus discretion in the history of international monetary regimes. In: Bordo MD, Eichengreen B (eds) A retrospective on the Bretton woods system. University of Chicago Press, Chicago/London, pp 109–154
Goldstone J (1984) Urbanization and inflation: lessons from the English Price revolution of the sixteenth and seventeenth centuries. Am J Sociol 89:1122–1160
Goodhart CAE (1988) The evolution of central banks. MIT Press, Cambridge
Goodhart CAE (1989) Money, information and uncertainty. Macmillan, London
Goodhart CAE (1998) The two concepts of money: implications for the analysis of optimal currency areas. Eur J Polit Econ 14(1998):407–432
Goodhart CAE (2011) The changing role of central banks. Financ Hist Rev 18(2):135–154
Ghosh AR, Gulde AM, Wolf HC (2003) Exchange rate regimes: choices and consequences. MIT Press, Cambridge
Grubb F (2003) Creating the U.S. dollar currency union, 1748–1811: a quest for monetary stability or a usurpation of state sovereignty for personal gain? Am Econ Rev 93(5):1778–1798
Grubb F (2006) The U.S. constitution and monetary powers: an analysis of the 1787 constitutional convention and the constitutional transformation of the U.S. monetary system. Financ Hist Rev 13(1):43–71
Grubb F (2008) The continental Dollar: how much was really issued? J Econ Hist 68(1):283–291

Grubb F (2012) State redemption of the continental Dollar, 1779-90. William Mary Q 69(1):147–180
Grubb F (2013) The Continental Dollar: how the American Revolution was financed with paper money-Initial design and ideal performance. NBER Working Paper n. 19577
Grubb F (2016a) Is paper money just paper money? Experimentation and variation in the paper monies issued by the American colonies from 1690 to 1775. Res Econ Hist 32:147–224
Grubb F (2016b) Colonial New Jersey paper money, 1709–1775: value decomposition and performance. J Econ Hist 76(4):1216–1232
Hamilton JD (1987) Monetary factors in the great depression. J Monet Econ 19(2):145–169
Hamilton JD (1992) Was the deflation during the great depression anticipated? Evidence from the commodity futures market. Am Econ Rev 82(1):157–178
Hanes C, Rhode PW (2013) Harvests and financial crises in gold standard America. J Econ Hist 73(1):201–246
Helleiner E (2003) The making of national money. Territorial currencies in historical perspective. Cornell University Press, Ithaca
Hetzel R (2008) The monetary policy of the Federal Reserve. A history. Cambridge University Press, Cambridge
James H (1996) International monetary cooperation under Bretton Woods. Oxford University Press, Oxford/New York; International Monetary Fund, Washington DC
James H (1997) Monetary and fiscal unification in nineteenth-century Germany: what can Kohl learn from Bismarck? Essays in International Finance n. 202, Department of Economics, Princeton University
Jaremski M (2019) Today's economic history and tomorrow's scholars. Cliometrica. https://doi.org/10.1007/s11698-019-00188-9
Jordà O, Schularick M, Taylor AM (2011) Financial crises, credit booms, and external imbalances: 140 years of lessons. IMF Econ Rev 59(2):340–378
Jordà O, Schularick M, Taylor AM (2013) When credit bites back. J Money Credit Bank 45(s2):3–28
Jordà O, Schularick M, Taylor AM (2015) Leveraged bubbles. J Monet Econ 76(supplement issue): s1–s20
Kahn CM, Roberds W (2007) Transferability, finality and debt settlement. J Monet Econ 54(4):955–978
Karaman KK, Sevket P, Yildirim-Karaman S (forthcoming) Money and monetary stability in Europe 1300–1914. J Monet Econ. https://doi.org/10.1016/j.jmoneco.2019.07.007
Kuroda A (2008a) What is the complementarity among monies? An introductory note. Financ Hist Rev 15(1):7–11
Kleer RA (2008) Fictitious cash: English public finance and paper money, 1689–97. In McGrath CI, Fauske CJ (eds) Money, power and print: interdisciplinary studies on the financial revolution in the British Isles. University of Delaware Press, Newark, pp. 70–103
Kleer RA (2015) 'A new species of mony': British Exchequer bills, 1701–1711. Financial History Review 22(2): 179–203
Kleer RA (2017) Money, politics, and power. Banking and public finance in wartime England 1694–1696. Routledge, London/New York
Kuroda A (2008b) Concurrent but non-integrable currency circuits: complementary relationships among monies in modern China and other regions. Financ Hist Rev 15(1):17–36
Le Rider G (2001) La naissance de la monnaie. In: Pratíques monétaires de l'Orient ancien. Presses Universitaires de France, Paris
Lindert P (1985) English population, wages, and prices: 1541–1913. J Interdiscip Hist 15:609–634
Lopez JA, Mitchener KJ (2018) Uncertainty and hyperinflation: European inflation dynamics after World War I. NBER Working Paper, 24624
Ma D (2013) Chinese money and monetary system 1800–2000. Overview. In: Caprio G et al (eds) Handbook of key global financial markets, institutions, and infrastructure. Elsevier, London/San Francisco, pp 57–64
Margo R (2018) The integration of economic history into economics. Cliometrica 12:377–406

Mayhew N (1995) Population, money supply, and the velocity of circulation in England, 1300–1700. Econ Hist Rev 48:238–257
Mayhew N (2013) Prices in England 1170–1750. Past Present 219:3–39
McCallum BT (1992) Money and price in colonial America: a new test of competing theories. J Polit Econ 100(1):143–161
McKinnon RI (1993) The rules of the game: international money in historical perspective. J Econ Lit 31:1–44
Meissner CM (2013) Capital flows, credit booms, and financial crises in the classical Gold Standard era. NBER Working Paper n. 18814
Meltzer AH (2003) A history of the Federal Reserve vol 1 1913–1951. The University of Chicago Press, Chicago
Meltzer AH (2010) A history of the Federal Reserve vol 2 book 1 1951–1969. The University of Chicago Press, Chicago
Michener R (1987) Fixed exchange rates and the quantity theory in colonial America. Carn-Roch Conf Ser Public Policy 27:233–307
Michener R (1988) Backing theories and the currencies of eighteenth-century America: a comment. J Econ Hist 48(3):682–692
Michener R (2015) Redemption theories and the value of American colonial paper money. Financ Hist Rev 22(3):315–335
Michener R, Wright R (2005) State 'currencies' and the transition to the U.S. dollar: clarifying some confusions. Am Econ Rev 95(3):682–703
Michener R, Wright R (2006) Development of the US monetary union. Financ Hist Rev 13(1):19–41
Mishkin FS (1999) International experiences with different monetary policy regimes. J Monet Econ 43:579–605
Miskimin HA (1984) Money and power in fifteenth-century France. Yale University Press, New Haven/London
Monnet E (2018) Controlling credit. Central banking and the planned economy in postwar France 1948–1973. Cambridge University Press, Cambridge
Mouré K (2002) The gold standard illusion: France, the Bank of France, and the international gold standard, 1914–1939. Oxford University Press, Oxford
Muldrew C (2001) 'Hard food for Midas': cash and its social value in early modern England. Past Present 170:78–120
Muldrew C (2007) Wages and the problem of monetary scarcity in early modern England. In: Lucassen J (ed) Wages and currency. Peter Lang, New York, p 393
Muldrew C, King S (2003) Cash, wages and the economy of makeshifts in England, 1650–1800. In: Scholliers P, Schwarz L (eds) Experiencing wages: social and cultural aspects of wage forms in Europe since 1500. Berghahn Books, New York, pp 155–180
Munro J (1991) The central European mining boom, mint outputs, and prices in the Low Countries and England, 1450 – 1550. In: Van Cauwenberghe EHG (ed) Money, coins and commerce: essays in the monetary history of Asia and Europe (from antiquity to modern times). Leuven University Press, Leuven, pp 119–183
Munro J (1992) Bullion flows and monetary policies in England and the Low Countries, 1350–1500. Ashgate, Aldershot
Munro J (2003) The monetary origins of the 'Price Revolution': South German silver mining, merchant-banking, and Venetian commerce, 1470–1540. In: Flynn D, Giráldez A, von Glahn R (eds) Global connections and monetary history 1470–1800 (ed) D Flynn, A. Giráldez and R. von Glahn. Aldershot and Brookfield, Ashgate
Munro J (2008) Price revolution. In: Durlauf SN, Blume LE (eds) The new Palgrave dictionary of economics, vol 6. Palgrave Macmillan, London, pp 632–634
Munro J (2016) The technology and economics of coinage debasements in medieval and early modern Europe: with special references to the Low Countries and England. In: Munro J (ed) Money in the pre-industrial world. Routledge, London, pp 15–32

Neal L (2000) How it all began: the monetary and financial architecture of Europe during the first global capital markets, 1648–1815. Financ Hist Rev 7:117–140
Neal L (2012) "I am not master of events": the speculations of John law and Lord Londonderry in the Mississippi and South Sea bubbles. Yale University Press, New Haven
Needham D (2014) UK monetary policy from devaluation to Thatcher 1967–1982. Palgrave Macmillan, London
Nicolini E, Ramos F (2010) A new method for estimating the money demand in pre-industrial economies: probate inventories and Spain in the eighteenth century. Eur Rev Econ Hist 14:145–177
North D (1994) Government and the cost of exchange in history. J Econ Hist 44:255–264
Obstfeld M, Shambaugh JC, Taylor AM (2005) The trilemma in history: tradeoffs among exchange rates, monetary policies, and capital mobility. Rev Econ Stat 87(3):423–438
Ohkura T, Shimbo H (1978) The Tokugawa monetary policy in the eighteenth and nineteenth centuries. Explor Econ Hist 15:101–124
Palma N (2018) Reconstruction of money supply over the long run: the case of England, 1270–1870. Econ Hist Rev 71(2):373–392
Patat JP, Lutfalla M (1990) Monetary history or France in the twentieth century. Palgrave Macmillan, London
Pomerantz K (2000) The great divergence: China, Europe and the making of the modern world economy. Princeton University Press, Princeton
Prados de la Escosura L, Álvarez Nogal C (2013) The rise and fall of Spain, 1270–1850. Econ Hist Rev 66(1):1–37
Quinn S (1997) Goldsmith-banking: mutual acceptance and interbanker clearing in restoration London. Explor Econ Hist 34:411–432
Quinn S, Roberds W (2003) Are on-line currencies virtual banknotes? Fed Reserv Bank Atl Econ Rev 88(2):1–15
Quinn S, Roberds W (2008) The evolution of the check as a means of payment: a historical survey. Fed Reserv Bank Atl Econ Rev 93(4):1–28
Quinn S, Roberds W (2014) How Amsterdam got fiat money. J Monet Econ 66:1–12
Redish A (2000) Bimetallism. An economic and historical analysis. Cambridge University Press, Cambridge
Rheinart C, Rogoff K (2004) The modern history of exchange rate arrangements: a reinterpretation. Q J Econ 119(1):1–48
Robson E (1999) Mesopotamian mathematics 2100–1600 BC. Technical constants in bureaucracy and education. Oxford University Press, Oxford
Robson E (2007) Mesopotamian mathematics. In: Katz VJ (ed) The mathematics of Egypt, Mesopotamia, China, India and Islam. A sourcebook. Princeton University Press, Princeton, pp 58–185
Rockoff H (2000) How long did it take the United States to become an optimum currency area? NBER Historical Working Paper n. 124
Rolnick AJ, Weber WE (1986) Gresham's law or Gresham's fallacy? Fed Reserv Bank Minneapol Q Rev 10(1):17–24
Rolnick AJ, Weber WE (1997) Inflation and output under fiat and commodity standards. J Polit Econ 105(6):1308–1321
Rolnick AJ, Velde FR, Weber WE (1996) The debasement puzzle: an essay on medieval monetary history. J Econ Hist 56(4):789–808
Romer CD, Romer DH (2013) The missing transmission mechanism in the monetary explanation of the great depression. Am Econ Rev 103(3):66–72
Rose AK (2011) Exchange rate regimes in the modern era: fixed, floating, and flaky. J Econ Lit 49(3):652–672
Rousseau PL (2002) Jacksonian monetary policy, specie flows and the panic of 1837. J Econ Hist 62(2):457–488
Rousseau PL (2006) A common currency: early US monetary policy and the transition to the dollar. Financ Hist Rev 13:97–122

Rousseau P (2007) Backing, the quantity theory, and the transition to the US dollar, 1723–1850. Am Econ Rev 97(2):266–270
Rousseau P (2015) Politics on the road of US monetary union. In: Humpage OF (ed) Current Federal Reserve policy under the lens of economic history. Essays to commemorate the Federal Reserve System's centennial. Cambridge University Press, Cambridge, pp 151–173
Sargent TJ (1982) The ends of four big inflations. In: Hall RE (ed) Inflation: causes and effects. University of Chicago Press, Chicago/London, pp 41–98
Sargent TJ, Smith BD (1997) Coinage, debasements, and Gresham's law. Econ Theory 10:197–226
Sargent TJ, Velde FR (1995) Macroeconomic features of the French revolution. J Polit Econ 103(3):474–518
Sargent TJ, Velde FR (1999) The big problem of small change. J Money Credit Bank 31(2):137–161
Sargent TJ, Velde FR (2003) The big problem of small change. Princeton University Press, Princeton
Schnabel I (2004a) The German twin crisis of 1931. J Econ Hist 64(3):822–871
Schnabel I (2004b) Reply to Thomas Ferguson and Peter Temin's "comment on 'the German twin crisis of 1931'". J Econ Hist 64(3):877–878
Schnabel I (2009) The role of liquidity and implicit guarantees in the German twin crisis of 1931. JInt Money Financ 28(1):1–25
Schularick M, Taylor AM (2012) Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. Am Econ Rev 102(2):1029–1061
Selgin G (1996) Salvaging Gresham's law: the good, the bad, and the illegal. J Money Credit Bank 28(4):637–649
Selgin G (2007) The suppression of state banknotes: a reconsideration. Econ Inq 38(4):600–615
Selgin G (2008) Good money: Birmingham button makers, the Royal Mint, and the beginnings of modern coinage, 1775–1821: private enterprise and popular coinage. University of Michigan Press, Ann Arbor
Siklos P (1990) Hyperinflations: their origins, development and termination. J Econ Surv 4(3):225–248
Smith BD (1985) Some colonial evidence on two theories of money: Maryland and the Carolinas. J Polit Econ 93(6):1178–1211
Smith BD (1988) The relationship between money and prices: some historical evidence reconsidered. Fed Reserv Bank Minneapol Q Rev 12(Summer):18–32
Spufford P (1991) Money and its use in medieval Europe. Cambridge University Press, Cambridge
Straumann T, Woitek U (2009) A pioneer of a new monetary policy? Sweden's price-level targeting of the 1930s revisited. Eur Rev Econ Hist 13:251–282
Sussman N (1993) Debasements, royal revenues, and inflation in France during the hundred years' war, 1415–1422. J Econ Hist 53(1):44–70
Sussman N (1998) The late medieval bullion famine reconsidered. J Econ Hist 58(1):126–154
Sussman N, Zeira J (2003) Commodity money inflation: theory and evidence from France in 1350–1436. J Monet Econ 50(8):1769–1793
Sylla R (1982) Monetary innovations in America. J Econ Hist 42(1):21–30
Sylla R (2006) The transition to a monetary union in the US. Financial. Hist Rev 13(1):73–95
Sylla R (2014) Early US struggles with fiscal federalism: lessons for Europe? Comp Econ Stud 56(2):157–175
Tavlas GS (1993) The 'new' theory of optimum currency areas. World Econ 16(6):663–685
Taylor AM (1998) On the costs of inward-looking development: price distortions, growth, and divergence in Latin America. J Econ Hist 58(1):1–28
Timini J (2018) Currency unions and heterogeneous trade effects: the case of the Latin monetary union. Eur Rev Econ Hist 22(3):322–348
Toniolo G (ed) (1988) Central banks' independence in historical perspective. De Gruyter, Berlin/New York
Toniolo G (2005) Central bank cooperation at the Bank for International Settlements, 1930–1973. Cambridge University Press, Cambridge

Toniolo G, Conte L, Vecchi G (2003) Monetary unification, institutions and financial market integration: Italy, 1862–1905. Explor Econ Hist 40:443–461
Toniolo G, White E (2015) The evolution of the financial stability mandate: from its origins to the present day. NBER Working Paper n. 20844
Van de Mieroop (2004) The ancient Mesopotamian city. Oxford University Press, Oxford
Velde FR (2007) John Law's system. Am Econ Rev 97(2):276–279
Velde FR (2008) Government equity and money: John Laws system in 1720 France. Princeton University Press, Princeton
Velde FR, Weber WE, Wright R (1999) A model of commodity money, with applications to Gresham's law and the debasement puzzle. Rev Econ Dyn 2:291–323
Volckart O (2018) Technologies of money in the middle ages: the "principles of minting". London School of Economics and Political Science, Economic History Working Paper n. 275
von Glahn R (1996) Fountain of fortune. Money and monetary policy in China, 1000–1700. University of 1784 California Press, Oakland
von Glahn R (2007) Foreign silver coins in the market culture of nineteenth century China. Int J Asian Stud 4(1):51–78
von Glahn R (2013) Chinese finance 1348–1700. In: Caprio G et al (eds) Handbook of key global financial markets, institutions, and infrastructure. Elsevier, London/San Francisco, pp 47–56
von Glahn (2016) The economic history of China: from antiquity to the nineteenth century. Cambridge University Press, Cambridge
Walsh C (2010) Monetary theory and policy. MIT Press, Cambridge
Wehrheim L (2019) Economic history goes digital: topic modeling the journal of economic history. Cliometrica 13:83–125
Whaples R (2002) The Supply and Demand of Economic History: Recent Trends in the Journal of Economic History. J Econ Hist 62(2):524–532
White E (1995) The French revolution and the politics of government finance 1770–1815. J Econ Hist 55:227–255
Wray LR (2012) Introduction. An overview of heterodox approaches to money and financial institutions. In: Wray LR (ed) Theories of money and banking. Development of heterodox approaches to money and banking, vol 1. Edward Elgar, Cheltenham, pp 1–15
Wray LR (2014) From the State Theory of Money to Modern Money Theory: an alternative to economic orthodoxy. Levy Economics Institute of Bard College, Working Paper n. 792

Part I

Historical Origins of Money

Origins of Money and Interest: Palatial Credit, Not Barter

2

Michael Hudson

Contents

Abstract

Neolithic and Bronze Age economies operated mainly on credit. Because of the time gap between planting and harvesting, few payments were made at the time of purchase. When Babylonians went to the local alehouse, they did not pay by carrying grain around in their pockets. They ran up a tab to be settled at harvest time on the threshing floor. The ale women who ran these "pubs" would then pay most of this grain to the palace for consignments advanced to them during the crop year. These payments were financial in character, not on-the-spot barter-type exchange. As a means of payment, the early use of monetized grain and silver was mainly to settle such debts. This monetization was not physical; it was administrative and fiscal. The paradigmatic payments involved the palace

M. Hudson (✉)
University of Missouri-Kansas City, Kansas City, MO, USA

Levy Economics Institute of Bard College, New York, NY, USA

Institute for Long-term Economic Trends, Peking University, Beijing, China
e-mail: hudson.ISLET@gmail.com

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_1

or temples, which regulated the weights, measures, and purity standards necessary for money to be accepted. Their accountants developed money as an administrative tool for forward planning and resource allocation, and for transactions with the rest of the economy to collect land rent and assign values to trade consignments, which were paid in silver at the end of each seafaring or caravan cycle.

Keywords
Palatial economies · Debt · Grain and silver monetization

Neolithic and Bronze Age economies operated mainly on credit. Because of the time gap between planting and harvesting, few payments were made at the time of purchase. When Babylonians went to the local alehouse, they did not pay by carrying grain around in their pockets. They ran up a tab to be settled at harvest time on the threshing floor. The ale women who ran these "pubs" would then pay most of this grain to the palace for consignments advanced to them during the crop year. These payments were financial in character, not on-the-spot barter-type exchange.

As a means of payment, the early use of monetized grain and silver was mainly to settle such debts. This monetization was not physical; it was administrative and fiscal. The paradigmatic payments involved the palace or temples, which regulated the weights, measures, and purity standards necessary for money to be accepted. Their accountants developed money as an administrative tool for forward planning and resource allocation, and for transactions with the rest of the economy to collect land rent and assign values to consignments, which were paid in silver at the end of each seafaring or caravan cycle.

Money's Role in the Palatial Economies of Mesopotamia and Egypt

The origins of monetary debts and means of payment are grounded in the accounting practices innovated by Sumerian temples and palaces c. 3000 BC to manage a primarily agrarian economy that required foreign trade to obtain metal, stone, and other materials not domestically available. These large institutions employed staffs of weavers and other craft personnel, who were fed by crops grown either on palace or temple land or that of sharecroppers paying grain-rent or fees to these institutions and supplied with wool from temple and palace herds managed by entrepreneurs or owned outside of these institutions.

Building public infrastructure required feeding and supplying corvée labor and craftsmen with food, tools, and beer, as well as provisioning celebratory festivals. In order to calculate budgets for forward planning and tally surpluses or shortfalls, these flows had to be measured and accounts presented to the palace for managing

cropland and herds, brewing and selling beer, baking bread, and producing handicrafts for use within these institutions and for local or long-distance trade.

Textiles and other products were consigned to traders to obtain silver, copper, and other raw materials, while land and professional functions or enterprises were consigned to entrepreneurs to manage in exchange for a stipulated revenue, typically calculated in advance as a flat fee based on normal experience. This administrative system is described by Renger (1979, 1984), Bongenaar (2000), and Garfinkle (2004, 2012). The papers in Hudson and Wunsch (2004) survey account-keeping and monetization of the Mesopotamian and Egyptian economies from the inception of written accounts in the late fourth millennium BC to the Neo-Babylonian period.

The scale on which the large institutions operated required forward planning to schedule and track the flow of food and raw materials through their fields and workshops. The first need was to assign standardized values to key commodities. This problem was solved by creating a grid of administered prices, set in round numbers for ease of computation and account-keeping. Grain was designated as a *unit of account* to calculate values and co-measure labor time and land yields for resource allocation involving the agricultural and handicraft sphere, as well as the means of payment.

The second need of these large institutions was to organize *means of payment* for taxes and fees to their officials, and for financing trade ventures. Silver served as the money-of-account and also as the means of payment for trade and mercantile enterprise. A bimonetary system was created for paying the palace and temples and for valuing disparate commodities and functions, by setting the shekel-weight of silver (8 grams) as equal to a *gur* "quart" of grain or 300 sila. Acceptability of grain and silver for settling official debt balances catalyzed their usage as money throughout the economy.

None of this is comensurability is found in "primitive" money. Philip Grierson (1977: 19ff., endorsed by Goodhart 1998), for instance, seeks the origins of what has come to be called "state money" in wergild payments for personal injury. Along with dowries or bride-price, such fines were denominated in customary market baskets that might include animals or slave girls, items of clothing and jewelry, not a particular commodity. They thus were pre-monetary and special-purpose.

Likewise the "spit money" and other food money cited by Laum (1924: 27ff., 158f.) was pre-monetary in character, not a common denominator to value disparate commodities. Although Laum was a follower of Knapp's State Theory of money, he saw the archaic state as a religious cult not playing a commercial or financial role. In his view, the valuation of goods finds its origins in the selection of sacrificial animals "in the cult, not in commerce (which knows no basis for typology, but remains purely individualistic)." Donations of animals defined status, but did not provide a standardized economic and monetary unit. Fines and contributions were levied without reference to a standardized commodity whose price was set by palaces or temples as the basis for account-keeping, commercial exchange, and credit. That innovation occurred in Mesopotamia early in the third millennium BC.

Accounting and the Origins of Money

In contrast to the grain and silver that served as twin measures of value to evaluate Mesopotamian production and distribution, no monetary common denominators are found in Mycenaean Linear B accounting c. 1400–1200 BC. Tribute lists and deliveries from agricultural centers and workshops were "in kind," with no indication of money as either a measure of value or a general means of payment. Finley (1981: 198) cites Ventris and Chadwick (1956: 113) to the effect that "they 'have not been able to identify any payment in silver or gold for services rendered,' and that there is no evidence 'of anything approaching currency. Every commodity is listed separately, and there is never any sign of equivalence between one unit and another.'"

Van De Mieroop (2004: 49) cites the challenge to ancient accountants: to record not merely "a single transfer, but the combination of a multitude of transfers into a summary. When information piles up and is not synthesized, it becomes useless: a good bureaucrat needs to be able to compress data. The summary account requires that the scribe combine information from various records." Mesopotamia's palaces and temples solved this problem by designating grain and silver as reference points to co-measure the wide range of transactions within their own institutions and with the rest of the economy for grain, textiles, beer, boat transport and the performance of ritual services.

Establishing a set of price equivalencies enabled values to be assigned and payments to be made in terms of any commodity listed on such a schedule. Englund (2004: 38) cites major commodities such as copper, wool and sesame oil being assigned values in an overall price grid, mutually convertible with grain, silver or each other: "The concept of value equivalency was a secure element in Babylonian accounting by at least the time of the sales contracts of the ED IIIa (Fara) period, c. 2600 BC."

Cripps (2017) has reviewed of prices for silver, grain and other commodities and found that administrative barley: silver price ratios among Sumerian cities "vary considerably with both the geographic origin of a text and the administrative context in which these ratios occur, whether or not we understand them as prices or equivalents." However, these variations do not seem to reflect market supply and demand, but are administered. "The value of barley relative to silver arguably varies for quite other reasons than those of abundance or shortage due to natural events, or because of changes in the market and therefore the demand for and supply of one or the other of these commodities." (See Englund 2012 for a general discussion.)

The 1:1 shekel/*gur* ratio enabled monthly and annual income and expense statistics to be expressed in terms of the most basic common denominator, and was used to denominate fees, taxes and other debts owed to the large institutions. But more steps were necessary to fill out the monetary process. To provide a standard of value and serve as the means of payment, grain and silver had to be measured or weighed in standardized units. To facilitate calculation for internal resource allocation within the large institutions, these units were based on the administrative

calendar that temples created in order to allocate resources on a regular monthly basis.

That in turn required replacing lunar months of varying length with standardized 30-day months (Englund 1988). Each monthly unit of grain was measured in volumetric *gur* units divided into 60ths, apparently for consumption as rations to the workforce twice daily during each administrative month. Lambert (1960) describes how Babylonian accounts translated food rations into labor time for each category of labor – males, females and children. This sexagesimal system of fractional divisions enabled the large institutions to calculate the rations needed to produce textiles or bricks, build public structures or dig canals during any given period of time. Weights for silver and other metals followed suit, by dividing the mina into 60 shekels.

This silver and grain money served as the price coefficient by which the temples and palaces valued the products of their work force and the handicrafts they consigned to merchants. The interest rate on commercial advances denominated in silver was set in the simplest sexagesimal way: 1/60th per month, doubling the principal in 5 years (60 months). This standardized rate was adopted by the economy at large.

Money and Prices

By the end of the third millennium the large institutions were stating the value of foreign trade and other palatial enterprise in terms of silver, which emerged as the "money of the world." Gold was used in less public contexts, such as for capital investment in Assyria's foreign trade ventures after 2000 BC. Its price vis-à-vis silver varied from city to city and from period to period. But any attempt to link price changes to variations in the money supply would be anachronistic as far as Bronze Age Mesopotamia is concerned. That is because "money" simply took the form of commodities acceptable for payment to temples and palaces at guaranteed prices. These large institutions gave monetary value to wheat, wool and other key products by accepting them in payment for taxes and fees.

Monetary values had to be stable in order for producers to plan ahead and minimize the risk of disruptive shifts in prices, and hence the ability to pay debts. Official price equivalencies thus served as an adjunct to fiscal policy while avoiding instability. §51 and gap §**t** (sometimes read as §96) of Hammurabi's laws (c. 1750 BC) specified that any citizen who owed barley or silver to a *tamkarum* merchant (including palace collectors) could pay in goods of equivalent value, e.g., in grain, sesame or some other basic commodity on the official price grid (Roth 1997: 91 and 97).

This ruling presumably was important for agricultural entrepreneurs and herd managers who borrowed from the well to do. But most of all, along with §§48–50 of Hammurabi's laws, rulings that stabilized the grain/silver exchange rate "are all meant to give a weak debtor (a small farmer or tenant) some legal protection and help," and are "'given teeth' by stipulating that if [the creditor] takes more

he will forfeit 'everything he gave,' that is, his original claim" (Veenhof 2010: 286f.). Babylonian debtors thus were saved from being harmed at harvest time when payments were due and grain prices were at their seasonal low against silver outside the large institutions. The palace's exchange-rate guarantee enabled cultivators who owed fees, taxes and other debts denominated in silver to pay in barley without having to sell it for silver.

What was called a "silver" debt thus did not mean that actual silver had to be paid, but simply that the interest rate was 20%. If creditors actually wanted silver, they would have to convert their grain at a low market price at harvest time when crops were plentiful. Deliveries to the palace's collectors were stabilized, minimizing the effect of price fluctuations outside of the palatial sector, such as outside the city gates in the quay area along the Euphrates. The effect was much like modern farmers signing "forward" contracts so as not to get whipsawed by shifting market prices.

There was little thought of preventing prices from varying for transactions not involving the large institutions. Prices for grain rose sharply in times of crop failure, droughts or floods, as when Ur was obliged to buy grain from the upstream town of Isin at the end of the Ur III empire c. 2022 BC. But these price shifts were the result of scarcity resulting from natural causes, not a monetary phenomenon.

A monetary drain was avoided in such cases by royal "restorations of order" (Sumerian *amargi*, Babylonian *andurarum*) cancelling agrarian debts when circumstances made them unpayable. To maintain general economic balance in the face of arrears that constantly mounted up, new rulers proclaimed these clean slates upon taking the throne. No money was required from personal debtors (although commercial debts were left in place). The details are spelled out in greatest detail c. 1645 BC in Ammisaduqa's edict §§17f. (translated in Pritchard 1985 [1955]).

Despite variation in market prices for transactions outside of the large institutions, Babylonia's bimonetary standard had no Gresham's Law of "cheap" or "bad" money driving out good money. Grain did not drive out silver. When entrepreneurs in the agricultural sector sought to pay official debts in grain at harvest time, this was part of a structured stable relationship. There was no creation of fiat money by Bronze Age temples and palaces to spend into the economy, and no monetary inflation. Early "money" was simply the official price schedule for paying debts to the large institutions, along lines much like the American "parity pricing" policy to support farm prices after the 1930s. The fact that wool prices, for instance, varied in response to market conditions but nominally remained fixed by royal fiat for 150 years shows that this standardized price referred to debt payments owed to the palace and its collectors.

Rulers promised to promote prosperity by providing consumer goods such as vegetable oil and other commodities at relatively low prices – with what seems to have been an element of idealism. Around 1930 BC, §1 of the laws of Eshnunna (north of Ur, on the Tigris River) was typical in announcing the official rate of 300 silas of barley for 1 shekel of silver, or 3 silas of fine oil, 12 silas of regular oil, 15 silas of lard, 360 shekels (=6 minas) of wool, 2 gur of salt, 600 silas of salt, 3 minas of raw copper or 2 minas of wrought [i.e., refined] copper (Roth 1997: 59).

These low prices were not to be achieved by reducing the money supply. Unlike the grain/silver ratio, such price promises were not monetary rules.

The Two Ideologies of Monetary Origins

The root of the word *numismatics* (the study or collection of coins) is *nomos*, "law" or "custom." Aristotle wrote (*Ethics,* 1133) that money "is called *nomisma* (customary currency), because it does not exist by nature but *nomos*." It is "accepted by agreement (which is why it is called *nomisma*, customary currency." Government priority in supplying money always has been primarily legal and fiscal.

The policy implication of this "state theory" for modern times is that governments do not have to borrow from private banks and bondholders when they run deficits. They can monetize their spending by fiat money – and, as Aristotle added, "it is in our power to change the value of money and make it useless," regardless of its actual use value. (See also Aristotle, *Pol.* 1257b10–12.)

Georg Friedrich Knapp's *State Theory of Money* (1924 [1905]) described how money was given value by its fiscal role: the state's willingness to accept it in payment of taxes. Innes (1913, 1914) added an important dimension by describing the origins of money in paying debts. This linked money to the credit process, not to a commodity as such. (Wray 2004 reviews the historiography of early money and reprints Innes's articles.) Karl Polanyi (1944; Polanyi et al. 1957) led a school emphasizing that "redistributive" economies with administered price equivalencies took precedence over market exchange setting prices by supply and demand (Hudson 2018). Renger (1979, 1984) elaborated the administrative character of Mesopotamia's palatial economies.

These varieties of the State Theory of money (also called Chartalism) downplayed the role of personal and purely commercial gain seeking that had dominated most earlier views of money's origins. Elaborating what Adam Smith described as an instinct among individuals to "truck and barter," Carl Menger (1892 [1871]) put forth the classical version of the Commodity Theory of how money originated Without making any reference to paying taxes or other debts to public authorities, he postulated that money was an outgrowth of barter among individual producers and consumers. According to this view, a preference for metal emerged as the most desirable medium for such trade, thanks to its ability to serve three major functions:

1. A compact and uniform store of value in which to save purchasing power, compressing value (savings) into a relatively small space, and not spoiling (unlike grain).
2. A convenient means of payment, divisible into standardized fractional weight units (assuming that their degree of purity or alloy is attested).
3. A measure of value. Because of the above functions, silver and gold have been widely acceptable commodities against which to measure prices for other products.

The Commodity or Barter Theory depicts money as emerging simply as a commodity preferred by Neolithic producers, traders and wealthy savers when bartering crops and handicrafts amongst themselves. In this origin myth bullion became the measure of value and means of payment without palace or temple oversight, thanks to the fact that individuals could save and lend out at interest. So money doubled as capital – provided by individuals, not public agencies.

Differing views regarding the origins of money have different policy implications. Viewing money as a commodity chosen by individuals for their own use and saving implies that it is natural for banks to mediate money creation. Banking interests favor this scenario of how money might have originated without governments playing any role. The political message is that they – backed by wealthy bondholders and depositors – should have monetary power to decide whether or not to fund governments, whose spending should be financed by borrowing, not by fiat money creation. As a reaction against the nineteenth and early twentieth centuries' rising trend of public regulation and money creation, this school describes money's value as based on its bullion content or convertibility, or on bank deposits and other financial assets.

Governing authorities are missing from this "hard money" view, which its proponents have grounded in an etiological scenario of prehistoric individuals bartering commodities among themselves. The policy implication is that it is irresponsible for governments to create their own money.

Both theories of money's functions, origins source of value are rooted in the debate over whether money should be public or private, and whether it should be backed by public fiat or bullion. Charles Goodhart (1998: 411) shows that the barter or "metallist" theory that money was developed to facilitate individualistic exchange does not even apply to modern times. He provides a bibliography of the long debate to highlight the politically partisan motivations behind today's metallist bias.

Shortcomings of the Barter Theory of Monetary Origins

The long-dominant college textbook by Paul Samuelson (1973: 274f.) summarizes the logic of the Barter Theory taught to generations of economics students:

> Inconvenient as barter obviously is, it represents a great step forward from a state of self-sufficiency in which every man had to be a jack-of-all-trades and master of none. . . . If we were to construct history along hypothetical, logical lines, we should naturally follow the age of barter by the age of commodity money.

Ignoring credit arrangements and excluding any reference to palaces or temples in the Near Eastern inception of monetization, Samuelson then tries to ground this speculation in ostensibly empirical evidence by turning to pseudo-anthropology:

> Historically, a great variety of commodities has served at one time or another as a medium of exchange: . . . tobacco, leather and hides, furs, olive oil, beer or spirits, slaves or wives . . . huge rocks and landmarks, and cigarette butts. The age of commodity money gives way to the age of paper money . . . Finally, along with the age of paper money, there is the age of bank money, or bank checking deposits.

Depicting commodity money as primordial and natural, this view sees the direction of history as culminating in today's commercial banking. It puts credit at the *end* of the Barter-Money-Credit sequence, not at the beginning. By the time Samuelson wrote, the prehistory of money had become an arena in which free-market economists fought with advocates of government regulation over whether the private or public sector should be dominant, and whether governments should oversee credit and create their own money or leave it in private hands. Financial interests applaud the implication that it is natural to leave credit and pro-creditor rules of debt collection to bankers, bondholders and the wealthy, minimizing government "interference."

Neither prehistorians nor anthropologists provide supporting evidence for this Barter Theory. "No example of a barter economy, pure and simple, has ever been described, let alone the emergence from it of money," anthropologist Caroline Humphrey (1985: 48) has emphasized, stating that "all available ethnography suggests that there never has been such a thing" (cited in Graeber 2014: 29; for further critiques see Wray 2004 and Ingham 2004). As for the cuneiform record, it shows that the major initial monetary activity of most Mesopotamians was to pay taxes, fees or to buy products that palaces and temples made or imported, on credit provided or regulated by these large institutions.

As far as convenience is concerned, the simplest and least costly way to conduct exchange is to circumvent direct payment in metal. Having to weigh money for retail or even larger exchanges would have maximized transaction costs. Yet when anti-government ideologues argue that commodity money and bank credit minimize transactions costs (Ober 2008: 49f., echoing the ideas of North 1992), they compare coinage only to barter, not to credit, e.g., settled on the threshing floor at harvest time. The Barter Theory excludes the thought that palatial credit creation and regulation served to minimize transactions costs and indeed, to preserve economic stability.

The Barter Theory's lack of evidence did not trouble Menger, because his logic was purely speculative: "even if money did not originate from barter, could it have?" Prehistorians and anthropologists would answer, "No, it couldn't have happened that way." Money always has been embedded in a public context, and hardly could have evolved without public catalysts and ongoing oversight to make it acceptable.

For starters, any practical payment system for credit and trade requires accurate weighing and measuring. This calls for public oversight as a check on fraudulent practice. Trust cannot be left to individuals engaging in barter or credit on their own. Crooked merchants historically have used light weights when selling goods or

lending out money so as to give their customers less, and heavy weights when buying or collecting debts so as to gain an unduly large amount of silver or other commodities (See Powell 1999 for discussion.)

Biblical denunciations of merchants using false weights and measures find their antecedents in Babylonia. Hammurabi's laws (gap **x** [Roth 1997: 98], sometimes referred to as §94 and §95) stipulates that merchants who lend grain or money by a small weight but demand payment using a larger measure should forfeit whatever they had lent. Ale-women found guilty of using crooked weights and measures in selling beer were to be cast into the water (§108 [Roth 1997: 101]). Many other rulings deal with creditor abuses, which date back to the rule of Urukagina of Lagash (c. 2350 BC).

Such abuses are timeless. The seventh century BC prophet Amos (8: 5ff.) depicts the Lord as denouncing wealthy Israelites "who trample the needy and do away with the poor of the land" by scheming, "skimping the measure (making the ephah small), boosting the price (making the shekel great), and cheating with dishonest scales." Likewise the prophet Micah (6:11) denounces merchants using "the short ephah, which is accursed? Shall I acquit a man with dishonest scales, with a bag of false weights?" (Kula 1986 reviews Biblical and Koranic examples).

Leviticus (19: 35f.) describes the Lord as directing Moses to instruct his followers not to use dishonest standards when measuring length, weight or quantity. Deuteronomy 25: 13–15 admonishes: "Thou shalt not have two differing weights in your bag – one heavy, one light. Thou shalt not have two differing measures in your house – one large, one small. You must have accurate and honest weights and measures … For the Lord your God detests … anyone who deals dishonestly."

Regulating weights and measures was a step far beyond primitive barter among individuals. It needed official organization and supervision of exchange and credit. As noted above, the sexagesimal weights to denominate minas and shekels reflect the priority of transactions within Mesopotamian palaces and temples, deriving from their grain-based accounting system to schedule and distribute food. Jewish temples likewise provided standardized measures (Exodus 30.13 and 38.24–27, and Leviticus 27.25; for royal measures see 2 Samuel 14.26), as did the Athenian *agoranomoi* (public market regulators). Throughout antiquity markets were located in the open spaces in front of city gates or temples, providing easy access to official weights and measures to prevent fraud.

In addition to public oversight of weighing and measuring, quality standards were required for alloys of silver and gold. Sales and debt contracts from the second and first millennia BC typically specified payment in silver of 7/8 purity (.875 fine, the equivalent of 21 carats). To avoid adulteration, silver was minted in temples to guarantee specified degrees of purity. The word "money" derives from Rome's Temple of Juno Moneta, where silver and gold coinage was struck during the Punic Wars, mainly to arm soldiers, build a navy and pay mercenaries – not for barter exchange.

Formal coinage was not required for these functions. Weighed metal was sufficient, often stamped by temples to attest to its degree of purity. Long before

coins were struck in the first millennium BC, raw silver (*hacksilver*) and weighed jewelry served the function that coinage did in classical times. Although coinage is not attested before the seventh century BC, Balmuth (1967, 1971) may have located second-millennium Near Eastern antecedents, citing an inscription by the Assyrian ruler Sennacherib (705–681 BC) saying that he "fashioned molds of clay and poured bronze therein, as in casting (fashioning) half-shekel pieces" (Balmuth 1971: 2). The earlier Ugaritic epics of Aqht and Krt describe "flows of tears . . . as resembling 1/4 shekels or pieces-of-four and 1/5 shekels or pieces-of-five." "Like markets, coinage was there before the Greeks," summarizes Powell (1999: 22); "the only significant difference that coinage makes in money transactions is the guarantee of quality; in Babylonia, as elsewhere, silver coins were cut up just like other silver and put in the balance pan."

There were no public debts to serve as a monetary base for bank reserves as in today's world. Throughout antiquity temples were society's ultimate bankers and sources of money in emergencies. Sacred statues were adorned with golden ornaments that could be melted down in times of need to pay mercenaries (or perhaps pay ransom or tribute; see Oppenheim 1949), much as were the Winged Victory statues of Athens during the Peloponnesian War.

The Monetary Role of Silver

Archaic palaces played the major role in importing silver and supplying it to the economy at large. Silver owed its status not to its technological use value in production, but to its role in settling debt balances owed to the palace, as well as the paradigmatic religious donation or commission to the temples.

Most silver was obtained by the palace mobilizing Mesopotamia's crop surplus to supply weaving and other workshops producing handicrafts to export. Silver and also gold from Cappadocia (in Asia Minor) was sold to Elam and the Indus Valley (via the island entrepôt of Dilmun/Bahrain) for tin. Late third-millennium BC records show that when merchants "receive silver and copper [from the palace] they are being paid to undertake a commission, not being issued with a commodity for disposal on the open market" (Postgate 1992: 220).

Denominating prices in silver forced reliance on scales. To prevent the awkwardness of weighing relatively small pieces, silver was cast into jewelry, such as bracelets made with easily broken-off segments measured in shekels. Powell (1977) notes that the Middle Babylonian word for 1/8 shekel, *bitqu* (literally "cutting"), suggests silver rings and coils, and may originally have denoted "a piece of standard size cut off from such a silver coil." Such jewelry money gave way to coinage in classical antiquity as officially stamped weights of precious metal.

Throughout the history of Sumer "the management of silver and gold, of textiles, and of other precious or 'luxury' goods is largely dominated by the royal palace," notes Garfinkle (2012: 244f., 226). Even in Lagash c. 2350 BC the temples "did not actively control the politically important treasuries" but were under control

of the palace. In neighboring Umma numerous branches of the economy converted their primary goods annually "into tiny sums of silver, which were collected by the province and then delivered to the state in the form of donations to a religious festival. … Luxury goods – textiles, silver and gold, meat, and special edible delicacies – are primarily found in one specific context, namely the palace."

Silver was used primarily by the palace and the entrepreneurs managing its trade and other enterprise. The result was a bifurcated economy, in which entrepreneurial trade and management operated on a silver standard atop a rural economy on a grain standard. Grain and wool were the main means of denominating and paying agricultural fees and debts, to be paid on the threshing floor or in the shearing season.

What is not well understood is how silver got into the hands of Mesopotamia's general population. Some would have been obtained by selling crops, textiles or other handicrafts to the palace (Sallaberger 2013, cited in Garfinkle 2012: 245), or to entrepreneurs who earned silver on their trade and management of public infrastructure. The palace paid some silver to mercenaries, and there are hints of rulers handing out silver tokens to soldiers after victory and perhaps at royal festivals. But most of the influx of silver from foreign trade was re-invested in more trade ventures or lent out in rural usury for current income and, ultimately, to acquire land. Silver lying around not lent out was called "hungry" for profit-making opportunities.

There is little hint of speculative credit, and no sudden gluts of silver such as occurred in classical antiquity when Alexander the Great looted the temples and palaces of their bullion in the lands he conquered, coined the booty and put it into circulation by paying his army. There also was no monetary drain, thanks to the fact that grain could be used as a substitute for silver at a stipulated exchange rate for payments to the large institutions. Personal debts mounted up rapidly as a result of agrarian usury, and inability to pay them often resulted from crop failure, drought, or the debtor's illness or other misfortune.

Royal proclamations cancelling agrarian debts preserved economic viability on the land. Public oversight of money thus went hand in hand with public management of debt, including the setting of interest rates and the customary royal amnesties for agrarian and personal debts. Ultimately underlying the conflict between the Barter and State theories of money is thus whether public policy should favor creditors or debtors. As creditors, banks seek "hard" debt collection rules. But governments recognize that most of their population are in debt and need to be protected from forfeiting their income and property to creditors, which would impoverish the economy at large.

Archaic Money and Interest-Bearing Debt

Interest always has been an inherently monetary phenomenon and officially regulated. The standard Mesopotamian interest rate for commercial loans denominated in silver was set to dovetail into palatial accounting practice at the

"unit fraction": one shekel (60th) per mina per month, 12 shekels a year (the equivalent of 20% annual interest in decimalized terms), doubling the principal in 5 years. Interest rates throughout antiquity emulated this practice for ease of calculation in terms of the "unit fraction," e.g., an ounce per pound in Rome. Rates were set simply for reasons of mathematical simplicity in Mesopotamia's sexagesimal system of fractional weights and measures. These rates remained traditional for centuries, not being related to productivity, profit levels, or risk.

The ancient words for interest – *mash* (goat) in Sumerian and Akkadian, and *tokos* and *faenus* (calf) in Greek and Latin are used in the metaphoric sense for "that which is born or produced." What was "born" was not goats or calves, but interest, on the new moon each month (Hudson 2000 discusses the semantics).

Much as the Barter Theory of money hypothesizes trade as leading to the emergence of money without any need for a public interface, its adherents have put forth an individualistic pre-monetary productivity theory of interest. According to this origin myth, early interest was paid by individual debtors to well-to-do creditors "in kind," out of seeds or animals. (Böhm-Bawerk's *Capital and Interest* (1890 [1884]) surveyed and refuted what he called "naïve productivity" theories of interest.) This scenario depicts the origins of interest-bearing debt as being productive and hence economically justified – and occurring without silver or other official money.

The classic attempt to depict such pre-monetary interest already in the Neolithic as reflecting productivity (and implicitly, profit) rates c. 5000 BC – subject to the risk of nonpayment – is Fritz Heichelheim's *Ancient Economic History, from the Palaeolithic Age to the Migrations of the Germanic, Slavic and Arabic Nations* (1958: 54): "Dates, olives, figs, nuts, or seeds of grain were probably lent out . . . to serfs, poorer farmers, and dependents, to be sown and planted, and naturally an increased portion of the harvest had to be returned in kind." In addition to fruits and seeds, "animals could be borrowed too for a fixed time limit, the loan being repaid according to a fixed percentage from the young animals born subsequently. . . . So here we have the first forms of money that man could use as a capital for investment, in the narrower sense." Such "food-money" supposedly was lent out in the form of seeds and animals, at interest rates reflecting their reproduction rates. This scenario depicts "money" as originating not as taxes or other payments to palaces or temples but as capital in the form of seeds and animals, capable of producing an economic surplus as interest paid in kind, at a rate reflecting physical productivity.

The problem with this mythology is that the traditional communities known to anthropologists do not lend or borrow cattle, either for calf-interest or other payment (Sundstrom 1974: 34 and 38, and Hoebel 1968: 230). When seeds are advanced, it typically is by absentee landowners to sharecroppers. Debtors are obliged to pledge (and forfeit) their livestock to creditors out of need to survive and pay usury out of their *own* resources, not from investing the creditor's livestock or seeds at a profit.

Like the Barter Theory of money, the Productivity Theory of interest takes interest out of its historical context, treating money simply as a commodity owned

by individuals, without public oversight or regulation. This is assumed to be the "natural" condition and, as such, applicable to today's world – with government money creation and regulation depicted as unnatural, not original.

If Heichelheim's scenario were valid, interest rates would have varied with the productivity of the cattle, seeds or mercantile profit rates. But interest rates remained standardized over many centuries, being set independently from the production process or profit rates on trade.

Barter-based "naïve productivity" theories of interest envision transactions among individuals acting on their own account, with borrowers hoping to make a gain out of which to pay interest. This reverses the historical line of development. The paradigmatic interest-bearing debts were owed to Mesopotamia's palaces and temples. Interest charges did not reflect physical productivity but were specifically monetary, paid in silver at a stipulated rate – for instance, for the advance of export goods to long-distance traders, paid out of their mercantile profits. Most fatal to productivity theories of interest is that the majority of Mesopotamian agrarian debts did not result from actual loans but accrued as arrears (see Wunsch 2002). Agrarian interest often was charged only after the "due date" was missed. In such cases one could say that interest was paid for the *failure* of productivity to keep up with normal expectations.

Mesopotamia did not have banking in the modern sense of taking in deposits and lending them out at a profit. Even in Neo-Babylonian times "banking families" such as the Egibi were simply wealthy families. They paid depositors the same rate (equivalent to 20%) as they charged customers, so there was no intermediation markup as in modern banking (Bogaert 1966).

The major policy tool for rulers to stabilize the economy and save their subjects from debt bondage was to proclaim Clean Slates wiping out the overgrowth of debt in excess of the ability to pay. Productive "silver loans" to commercial traders and managers were not subject to these amnesties. The exemption of credit from such royal Clean Slate proclamations shows a policy distinction between productive and unproductive credit – the contrast that medieval Church Fathers would draw between interest and usury.

Classical Antiquity's Changing Context for Money and Credit

Interest-bearing debt is found spreading westward to the Mediterranean lands around the eighth century BC, mainly via Syrian and Phoenician traders establishing trading enclaves (Hudson 1994). They brought with them weights and measures that were adopted by Greeks and Italians. A. E. Berriman's *Historical Metrology* (1953) points out that the carat originally was the weight of a carob grain, *Ceratonia siliqua*, a tree native to the Mesopotamian meridian, weighing 1/60th of a shekel. The Greek term is *keration* ("small grain").

Greek and Roman elites also adopted the Near Eastern practice of setting interest rates in accordance with the local unit-fraction, e.g., Rome's duodecimal system dividing the pound into 12 troy ounces. One ounce per pound per year (1/12th) was

the equivalent of an 8 1/3% rate of interest. That was much lower than Mesopotamia's agrarian interest rate of one-third of the principal (or one-fifth for commercial loans), but debts in Rome and Greece were inexorable and hence ultimately more burdensome.

Classical Greek experience confirms a number of generalities that can be drawn from earlier Near Eastern monetary development. Describing how the commercial Isthmus city of Corinth adopted coinage c. 575–550 BC (a generation after its origins in Aegina), Salmon (1984: 170f.) supports the conclusion of the numismatist C.M. Kraay (1976: 317–322): "coinage cannot have been intended to facilitate trade, either at a local level or on a wider scale." Early money was to finance credit transactions, not the exchange of goods (Salmon 1984: 171f.): "From the earliest issues to the second half of the fourth century, at least in Corinth, the association between coins and trade was mainly that they offered a means of providing credit. If they had acted as an item of trade themselves we should have expected them to travel much further, and in far greater quantities, from Corinth than they in fact did. Their main function was to be lent at Corinth for purchase of items to be traded."

Money was mainly for paying taxes and fees; Salmon continues: Corinthian "coins were first issued in order to serve the purposes of the minting authorities. Cities would find it convenient if payments made to them – taxes, fines, etc. – were in the form of coins whose purity and weight were fixed; while payments made by the state from time to time for building schemes, mercenaries, and other purposes could be much simplified if trustworthy coins were available."

However, taxation developed only slowly in Greek cities. Greece and Rome obtained bullion not from tax revenue or public enterprise but from war booty, by levying tribute or, in Athens, from local silver mines. Spending was the key, mainly to pay soldiers and hire mercenaries. In the Ionian cities of Asia Minor, money's primary role was to pay "allowances to the sailors manning the huge fleet being prepared by the rebels" (Figueira 1981: 157). "Hectataeus of Miletus did not propose a large capital levy or other forms of taxation to build up the allied fleet, but a confiscation of the treasures at Branchidae (Hdt. 5.36.3-4). This may suggest that taxation was primitive in early fifth-century Ionia."

Military conquest remained the major source of monetary metal from Alexander the Great's looting of temples and palaces down through the end of antiquity in Rome. Armies brought minters along to melt down the booty and distribute it to their commanders and troops, with a tithe to the city-temple. When there were no more realms for imperial Rome to conquer and extract tribute, the inability to tax the oligarchic economy led to debasement of the coinage. Replacing the State Theory of money by treating money simply as a commodity led to a monetary drain – ultimately forcing resort to barter.

The main difference between Greek and Roman economies and those of the Ancient Near East was the absence of debt relief, resulting in a long series of political crises extending from the seventh-century BC "tyrants" (populist reformers) from classical Sparta and Corinth down to Rome in the first century BC. Mid-nineteenth-century historians attributed these debt crises to the introduction

of coinage around the seventh and sixth centuries BC, when Greek city-states issued coins imprinted with their city-images, such as the owls of Athens. But moneychangers still weighed coins from the various cities, in keeping with the use of weighed bullion that predated coinage by about 2000 years.

The economic impact of coinage thus did not differ much from that of hacksilver. So it was not money, coinage, or even interest-bearing debt by themselves that caused the polarization under antiquity's creditor oligarchies. The problem was the way in which society handled the proliferation of interest-bearing debt.

As credit was increasingly privatized, debt became a dynamic powerful enough to dissolve the checks and balances that had shaped the social context in which money first developed. Mesopotamia had usury and debt bondage, but its rulers managed to avoid the irreversible disenfranchisement and ultimate serfdom that plagued the Mediterranean lands. The Near Eastern aim was to preserve a land-tenured citizenry supplying the palace with corvée labor and military service. Despite the palace's role as the major creditor, it protected debtors by debt amnesties that undid the polarizing effect of interest-bearing debt. Most debts in early Mesopotamia were owed to the palace, so rulers basically were cancelling debts owed to themselves and their collectors when they proclaimed Clean Slates that saved their economies from widespread debt bondage that would have diverted labor to work for creditors at the expense of the palace.

But as debts came to be owed mainly to Greek and Roman oligarchies, debts no longer were canceled except in military or social emergencies to maintain the *demos*-army's loyalty. What came to be "sanctified" was the right of creditors to foreclose, not cancelling debts to restore economic balance.

Money and debt in Greece and Rome thus followed a different trajectory from its origins in Mesopotamia. Oligarchies gained sufficient power to stop civic debt cancellations. Rural usury in Greece and Rome expropriated indebted citizens from their land irreversibly, typically to become mercenaries in armies formerly manned by self-supporting citizens. Land ownership was much more concentrated than in Bronze Age Mesopotamia or even in the contemporary Neo-Babylonian economy.

Today's mainstream ideology maintains this shift to hard pro-creditor law and depicts nonpayment of debts as leading to chaos. Yet Clean Slates are what *saved* Near Eastern economies from the chaos of economic polarization and widespread bondage. Mesopotamia's economic takeoff could not have been sustained if rulers had adopted modern creditor-oriented rules.

Nor was classical antiquity's takeoff sustained. By the closing centuries of the Roman Empire, wealthy elites had monopolized the land and stripped the economy of money, spending most of what they had on imports that drained monetary silver and gold to the East – leaving a barter economy in its wake as the "final" or "third" stage of monetization: impoverishment and polarization in which money was stripped away.

This post-Roman oligarchic collapse into local self-sufficiency and barter reverses the once-held idea that exchange evolved *from* barter via monetization to credit economies. Yet textbooks still repeat that sequence without recognizing the early role of credit, without mentioning the palaces and temples where

monetization first evolved, or citing the tendency of debts to be mathematically self-expanding when not overridden by debt writedowns and clean slates. If such economic theorizing really were universal, history simply could not have occurred in the way it did.

Also reversed today is understanding of how the charging of interest originated. Instead of reflecting productivity, profitability, or risk, interest rates were officially administered and remained remarkably stable in each region throughout antiquity. Today's governments continue to regulate interest rates. Yet mainstream economic theory continues to propose interest-rate models based not Treasury fiscal and monetary policy, but on profit rates, "risk," and consumer "choice."

Summary: The Shifting Historiography of Money's Origins

Origin myths at odds with the historical record are the result of the conflict between vested interests and reformers over whether the monetary and credit system should be controlled by banks or by governments. Are credit and debt to be administered by laws favoring creditors or should the prosperity of the indebted population at large be protected? The way in which economic writers answer this question turns out to be the key to their preference regarding the Barter or State Theories of the origins and character of money, credit, and interest.

Assyriological and anthropological research confirms that money and monetary interest were not created by individuals trucking and bartering crops and handicrafts or lending crops and animals with each other. Archaic economies operated on credit, creating money as means of paying debts, mainly to Mesopotamia's palaces and temples. Interest emerged as the means of financing long-distance trade and advancing land to its cultivators or managers, administered mainly by palace officials.

Recognition of this palatial origin of money and interest is at odds with the drive by commercial bankers to depict their own control of money and credit as being natural and primordial. Ever since Roman law was written to favor creditors, history has been written to defend the view that private credit and the "sanctity" of debts being paid is natural. The resulting mythology to explain the origins of money and interest reflects public relations lobbying by bankers and other creditors.

Goodhart (9) highlights the relevance to modern times of misinterpreting the history of money: It underlies creation of the euro. The eurozone was created without a central bank to monetize budget deficits for EU member governments. The anti-state ideology underlying the euro thus stands in opposition to the State Theory of money. Central bank credit is to be created only to bail out commercial banks for losses on their own credit creation and bad investments, not for governments to spend directly into the economy.

What makes today's monetary system opposite from that of Bronze Age Mesopotamia is an ideology that recognizes no role for money and credit creation

except to benefit creditors. Understanding how the origins of money went hand in hand with checks and balances to protect economies from being polarized and impoverished by debt would call for treating money and credit as part of the overall economic system, not merely a matter of "individual choice." To view contractual monetary and debt arrangements between individual lenders and borrowers without regard for how the overgrowth of debt may disrupt the economy is not only a travesty of economic history but is largely responsible for today's short-termism in re-enacting the debt crises that plagued classical antiquity.

References

Balmuth M (1967) Monetary forerunners of coinage in Phoenicia and Palestine in Antiquity. In: Kindler A (ed) The patterns of monetary development in Phoenicia and Palestine in Antiquity, Jerusalem

Balmuth M (1971) Remarks on the appearance of the earliest coins. In: Mitten DG, Pedley JG, Scott JA (eds) Studies presented to George M. A. Hanfmann, Cambridge, MA, pp 1–7

Bell SA, Nell EJ (eds) (2003) The state, the market and the euro: chartalism versus metallism in the theory of money. Edward Elgar, Cheltanham, pp 1–25

Berriman AE (1953) Ancient Metrology (London: J. M. Dent & Sons)

Bogaert R (1966) Les Origines Antiques de la Banque de Depôt. Sijthoff, Leiden

Bongenaar ACVM (ed) (2000) Interdependency of institutions and private entrepreneurs. MOS studies, vol 2. Nederlands Institut voor het Nabije Oosten, Istanbul

Charpin D (1987) Les Decréts Royaux a l'Epoque Paleo-babylonienne, à Propos d'un Ouvrage Recent. AfO 34:36–44

Charpin D (1990) Les Edits de 'Restauration' des Rois Babyloniens et leur Application. In: Nicolet C (ed) Du Pouvoir dans L'Antiquite: Mots et Realities. Droz, Geneva, pp 13–24

Charpin D (1998) Les prêteurs et la palais: Les édits de mîšharum des rois de Babylone et leurs traces dans les archives privées. In: Bongenaar, ACVM (ed) Interdependency of institutions and private entrepreneurs (proceedings of the second MOS symposium, Leiden 1998), Nederlands Historisch-Archaeologisch Institute, Te Istanbul, 2000, pp 185–211

Cripps EL (2017) The structure of prices in the neo-Sumerian economy (I). Silver Price Ratios, Barley. http://cdli.ucla.edu/?q=cuneiform-digital-library-preprints

Dalton G (1965) Primitive money. Am Anthropol 67:44–65

Dercksen JG (ed) (1999) Trade and finance in ancient Mesopotamia. MOS studies, vol 1 [Leiden 1997]. Nederlands historisch-Archaelogisch Instituut, Istanbul

Dercksen et al (2016) "Kaspum lasqul, 'Let me weigh out the silver.' Mesopotamian and Anatolian Weights during the Old Assyrian Period," in Kleber and Pirngruber (eds): 13–25. K. Kleber and R. Pirngruber (eds) (2016), Silver, Money and Credit. A Tribute to Robertus J. van der Spek on Occasion of his 65th Birthday on 18th September 2014 (Leiden: Nederlands Instituut voor het Nabije Oosten)

Englund R (1988) Administrative timekeeping in ancient Mesopotamia. J Econ Soc Hist Orient 31:121–185

Englund R (2004) Proto-cuneiform account-books and journals. In: Hudson, Wunsch (eds) Creating economic order: record-keeping, standardization and the development of accounting in the ancient Near East. CDL Press, Bethesda, pp 23–46

Englund R (2012) Chapter 21: Equivalency values and the command economy of the Ur III period in Mesopotamia. In: Urton G, Papadopoulos J (eds) The construction of value in the ancient world. Cotsen Institute of Archaeology Press, Los Angeles, pp 427–595

Figueira TJ (1981) Aegina: society and politics. Arno Press, New York
Finkelstein JJ (1961) Ammisaduqa's edict and the Babylonian 'law codes'. JCS 15:91–104
Finkelstein JJ (1965) Some new misharum material and its implications, in AS 16. Studies in Honor of Benno Landsberger on his Seventy-Fifth Birthday, pp 233–246
Finkelstein JJ (1969) The edict of Ammisaduqa: a new text. RA 63:45–64
Finley M (1981) Economy and society in ancient Greece. Chatto and Windus, London
Garfinkle SJ (2004) Shepherds, merchants, and credit: some observations on lending practices in Ur III Mesopotamia. J Econ Soc Hist Orient 47:1–30
Garfinkle SJ (2012) Entrepreneurs and enterprise in early Mesopotamia: a study of three archives from the Third Dynasty of Ur (2112–2004 BC). CDL Press, Bethesda
Gelb IJ (1965) The ancient Mesopotamian ration system. J Near East Stud 24:230–243
Goodhart C (1998) The two concepts of money: implications for the analysis of optimal currency areas. Eur J Polit Econ 14:407–432
Graeber D (2014) Debt: the first 5000 years. Melville Press, Brooklyn
Grierson P (1977) The origins of money. Athlone Press, London
Hartman LF, Oppenheim L (1950) On beer and brewing techniques in ancient Mesopotamia. Supplement to the Journal of the American Oriental Society, 10
Hawkins JD (1988) Royal statements of ideal prices: Assyrian, Babylonian, and Hittite. In: Canby JV et al (eds) Ancient Anatolia: aspects of change and cultural development. Essays in honor of Machteld J. Mellink. Madison, pp 93–102
Heichelheim FM (1958) An ancient economic history, from the palaeolithic age to the migrations of the Germanic, Slavic and Arabic nations, vol I. Rev. ed., Sitjhoff, Leiden
Hoebel EA (1968[1964]) The law of primitive man: a study in comparative legal dynamics. Atheneum, New York
Hudson M (1992) Did the Phoenicians introduce the idea of interest to Greece and Italy – and if so, when? In: Kopcke G, Tokumaru I (eds) Greece between East and West: 10th–8th centuries BC. Philipp von Zabern, Mainz, pp 128–143
Hudson (1994) Land Monopolization, Fiscal Crises and Clean Slate 'Jubilee' Proclamations in Antiquity. In: Hudson, Miller and Feder (eds) A Philosophy for a Fair Society, London, pp 33–79
Hudson M (1996) Privatization in history and today: a survey of the unresolved controversies, and the dynamics of privatization, from the Bronze Age to the present. In: Baruch Levine (ed) Privatization in the Ancient Near East and Classical Antiquity. Peabody Museum, Harvard, Cambridge, Mass, pp 1–72
Hudson M (2000) How interest rates were set, 2500 BC – 1000 AD: Máš, tokos and fænus as metaphors for interest accruals. J Econ Soc Hist Orient 43:132–161
Hudson M (2002) Reconstructing the origins of interest-bearing debt and the logic of clean slates. In: Hudson M, Van De Mieroop M (eds) Debt and economic renewal in the ancient Near East. CDL Press, Bethesda, pp 7–58
Hudson M (2003) The Cartalist/Monetarist debate in historical perspective. In: Bell S, Nell E (eds) The state, the market and the euro. Edward Elgar, Cheltenham
Hudson M (2004a) The development of money-of-account in Sumer's temples. In: Hudson M, Wunsch C (eds) Creating economic order: record-keeping, standardization and the development of accounting in the ancient Near East. CDL Press, Bethesda, pp 303–329
Hudson M (2004b) The archaeology of money in light of Mesopotamian records. In: Wray R (ed) Credit and state theories of money: the contributions of A. Mitchell Innes. Edward Elgar, Cheltenham
Hudson M (2018) . . .and forgive them their debts: credit and redemption from Bronze Age finance to the Jubilee Year. (Dresden, ISLET Verlag 2018)
Hudson and Wunsch (2004) Michael Hudson and Cornelia Wunsch, ed., Creating Economic Order: Record-Keeping, Standardization and the Development of Accounting in the Ancient Near East (CDL Press, Bethesda, Md.)
Humphrey C (1985) Barter and economic disintegration. Man, New Series 20:48

Ingham G (2004) The nature of money. Polity Press, Cambridge
Innes AM (1913) What is money? Banking Law J May:377–408, reprinted in Randall Wary L (ed) Credit and state theory of money. The contributions of A. Mitchell Innes. Edward Elgar, Cheltenham
Innes AM (1914) The credit theory of money. Banking Law J Dec/Jan:151–168, reprinted in Randall Wary L (ed) Credit and state theory of money. The contributions of A. Mitchell Innes. Edward Elgar, Cheltenham
Jursa M (2002) Debts and indebtedness in the Neo-Babylonian period: evidence from institutional archives. In: Hudson, Van De Mieroop (eds) Debt and economic renewal, pp 197–220
Knapp GF (1924[1905]) Knapp State theory of money: Macmillan & Co, London
Kraay CM (1976) Archaic and classical Greek Coins. Methuen, London
Kraus FR (1958) Ein Edikt des Königs Ammi-saduqa von Babylon. SD, vol 5. Brill, Leiden
Kula W (1986) Measures and men. Princeton University Press, Princeton
Lambert M (1960) La naissance de la bureaucratie. Rev Hist Demogr 224:1–26
Lambert M (1961) Le premier triomphe de la bureaucratie. Rev Hist Demogr 225:21–46
Lambert M (1963) L'Usage de l'argent-metal a Lagash au temps de la IIIe Dynastie d'Ur. Revue d'Assyriologie 57:79–92
Larsen MT (1976) The old Assyrian city-state and its colonies. Akademisk Forlag, Copenhagen
Larsen MT (2015) Ancient kanesh a merchant colony in bronze age anatolia. Cambridge University Press, New York/Cambridge
Laum B (1924) Heiliges Geld: eine historiche Untersuchung über den sakralen Ursprung des Geldes. Mohr, Tübingen
Laum B (1952) Geschichte der öffentlichen Finanzwirtschaft im Altertum und Frühmittelalter. In: Gerloff W, Neumark F (eds) Handbuch der Finanzwissenschaft, 2nd edn, Tübingen, pp 211–235
Leemans (1950) The Old Babylonian merchant: his business and social position. Brill, Leiden
Manning JG, Morris I (2005) The ancient economy: evidence and models. Stanford University Press, Stanford
Menger C (1892) On the origins of money. Econ J 2:238–255. Originally published in 1871
Menger C (1871) Principles of economics. New York University Press, New York
Michel C (2013) Economic and social aspects of the old Assyrian loan contract. In: D'Agostino F (ed) L'economia dell'antica Mesopotamia (III-I millennio a.C.), Rome, pp 41–55
Ober J (2008) Democracy and knowledge. Innovation and learning in classical Athens. Princeton University Press, Princeton
Oppenheim L (1949) The golden garments of the gods. J Near East Stud 8:172–193
Polanyi K (1944) The great transformation. Beacon Press, Boston
Polanyi K, Arensberg CM, Pearson HW (eds) (1957) Trade and market in the early empires: economies in history and theory. Free Press, New York
Postgate JN (1992) Early Mesopotamian society and economy at the dawn of history. Routledge, London/New York
Powell MA (1977) Sumerian Merchants and the Problem of Profit. Iraq 39:23–29
Powell MA (1999) Wir müssen alle unsere Nische nuzten: monies, motives, and methods in Babylonian economics. In: Dercksen JG (ed) Trade and finance in Ancient Mesopotamia, Istanbul, pp 5–23
Pritchard JB (1985 [1955]) Ancient Near Eastern texts. Princeton University Press, Princeton. Paperback, 1958
Radner K (1999) Money in the Neo-Assyrian Empire. In: Dercksen JG (ed) Trade and finance in ancient Mesopotamia. MOS studies, vol 1, proceedings of the first MOS symposium [Leiden 1997]. Nederlands historisch-Archaelogisch Instituut, Istanbul, pp 127–157
Renger J (1979) Interaction of temple, palace, and 'private enterprise' in the old Babylonian economy. In: Lipinski E (ed) State and temple economy in the ancient Near East, vol I. Departement Orientalistiek, Leuven, pp 249–256
Renger J (1984) Patterns of non-institutional trade and non-commercial exchange in ancient Mesopotamia at the beginning of the second millennium BC. In: Archi A (ed) Circulation of goods in non-palatial context in the ancient Near East. Incunabula Graeca, vol LXXII, Rome, pp 31–115

Roth MT (1997) Law collections from Mesopotamia and Asia Minor, 2nd edn. Scholars Press, Atlanta

Sallaberger W (2013) The management of royal treasure: palace archives and palatial economy in the ancient Near East. In: Hill JA, Jones P, Morales AJ (eds) Experiencing power, generating authority: cosmos, politics and the ideology of kingship in Ancient Egypt and Mesopotamia. University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, pp 213–255

Salmon JB (1984) Wealthy Corinth. Clarendon Press, Oxford

Samuelson P (1973) Economics, 9th edn. McGraw Hill, New York

Steinkeller P (1981) The renting of fields in early Mesopotamia and the development of the concept of 'interest' in Sumerian. J Econ Soc Hist Orient 24:113

Steinkeller P (2002) Money lending practices in Ur III Babylonia: the issue of economic motivation. In: Hudson M, Van De Mieroop M (eds) Debt and economic renewal in the ancient Near East. CDL Press, Bethesda, pp 109–137

Sundstrom L (1974[1965]) The exchange economy of pre-colonial Africa. St. Martin's Press, New York, p 34, 38

Van De Mieroop M (1995) Old Babylonian interest rates: were they annual? In: Lipinski FE (ed) Immigration and emigration within the Near East. Orientalia Lovaniensia Analecta, vol 65. Departement Oriëntalistick, Louvain, pp 357–364

Van De Mieroop M (2002a) Credit as a facilitator of exchange in Old Babylonian Mesopotamia. In: Hudson M, Van De Mieroop M (eds) Debt and economic renewal in the ancient Near East. CDL Press, Bethesda, pp 163–174

Van De Mieroop M (2002b) A history of Near Eastern debt? In: Hudson, De Mieroop V (eds) Debt and economic renewal, pp 59–94

Van De Mieroop M (2004) Accounting in early Mesopotamia: some remarks. In: Hudson, Wunsch (eds) Creating economic order. ISLET, Dresden, pp 47–64

Van De Mieroop M (2005) The invention of interest. In: Goetzmann WN, Geert Rouwenhorst K (eds) The origins of value: the financial innovations that created modern capital markets. Oxford University Press, Oxford, pp 17–30

Veenhof KR (1972) Aspects of old Assyrian trade and its terminology. Brill, Leiden

Veenhof KR (2010) The interpretation of paragraphs **t** and **u** of the Code of Hammurabi. In: Hazirlayan Y, Dönmez S (eds) DUB.SAR E.DUB.BA.A: studies presented in honour of Veysel Donbaz, Istanbul

Ventris M, Chadwick J (1956) Documents in Mycenaean Greek: three hundred selected tablets from Knossos, Pylos and Mycenae with commentary and vocabulary. Cambridge University Press, Cambridge

von Böhm-Bawerk E (1890[1884]) Capital and Interest: A Critical History of Economic Theory. London, Macmillan & Co

Wray LR (ed) (2004) Credit and state theories of money: the contributions of A. Mitchell Innes. Edward Elgar, Cheltenham

Wunsch C (2002) Debt, interest, pledge and forfeiture in the Neo-Babylonian and early Achaemenid period: the evidence from private archives. In: Hudson M, Van De Mieroop M (eds) Debt and economic renewal in the ancient Near East. CDL Press, Bethesda, pp 221–255

The Role of Money in the Economies of Ancient Greece and Rome

3

Colin P. Elliott

Contents

Abstract

Coined money was first invented in the Greco-Roman world. The monetization of Greek and Roman societies was a complex, dynamic, and often experimental process in which the economics of money were inescapably connected with cultural, political, and social developments. How did money contribute to the spread of market exchange and the development of sprawling territorial empires? Certainly, the rise of Greek democracy coincides with the adoption of coinage. Furthermore, in the wake of Alexander the Great's conquests, the Eastern Mediterranean went through processes of both Hellenization and monetization – as Greek-style coinage systems promoted the development of financial institutions and market exchange. The course of monetary development, however, was not a uniform process. Many of the Celtic cultures of Western Europe simply incorporated Greek coins into their existing traditions of reciprocity. Greek city states on the Italian peninsula used coinage, but inhabitants of the Latin-speaking cities of central Italy, including Rome, were comparably under-monetized. Rome, in fact, did not adopt a coherent coinage system until the necessities of the Second Punic

C. P. Elliott (✉)
Department of History, Indiana University, Bloomington, IN, USA
e-mail: cpe@indiana.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_46

War forced Rome to finally adopt a Greek-style coinage system. By the late first century B.C., however, money served much of the Mediterranean world in one form or another. How prevalent was money use? Even under the *Pax Romana* in the first and second centuries A.D., the lines between "general purpose" and 'special purpose' money were blurred. Money may have only been useful to autarkic peasants whenever taxes and tributes were due. For urbanites, however, money served them regularly in interactions with strangers, neighbors, and even family. Barter, commodity money, and credit maintained a role throughout Greco-Roman antiquity. By the third-century A.D., the Roman monetary system came under strain as coinage standards became erratic. Eventually, all competing provincial and local coinages were abolished in favor of a single central coinage supported by strict legal tender laws. This experiment with fiduciary – like many in ancient Greek and Roman societies – was of mixed success. Money's complex history and diverse functions in the Greco-Roman continue to captivate economic historians.

Keywords
Roman Empire · Ancient Greece · Early coinage · Denarius

Introduction

A range of materials and goods, both physical and virtual, performed the functions of money in the societies of the Classical Mediterranean. Coins were first minted in the unique social, cultural, and political context of the Greek city-state – a fact which inescapably dictated the trajectory of monetization in the Western world. In some parts of the ancient Mediterranean, coinage spread rapidly, while in other places – such as Egypt and central Italy – coinage did not enjoy regular use until several centuries after it had first appeared in Greece. Gold, silver, copper, bronze, grain, and other commodities were used at in-kind moneys, even in places where coinage was available. In many ancient as societies, a combination of both coinage and kind were used. Credit instruments, too, were available in various parts of the Greco-Roman world – but their geographic extent and role in monetary systems remains a contentious issue and the evidence is unclear. Finally, scholars debate whether money acted an arbitrary power which changed Greek and Roman societies and their cultural institutions or whether money was a product of context – that is, whether Greek and Roman money reflects the cultures in which it was exchanged and, hence, took on meanings which had little to do with fiscal management, profit-seeking or otherwise "making money." The Greco-Roman world – in which coinage was first invented and subsequently deployed in experimental fashion – is in fact an ideal testing ground for hypotheses for the impact of monetization on complex societies. There is no question that monetization brought substantial changes to the ancient Mediterranean; at the same time, however, money was not autonomous from the cultures which adopted and used it. The interaction between the overriding cultural, social, and political constraints on the role of money and money's powerful ability to

transform economies makes the monetization of the Greco-Roman Mediterranean one of the most interesting and controversial topics in all of economic history.

Money and Coinage in the Archaic Mediterranean

The Greeks may have invented coinage, but they were by no means the first society to use money. Ideas about money may have travelled to Greece from Egypt. The surviving papyrus records of Pharaonic Egypt (3150–342 B.C.) indicate that Egyptians used fixed measures of copper, silver bullion (*hacksilber*), copper scrap, gold rings, oil, grain, and barley, among other commodities to make payments, store value and keep accounts. While many papyrus contracts discuss physical objects, in some cases commodities could be used to keep track of credits and debts. Egypt's exceptional record is an accident of ecology; the Nile marshes in which the papyrus plant grew were surrounded by arid deserts capable of preserving papyri for thousands of years. Much less is known about the early monetary history of other Mediterranean societies, including those in the Greek-speaking Aegean. Hints in the Homeric epics *The Iliad* and *Odyssey* suggest that cattle were used as a standard of value in Bronze Age Greece. Similarly, inhabitants of archaic Italy may have also used livestock as money (Hollander 2007, 6–7).

A clearer picture emerges once Greek societies began to use precious metal objects as money, as these objects are capable of surviving for thousands of years to confirm their use (although not their meaning). Small but standardized globules of electrum – a naturally occurring alloy of gold, silver, and small amounts of copper – were exchanged for an unknown period of time prior to the invention of coinage in Asia Minor. The earliest known electrum coins, dated to the mid- to late seventh century B.C. and found in Western Asia Minor, were buried with unstamped, but still standardized piece of electrum (van Alfen 2011). Writing some three centuries after the first coins appeared in the Greek-speaking world, Aristotle wrote that the stamp added to electrum globules saved the trouble of weighing bullion. In other words, the stamp did not *create* value, but *communicated* value which had already been given to the objects by those who used them. The process of standardization and verification by an authority, however, unquestionably added economic value to early coins, making them a better money than unstamped metal, primarily by lowering transaction costs. Pieces of precious metal could now be counted rather than weighed. Greeks seems to have been used to making payments, storing value, and accounting transactions in silver bullion; coinage, however, was a more convenient mechanism for carrying out these same sorts of transactions.

Convenience was merely one of many factors which contributed to the development and spread of coinage. Many of the stamps on early coins bear religious and perhaps political significance, and the coins themselves are typically found at sacred sites, implying that early money was a means of embodying obligations towards deities or perhaps elite members of a community, whether individually or at large (Kurke 1999). It seems no accident that early Greek coins are denominated as "spits" (*obeloi*) and "handfuls" of spits (*drachmai*); such objects appear to have been tokens

of animal sacrifice. In Archaic Greece, a sacrifice created an expectation that a deity give something in return, but the sacrificial ceremony itself is redistributive, as all share equally in eating the meat offered. According to Richard Seaford: "it was this traditional, sanctified equality that we saw as a factor in that communal confidence in multiple symbols of identical value that is a prerequisite for the communal adoption of coinage" (Seaford 2004, 294). A sense of obligation or debt to the gods may have therefore aided in the early adoption of coinage.

Existing systems of value and exchange – whether the gift-giving system of elites or the increasingly democratic elements emerging in Greek city states of the late sixth and early fifth centuries B.C. – were stretched to accommodate coined money. The transition appears to have been a messy one, as some early Greek texts imply antagonism between elites and non-elites, the latter of which were said to "not understand or respect the proper workings of gift exchange" (Kurke 1999, 103). Coins were part of a new regime of value which perhaps competed with (and ultimately superseded) aristocratic value systems, into which gift-exchange was embedded. It probably helped that early coins were made of precious metals, as many Greek societies believed that gold was imbued with moral, social, and religious meaning – having been tested and purified. It was the sovereign metal, and its shine was immortal: a metal of gods, kings, and men of virtue. Precious metals were not merely pragmatic, but sacred; their exchange connected human beings with the deities they worshipped and the elites who lead their cities in culturally understood ways.

The spread of coinage through the social strata of Greek societies had a major impact on the geographic spread of coinage in the ancient Mediterranean. Eventually, soldiers and mercenaries, who had previously been paid in kind, were paid in coins (Howgego 1995, 18). Some of the earliest coins in Athens are contemporary with the mid- to late sixth-century B.C. tyrannies. Tyrants minted local stocks of silver into coins – which suggests to historians that uncoined silver had been used as money prior to the invention of coinage – in order to buy mercenaries and spend on public projects. Such expenditures enabled tyrants to obtain and then hold political power. The *Wappenmünzen* or "heraldic coins" of late sixth-century Athens show a range of symbols, and their diversity continues to puzzle numismatists about who minted these early coins and for what purposes. One thing that is clear, however, is that as Athens' form of government changed from tyranny to democracy, coin images transition to civic symbols – first the head of the gorgon (an oblique reference to the goddess Athena) and later the head of Athena herself and her owl. As Athens emerged as the leader of an Aegean naval empire after Persian Wars in the early fifth century B.C., it used the silver obtained via taxes and tribute, as well as from the city's silver mines at Laurion, to mint heavy silver coins (*tetradrachms*) both for domestic and foreign payments. The democratic city-state use coins to pay for magistrates, jurors, the theater (subsidies were given to both those who produced and attended), and even attendance at the democratic assembly (to ensure that the poor could vote). For the Athenians, especially as they became embroiled in a war with Sparta in the second half of the fifth century B.C., the city's coinage became a symbol of its strength. Athens' coins not only bore images signifying civic pride and divine blessing, but the Athenian monetary system was itself a product of political hegemony, military might and abundant wealth.

Athens' political influence, military power as well as the reliability and ubiquity of the *tetradrachm*, made Athens' coinage the dominant money of the Eastern Mediterranean by the middle of the fifth century B.C. The *tetradrachm* carried such great symbolic weight that city-states allied with Athens' rival Sparta just after the Peloponnesian War (431–404 B.C.) insisted upon minting silver coins on an alternative three-*drachm* (*tridrachm*) standard, perhaps as a deliberate rebellion against the prior dominance of the Athenian *tetradrachm* and Athenian imperials more generally.

Despite rapidly evolving monetary practices in the sixth and fifth century B.C. Aegean, the spread of coinage to other parts of the Classical Mediterranean was uneven. Cities which were linked to Greek city-states via language, culture, or trading networks were much more likely to adopt coinage than those which did not share these cultural and economic links. Hence, Syracuse in Sicily and many of the Greek cities in Southern Italy readily adopted coinage. The Syracusians, for example, were some of the first to experiment with token coinages by replacing small silver coins with coins of bronze. Egypt, by contrast, continued to use other media for exchange and accounting, although copies of Athenian silver coins were minted to pay foreign mercenaries. Rome too seems to have issued limited silver coinages on Greek weight standards as early as the late fourth century B.C. These two-*drachm* pieces (*didrachms*) probably were minted to pay the Greek mercenaries who aided Rome in their wars against their Latin-speaking neighbors in central Italy (Crawford 1985, 29). During the third century, Rome's limited silver coins took on a less Greek-flavor, both in iconography and weight standards. Rome's normal domestic currency, however, was weighed bronze ingots and scraps (*aes rude*). Bronze ingots were soon replaced by large, heavy discs (*aes grave*) as well as rectangular bars (*asses signata*). Some of the earliest of these bars were inscribed with miniaturized cattle, leading to speculation that a five-pound bar was meant to represent the value of a healthy adult cow. Being cast rather than struck, the production of these bronze objects must have been limited and, hence, they were also probably used infrequently and for special purposes. Smaller piece of bronze fittingly divided by yet another measure of weight, the Roman ounce (*uncia*), facilitated day-to-day purchases. Production of these smaller denominations was also limited. The Roman state paid soldiers, for example, in weighed bronze (Harl 1996, 27). The "bewildering variety of forms" Roman money took by the third century B.C. – the *aes grave* discs, the *aes signatum* bars, coined silver, and small bronze "ounce" pieces – seem to have had no clear systematic relationship to each other initially, but this may have changed by the end of the third century B.C (Hollander 2007, 2). Thus, Rome managed to cobble together a monetary system out of a hodgepodge of monetary objects, including coins and weighed metal in various forms (Burnett 1987, 13–15).

Money and Conquest in the Hellenistic World and Roman Republic

By the middle of the fourth century B.C., coinage could be found throughout the Mediterranean Basin. Even in places where coinage was less prevalent, such as in Rome and Egypt, the concept of money was known and used. Within less than

two hundred years, however, coinage would become ubiquitous throughout the ancient Mediterranean. What forces were responsible for this change? The role of coinage as a payment for military service was perhaps the most important element in its ascendancy in the ancient Mediterranean.

In the middle of the fourth century B.C., King Philip II of Macedon acquired gold and silver mines in southwestern Thrace. Using this metal to mint vast quantities of coin, Philip hired mercenaries to secure hegemony over the rest of Greece. Through Celtic mercenaries, many of these gold coins (typically referred to as *staters*) made their way north and west to temperate Europe where they were imitated and used by local tribes as gifts and prestige objects. Coinage did not automatically lead to monetization everywhere it spread (de Jersey and Haselgrove 2006; Roymans and Aarts 2009). The British tribes encountered by Julius Caesar during his invasion of Britain in 55–54 B.C. used coins – again, not exactly as *money* – which had been modeled and abstracted from the *staters* of Philip II. Of course, it was Philip's son, Alexander III or Alexander "the Great" as he is more famously known, who not only secured his father's ambitions for Greece but ended up marching his army as far east as India. As Alexander moved through Persia, his generals and officials redistributed the wealth stored in palaces in the form of coined gold and silver (de Callataÿ 2012, 176–180). After Alexander's premature death in 323 B.C., his generals, governors, and soldiers waged war for their late commander's legacy for nearly a quarter of a century, each leader carving out his own individual kingdom out of the territory conquered by Alexander. As a result of the vast amount of soldiers hired in these wars, coined gold and silver – most of it struck in the name of Alexander – flooded the Eastern Mediterranean and beyond. This explosion of gold and silver in the late fourth century B.C. led to Alexander's coinage surpassing the Athenian *tetradrachms* for supremacy.

A desire to associate with Alexander – and hence engender a sense of legitimacy – promoted a conservative approach to coin standards and iconography in the Hellenistic successor states. Eventually, however, the political fragmentation which occurred after Alexander's death aided in the creation of currency zones. Some successor kingdoms used Athenian weight standards for their distinct coinages, but others created their own coin denominations and enforced the use of their coins through legal tender laws. The first and most notable of these closed currency systems was the one created in Egypt under Alexander's general Ptolemy I Soter and his successors. Ptolemy's monetary system was based upon the silver *tetradrachm*, but the ruler reduced the weight of the coin several times during his rule. Nevertheless, he and his successors tariffed these reduced-weight coins at the same value as the higher-quality *tetradrachms* of other Hellenistic polities. Those who entered the territory of the Ptolemies were required to exchange their higher quality foreign coins for lighter Ptolemaic *tetradrachms*, losing roughly 25% of real assets in silver. Why did anyone consent to such a poor exchange rate? Ptolemaic coins were the only legal tender for purchasing surplus Egyptian grain as well as the luxury goods and raw materials from beyond the Eastern Mediterranean which could be found in the cosmopolitan capitol of Alexandria. The Ptolemies were situated in the breadbasket of the eastern Mediterranean, and they shrewdly adjusted their monetary system to take full advantage.

Ptolemaic coinage brought banking, taxation in cash, and a need for currency exchanges to Egypt. Ptolemy II set up state-run banks (*demosiai trapezai*) which collected money taxes. It was also possible for individuals to obtain a banking license in order to hold accounts, provide loans, and change money. Banks, both royal and private, absorbed many of the financial functions of temples due to their importance as treasuries and locations for making payments to the state. In direct opposition to the introduction of coinage in Aegean city states, therefore – where coinage both embodied the overriding cultural-religious order and served the maritime trade of coastal city-states – the introduction of coinage in Egypt was merely one front of a larger cultural invasion. Coinage was Greek, the cultural significance attached to stamped discs of precious metal did not automatically or evenly translate to the whole of Egyptian society, especially in the rural communities in which preexisting alternative forms of exchange were deeply embedded into the rhythms of daily life. Egyptian elites, on the other hand – members of the bureaucracy, priests and any others who wished to show loyalty to the regime – adopted the cultural markers of their Greek overlords, including the use of coinage.

Greek conquests clearly left a lasting impact upon the currency systems of the Eastern Mediterranean and also temperate Europe. For the Western Mediterranean, perhaps the most decisive transition in the monetary economy occurred during the Second Punic War, fought between Carthage and Rome. By the late third century B.C., Rome had not only acquired complete dominance of the Italian Peninsula, but its defeat of Carthage in the First Punic War (264–241 B.C.) and aftermath won Rome control of Sicily, Sardinia, Corsica, the Balearic Islands, and most of the southeastern Iberian Peninsula. Rome's Spanish territories included key precious metal mines. Rome also collected war indemnities from Carthage in silver, enabling the production of silver coins (*quadrigati*) on the Greek *didrachm* standard.

The Carthaginian general Hannibal's unexpected and unprecedented invasion of Italy in 218 B.C. brought a comprehensive crisis to Roman Italy. Rome's hodgepodge monetary system and now restricted supplies of metal simply could not be repurposed to raise the soldiers needed to defeat the ingenuitive Hannibal. Attempting to raise soldiers by paying them in coins, Rome soon had to debase silver *quadrigati* and reduce the weight of bronze *asses*. Rome even released emergency issues of gold coins marked with numerals ("LX," "XL," and "XX") to indicate their value in now token bronze *asses*. One of the major casualties of the Second Punic War was the old Roman monetary system of cast bronze objects (including coins), coined silver on Greek standards and commodity money.

Rome's fortunes in the war pivoted between A.D. 212 and 209, when Syracuse in Sicily as well as many of the Greek cities in the south of Italy were re-captured by Rome. The reversal of fortune brought in fresh supplies of silver, allowing the Roman state to mint a new coin of pure silver called the *denarius* ("piece of ten"). The *denarius* was connected directly to the Roman bronze coinage with a mark of "X" to indicate its value of ten bronze *asses*. Most other silver coins were also given value marks in *asses*. The pure silver *denarius* and value marks across the other denominations unquestionably promoted confidence in the new system. The gold coins were soon abandoned and the silver *denarius* became the bedrock of a new Roman monetary system.

As the Second Punic War wound down, the *denarius* system quickly came to dominate Italy and Sicily, pushing the various Greek silver coinages out of circulation. The spread of the new monetary system seems to have been partly due to the military mints set up in Italy, Sicily, and Sardinia, which would have overwhelmed these areas with coins on the Roman standard. As soldiers returned to their villages with coins and a new appreciation for the use of coined money, a broader education of the Roman population soon generated new demand for coined money (Hollander 2005). Even the monetized cities of Southern Italy with their long tradition of Greek coinage standards abandoned these systems and began using *denarii* and *asses*. The new Roman system was, after all, somewhat familiar to Greeks: a main precious metal denomination (the *denarius* in this case) was connected to smaller token units of bronze (Von Reden 2010, 51–52). The Romans, perhaps unintentionally, had stumbled upon a durable and coherent monetary system.

The growth of cities in Italy and their demand for surplus goods produced in the countryside may have contributed to the spread of Roman coinage. After the Second Punic War, small Italian villas produced cereals, olives, and wine for markets in Rome and other cities (Roselaar 2010, 164). By the end of the second century B.C., some villas operated as cash-crop-producing enterprises which generated high incomes for their distant owners. Urban markets provided an outlet for these goods, promoting a cash economy in cities (Kay 2014, 108). Roman agronomists made sure to suggest that villas be located next to good roads, harbors, or cities in order to ensure that produce could be sold in urban markets. Government officials recognized the usefulness of markets and responded by constructing permanent infrastructure in cities as well as providing basic rules and enforcement mechanisms to regulate and support markets. Ironically, Rome, a late adopter of Greek-style coinage as the basis for monetary transactions, ended up crafting a coinage system, which would eventually come to dominate all the others in the centuries to follow.

Markets and Monetization

Curiously, the spread of Roman coinage slowed considerably after an initial burst during and immediately following the Second Punic War. The slowdown is all the more remarkable considering that Roman military expansion was essentially uninterrupted from the early second century through the late first century B.C. Why did military conquest, the force which seems to have been so crucial to monetization, not continue to spread Roman coinage in tandem with the annexations of the ascendant Roman Empire? In fact, several Hellenistic coinage systems in the Eastern Mediterranean endured well into the third century A.D. – centuries after these territories came under direct Roman rule. The various entrenched civic and religious associations among the Greek cities, and their coinages may partly explain the persistence of Greek money. Also, however, the Romans themselves thwarted the expansion of their own monetary system, preferring to leave existing coinage systems in place to ensure that taxes, tributes, and other rents continued to be collected without interruption. Even Rome's bitter enemy Carthage, for example, was allowed to continue minting gold

and silver coinage (shipping a great deal of it to Rome in the form of a heavy war indemnity) after the Second Punic War until the city was destroyed by Rome in 146 B.C. Finally, despite the newfound coherence of the Roman monetary system and its emergence in a military setting, soldiers still regularly received bonuses, rewards, and discharge payments in kind, including parcels of land (Hollander 2007, 99).

Just as with coinage, the Romans were also late adopters of markets. Many Greek city states had at their center an *agora* which, although initially used as a meeting place for political assemblies, became a commercialized space for making transactions by the Archaic period (Hall 2007, 47–48). The *agora* in Athens, for example, seems to have taken on its commercial functions around the same time as the city adopted coinage (Schaps 2003, 113). Market supervisors (*Agoranomoi*) were put in charge to inspect the quality of goods sold, provide just weights and measures, and ensure that any fixed prices were adhered to. Coinage, especially the smaller denominations, brought poorer people into the *agora*, allowing them to make day-to-day transactions on a somewhat anonymous basis with each other. A character in Aristophanes' play "Women at the Festival" (*Thesmophoriazousae*), for example, is a poor widowed mother who sells garlands in the market for money. The advent of coinage thus provided new opportunities for those outside of the reciprocity-oriented landed male elite to use urban markets in both Greek and Roman cities to better their lives.

The link between monetization and markets in the Greco-Roman world seems apparent, but much about this relationship is unknown and contested. While local markets were clearly abundant in urban settings, did monetization strengthen the connections between markets and stimulate economic integration? Seminal articles by Keith Hopkins (1980) and Chris Howgego (1992, 1994) paint a picture of a classical world in which the use of money was nearly ubiquitous by the apogee of Roman power in the second century A.D. It is now uncontroversial to argue that money permeated the ancient world and was understood in cultural terms as a marker of civilization (Lo Cascio 2007, 627–628). Arguments for high levels of monetization in the Roman world are bolstered by MIT economist Peter Temin's boldly titled *The Roman Market Economy* (2013), which uses economic theory to demonstrate the Roman Empire's regional specialization, division of labor, long-distance trade, economic prosperity and "universal" monetary integration. Temin's use of regression analyses shows a link between distance from Rome and grain prices – a novelty among ancient historians, some of whom subscribe to the idea that economic theory has little to offer to them. Not everyone, however, is convinced by arguments for an integrated monetary system in the ancient Mediterranean. Jean-Jacques Aubert, for example, portrays a Roman Empire in which barter and commodity money were commonplace and were perceived as more desirable than coinage (Aubert 2014, 110–111). Much of what Aubert argues is confirmed in the archaeological record of north-western Europe. While coins were available, they were not always used as money and even thought the Romans minted an unprecedented amount of coinage, it may not have been enough to make monetized market exchange a universal feature of the Empire. Sitta von Reden in her *Money in Classical Antiquity* (2010) shows that "there was no continuous expansion toward a single monetary zone" in the whole of Classical Antiquity (Von Reden 2010, 16). Instead, changes in "monetary

networks" – a slightly fuzzy term which encompasses not only the weight, fineness and iconography of coins but their political and ideological contexts – were intimately connected with changes in the political landscape of the ancient world. Arguments such as von Reden's emphasize the significance of non-market factors while also avoiding dogmatic thinking about the total absence of markets.

Under Roman rule, markets do seem to thrive, but the link with monetization is less clear as it is in the Greek world. Instead of market forces supplying the impetus for the spread of Roman money, it seems that political and military forces were the main agent. It was not until the early first century A.D., for example, that the Greek city of Thessaly was required to recon their tax payments in Roman *denarii*. The law, promulgated under Rome's first emperor Augustus, was not so much an effort to force Roman coinage and culture onto locals or to stimulate the spread of Roman currency; it was instead an practical measure for administering the newly separated province of Achaea and the Roman colonists who had taken up residence there. Ramsay MacMullen argues that the most important reason for Roman laws establishing the use of Roman monetary units, including the law of Augustus, was to ensure the easy conversion of local taxes into supplies and payments for Roman soldiers (MacMullen 2000, 6–7). The physical, circulating coinages of eastern Roman provinces even after such decrees continued to be locally produced denominations in their own weight and fineness standards.

In some ways, Rome's approach to the monetization of its empire was similar to that of the pre-Punic War republic writ large: a hodgepodge of monetary systems and coinages, some of which had conversion rates. Several scattered hints in the source material, such as an early third century A.D. letter from a soldier to his sister (*P. Meyer* 20) in which he mentions that thirty *denarii* are worth 120 *drachms*, help scholars quantify coin exchange rates. Most scholars believe that exchange rates between the various currencies in the Roman sphere of influence were legally enforced, regardless of whether the exchange rates arose organically by convention or were part of some official design. Some provinces, the most notable being Egypt once again, were allowed to continue their pre-Roman practices of enforcing closed currency zones, in which any foreign coins (including those of Rome!) were to be exchanged for debased *tetradrachms*. Silver *cistophori*, also underweight compared to the Attic *tetradrachm*, were minted in Ephesus and Pergamum through the mid-second century A.D. While a few polities continued to mint silver or billon (alloys with a minority silver component) coins, most local coinages were copper or copper alloys. The explanation for the prevalence of these token coins could be economic: perhaps they were minted to act as small change for the supply of local markets. Alternatively, the right to mint coinage had long been a sign of authority in eastern Mediterranean societies, and it makes sense that elites would continue the practice as a means of signaling their position in the local hierarchy. At the same time, by emblazoning local coinage with iconography which glorified Rome and her emperors, these locals made it clear that their authority was not to be conflated with insubordination or rebellion. Despite the inconvenience caused by having such disparate monetary standards across the empire, imperial authorities may have allowed or even appreciated such voluntary gestures of supplication and submission.

While evidence from the Greco-Roman world indicates clear historical correlations between monetization and the development of markets, such results seem to have been largely accidental rather than the deliberate aims of coherent "monetary policies." Peter Bang in his *The Roman Bazaar* (2008) argues forcefully that the economy of the Roman world was characterized by slow and disconnected transport (by both sea and land), out-of-date price information on unequilibrated markets, unstandardized goods, small-scale and mobile merchants, and, finally, risk-minimizing trade via social networks. The inherent fragility of the system limited economic growth while, at the same time, brought flexibility and resilience in the face of arbitrary state power, especially at the local level. The contradictions and complexity of this model challenges modern scholars, especially those accustomed to thinking about Rome as part of the narrative of European economic development, and demands that the Roman economy be situated as a unique historical phenomenon.

It is difficult to imagine any ancient state, including the Roman Empire, having much if any "monetary policy." Scott Meikle argues that "the institutions and relations of exchange-value were peripheral to ancient society, not central and dominating as they have become in modernity" – a statement which, if true, suggests that accurate prices would have been difficult to find in the ancient world (Meikle 2002, 247). How, therefore, would an "economically rational" government, whether at the imperial or municipal level, be able to regulate the money supply without reliable economic information? Ancient societies, if they even thought to institute monetary policies, would have had to overcome tremendous institutional obstacles to do so. These obstacles should be enough to confound suggestions that the Roman state, even at its apogee, engaged in any kind of statistically informed monetary policy, even to preserve exchange rates between denominations (Lo Cascio 1981). In order to engage in such a task, there would need to have been current, available reports on the number of coins circulating of each denomination, the precious metal content and weight in each issue, the current market prices of precious metals, and a host of other data which even modern states sometimes cannot accurately quantify.

The Roman coinage system which came to dominate the Mediterranean developed in an *ad hoc* fashion in order to solve practical political and military problems. The system was ultimately managed within a household economy model – perhaps the largest household economy in human history. Ancient money was probably managed along lines similar to other operations in the ancient world: through webs of favors and effective ties of obedience. There were many formal regulations, rules, and laws related to the use of money and market activity. What are we to make of these? Many studies of ancient marketplaces, both urban and rural, suggest that the main tasks of market officials was "law and order" rather than fair prices. Roman magistrates and overseers routinely show fear of any unofficial gathering, assembly or association of locals (Shaw 1981, 47). The emperor Trajan prohibited one of his governors from establishing a fire brigade in the city of Nicomedia in northwest Asia Minor as the emperor feared that such a group could promote political insurrection. As long as marketplaces were not associated with outright rioting and rebellion, local officials at best dispassionately pursued their duties or, perhaps more often, intentionally abused their authority for personal gain.

On rare occasions, emperors themselves manipulated the monetary system, such as the emperor Tiberius' intervention of A.D. 33. Newly revived laws fixing interest rates and restricting land-ownership led to a crash in the Italian real estate market and the Roman economy became plagued by deflation. In response to the money shortage, Tiberius put together a state-backed rescue package of loans via the Roman banking system to stabilize the price of land, restore credit, and ensure that the fortunes of heavily leveraged elites remained mostly intact. Many scholars have interpreted Tiberius' actions as evidence of sophisticated macroeconomic management of the Roman monetary system (Elliott 2015). While financial motivations were clearly important, the overriding reality is that Tiberius ultimately rescued the social capital of Roman elites (Crawford 1970, 46). Land ownership was an inflexible requirement for elites if they wished to maintain their membership in Rome's senatorial class.

It seems that the intended roles of money in the Roman Empire were therefore almost entirely practical: paying troops, exercising power, and displaying status. Nevertheless, intentionally or not, the Romans managed to create a somewhat interconnected monetary system in the process. Administering this system, however, turned out to be complicated if not eventually impossible for emperors and state officials. The system's eventual collapse could be interpreted as a sign that Roman officials simply did not have at their disposal the required theoretical knowledge or practical mechanisms to manage their unwieldy monetary system.

The Value of Money in the High Roman Empire

Rome's first emperor, Augustus, codified the essence of a monetary system which would last until the late third century A.D. with few modifications. First, the system was tri-metallic: gold coins (*aurei*) were worth 25 silver *denarii* which could in turn be split into copper or copper-alloyed denominations for small change. The introduction of a brass *sestertius* – worth one quarter of a *denarius* – was a novelty to the Roman system, a numismatic practice which was borrowed (as so many other Roman monetary adoptions) from the Hellenistic east (Wallace-Hadrill 1986, 81). This is not to say that the system was stable; the presence of four different metals and alloys exchanging at fixed values created a great deal of instability. Nominal exchange rates probably held as long as the market prices of the various metals did not fluctuate too greatly, assuming of course that the coins themselves maintained standardized weights and finenesses. The gold *aureus* indeed changed only a little during the first two centuries of the Roman Empire. The coin maintained its purity and only diminished in weight by about one tenth (Duncan-Jones 1994, 216). During the same time, the brass *sestertius* lost its zinc content and soon became a bronze coin. The copper *as* lost one or two tenths of its weight, but coin wear and tear makes it impossible for numismatists to establish weight standards with any kind of exactitude. The smallest denominations below the *as*, the *semis* (1/2 *as*), and *quadrans* (1/4 *as*) dropped out of production (Corbier 2005, 332). One might speculate that prices may have risen overtime and invalidated the necessity of such

small denominations. Alternatively, it may be that low-denomination coins were never much in demand in the first place due to low levels of monetization.

The silver *denarius*, however, shows a more troubling trend during the first and second centuries A.D. Although only losing about one seventh of its weight, around half of its silver content vanished by the A.D. 190s. Nero's reign in the mid first century A.D. witnessed the first substantial debasement of the silver *denarius* – a removal of around 20% of the coin's silver. Interestingly, Nero's moneyers went to great lengths to hide their adulterations by enriching the surface of coins (Butcher and Ponting 2005). Recent destructive analyses of Neronian coins have revealed copper-alloy cores inside a silvered exterior. Moneyers apparently feared that coin users would only accept a limited fraction of fiduciary value in silver coins. *Denarius* debasements grew more pronounced over time; yet apart from a decade or so in the late second century, the size and weight (3.36 g) of *denarii* were strictly maintained to preserve appearances. By the end of the second century A.D., around half the silver was being quietly removed from newly minted *denarii*. At a 50:50 ratio of silver to copper, sophisticated metallurgical techniques were required to draw silver to the surface of coins and thus hide any tell-tale pink-brown hue.

A political and financial catch-22 in the early third century A.D. forced a significant change in the system. After murdering his brother and co-emperor Geta, the emperor Caracalla promised his soldiers double their normal salary "in return for the sole rule," as the contemporary historian Herodian relates (Herodian 4.4.7). At the same time, the stocks of precious metal in the imperial treasury had been departing at regular intervals over the previous half-century, if not before, into the hands of potentially hostile neighbors on the outskirts of the Roman *limes* (Howgego 1992, 5–8). By means of such payments, Roman officials hoped to convince migrants and raiders that peace was more attractive than plunder. Finally, it appears that Roman silver mines in northwest Spain were failing to produce sufficient ore (Wilson 2007, 113). These combined problems may have forced Caracalla's moneyers to take the novel step in A.D. 215 of introducing a new coin denomination, the first in over two centuries. The *antoninianus* (or "radiate," as some numismatists prefer to call it, primarily due to the consistent image of an emperor wearing the radiate crown of the deity Sol Invictus on the obverse) was a large, half-silver coin widely believed to have been worth two silver *denarii*. Caracalla's *antoninianus* seemingly allowed the emperor to fulfill his promise of higher pay to the troops while also mitigating the depletion of silver from imperial coffers. The metrology of the coins shows this in clear terms: each of Caracalla's *antoniniani* includes roughly 2.65 g of silver while his *denarii* contain about 1.66 g (Harl 1996, 127–130). Thus, the average silver content in each *antoninianus* is less than one and a half times the silver contained in each *denarius*. Hence, the *antoninianus* was a "double" *denarius* in name only. Despite the fact that moneyers managed to keep the coin at half-silver, and thus avoided exposing their manipulation to the naked eye, the new denomination flopped anyway. After circulating a mere four years, Elegabalus discontinued production of the *antoninianus* in A.D. 219. Imperial mints resumed minting the more trusted *denarius*; however, these new *denarii* contained about 20% less silver compared to those which had been

issued just four years earlier. The bedrock of the Roman monetary system was plunged below a threshold from which it would never recover.

The *denarius* continued to decline in quality through the middle of the third century A.D. Trust in the coin had been built up over centuries by reasonably high standards or, whenever these standards faltered, cleverly devised deceptions. After Caracalla's reign, however, it took only a few decades of naked debasements to destroy coin-users' faith. Finally, in the chaotic year of A.D. 238 – a year of nearly uninterrupted civil war which saw six (or perhaps seven) different emperors claim the purple, desperation prevailed and authorities revived the *antoninianus*. This time, officials eliminated competing denominations and the *denarius* was effectively abandoned. Coin hoards show that coin-users withdrew *denarii* from circulation and instead used the over-valued *antoniniani* for market transactions (Bland 1996). The *antoninianus*, like its predecessor, was continuously and rapidly debased, losing virtually all of its silver in a decline which took less than two decades to occur. From A.D. 238, not only was fineness reduced, but overall weight was allowed to drop as well. This time, silver debasement was not hidden from coin users (Elliott 2014). The debasement of the *antoninianus* tested the extent to which fiduciary value could command coin-user's trust.

Modern scholars continue to debate whether money during this period was valued based upon metal content or fiduciarity. Part of the reason why there is no clear answer to this question is that the stated value and precious metal content of coins were nearly synchronous during the first two centuries A.D. Coin users may have complied with legal exchange rates in spite of state power, not because of it, simply because transaction costs were too high. The minor differences in metal content and official value probably did not justify hoarding, melting, charging a premium, or other hedging behavior. The imposition of an *agio* – a commission on currency exchanges paid to money-changers, local officials or, as was often the case, both – is an obvious transaction cost; moreover, the fee's quantitative nature allowed it to be easily factored into the choices of ancient currency users. Although they are far more difficult to quantify, there were certainly costs associated with melting coin into bullion. Even hoarding, especially when one wished to invest or spend money, carried real costs beyond mere opportunity costs. There is also the question of what one did with melted metal, as re-coining may have been inconvenient and would have required its own fees. Information asymmetry was also an issue, as minting authorities and privileged state payees, who knew that the coinage had been debased or reduced in weight, could use this knowledge to their advantage until coins had been properly assayed on markets. Such barriers – and the list here is certainly not comprehensive – would have kept exchange rates between coins more stable than they might have otherwise been on a "frictionless" market in which there were no costs to melting, hoarding or exchanging currency.

Transitions in the Role and Supply of Money in Late Antiquity

The mid-third century A.D. brought a series of blows to the authority and legitimacy of the Roman Empire: plague in the A.D. 250s, the abduction and murder of the emperor Valerian in A.D. 260, the fragmentation of the empire into separatist states,

numerous foreign incursions and invasions, and domestic brigandage. Troops regularly mutinied and martial disloyalty was a hallmark of the third century A.D. The fidelity of armies was a central concern of Roman generals and emperors. Soldiers found it easy to defect to rival generals for financial reasons – a rival offered them more money or their current general reneged on promises of payment. Coinage therefore *had* to be acceptable to soldiers, not merely "the public" at large. A few scattered monetary regulations survive, however, which give the impression that the currency system was not commanding acceptance.

An inscription of A.D. 210 from the city of Mylasa in Caria (southwestern Asia Minor) tells how the city enacted a law to punish rouge money changers (including slaves and freedmen) from doing business on what was in effect a black-market currency exchange (*OGIS* 515). In the early third century, two categories of money changers operated in Mylasa. First, a single bank possessed both a legal right to exchange and a contract granting the privilege of monopoly – the former item permitted its existence, while the latter secured its profitability against competitors. Second, in spite of the monopoly privilege of the establishment bank, independent money changers also operated in the city, although they lacked the stability of state-contracted income as well as official legality. Functionally, the privileged bank was an arm of the municipal authorities, yet it was privately run for profit. By contrast, despite their illegality, independent money changers provided a service in the voluntary sector of the economy and were required to meet the increasingly demanding (as coins became more erratic) needs of coin users if they expected to continue to sustain themselves. The substantial shifts in the metal value of silver *denarii* which occurred at the end of the second century A.D. placed a premium upon the services of currency specialists – a premium so high that it was worth the risk for such specialists and their customers to operate in black of grey markets (Elliott 2014, 142–149). Unfortunately, the inherently subversive and secretive nature of black-market merchants makes the prospect of finding more concrete evidence for their existence nearly impossible, except when official sources complain about unlawful market activity or attempt (almost always futilely) to eradicate black-market competition, such as happened in this case.

Several decades later (approximately A.D. 260) in Oxyrhynchus, Egypt, monopoly currency exchanges were combating additional troubles. Even with favored status and the force of the state behind them, conditions had deteriorated and contracted exchanges were unable to remain open. The mutually advantageous relationship between the Roman state and official currency exchanges had degenerated into a one-sided affair, as authorities eventually chose to take severe measures to ensure the success of local (if not imperial) debasement policy at the expense of their collaborators. Clearly, laws which commanded acceptance of official coinage had come into being by the third century A.D. at the latest. Earlier laws, such as the first-century B.C. *Lex Cornelia de Falsis*, however, merely protected the official coinage from poor quality imitations and counterfeits, whereas currency laws in the third century seem have sought the opposite: protections for the low-quality official coinage obtained by restricting the ability of private assays and exchange institutions from revealing the diminished quality of imperial coinage. The

more obvious conclusion from the A.D. 260 law, however, is that professional banks were apparently closing down, or perhaps may have been shifting their operations underground to take advantage of more profitable opportunities in a comparatively less regulated black market in currency exchange and assay. There is a comparable precedent for such behavior in both the late second century and the early fourth century. When the emperor Commodus enacted price controls in the wake of famine in A.D. 189/90, shortages of grain were compounded, as suppliers withdrew from the open market in favor of better prices on the black market. More famously, price controls in the early fourth century A.D. led to shortages of all kinds of goods. Finally, when the last pagan emperor Julian attempted to control grain prices in Antioch in the A.D. 360s, vendors fled the city to purchase grain on black markets in the countryside.

By the reign of Constantine in the early fourth century A.D., it was a capital crime not only to counterfeit coins, but also to mint or exchange coins without official sanction as well as to patronize such operations. Such laws on currency required enforcement, presumably by *aediles* and *agoranomoi*, who seem to have oversaw prices in urban areas. It is even more difficult to imagine that rural transactions were policed with any kind of consistency, except perhaps on periodic and officially permitted "market days" (*nundinae*). Furthermore, exceptions to official regulation, such as is witnessed in a law of A.D. 396 allowing Jews in Palestine to appoint their own market officials should they so desire, are extant. Again, while state infrastructure certainly existed which might regulate currency ratios, it is difficult to argue that monetary regulations were of major concern to market officials or regularly enforced.

In the A.D. 270s, after decades of mistrust among coin-users, and especially soldiers, the emperor Aurelian reformed the coinage and returned the monetary system to a precious metal standard. The metal content of the heavily debased *antoninianus*, now the only "silver" coin in the system, was given a value mark of "XX·I" or "twenty to one" – that is, twenty parts copper to one part silver (Haklai-Rotenberg 2011). The weight of the gold *aureus*, which had become erratic and unreliable in the third century A.D., was standardized and restored to the weight and purity it had been at the beginning of the century. As with the silver coinage, the gold coinage was also marked to indicate its metal content; coins were stamped "I·L" or "one equals fifty" to show that 50 gold coins was equal to one pound of gold bullion.

The effect of these reforms on markets was probably unforeseen. Concurrent with Aurelian's reform was a sudden and severe (tenfold!) jump in prices (Rathbone 1996, 1997). The debasements of subsequent rulers unraveled Aurelian's initial efforts. Also, the stock of precious metals had been drained after more than a century of outbound state payments, export of precious metals due to Gresham's Law and the fact that many Roman silver mines appear to have been thoroughly worked out must not have helped matters either. Aurelian had little gold and only enough silver to raise his "silver" standard to a fractional silver coinage instead of something more robust. Although he fully disclosed this fractional standard to coin-users, he also inherited a legacy of deception and manipulation. Coin-users were almost certainly cynical of his reforms, and, by this time, they were probably trading in commodities

for large transactions and using *antoniniani* and copper coinage for day-to-day transactions. Only a total systematic reform might entice money-users to return to using the official imperial coinage.

Such a reform occurred just two decades after Aurelian's death when monetary officials under the emperor Diocletian scrapped the Augustan system and began afresh. The new system brought the return of a coin of pure silver, crassly named the *argenteus*. The new coin was in effect a silver *denarius*, although its new name suggests that the previous centuries of debasement had destroyed the once trusted *denarius* brand. A restored gold *aureus* was also given a new name which reflected the martial nature of the regime: *solidus* – soldier's slang meaning something like "solid bit." The new gold coin showed up for the first time in Diocletian's *Price Edict* of A.D.301 – a set of maximum prices for goods and services which was meant to curb price inflation. The *Edict* also treated gold, whether coined or not, as a commodity rather than a true currency (Hendy 1985, 451). A new set of token bronze denominations completed the reform. The regional currencies of the Roman monetary system, many of which were centuries older than the oldest Roman coin denominations, were finally abandoned and the new system was used "for the whole world," as the *Currency Edict* posted in the Greek city of Aphrodisias in Asia Minor in the early fourth century A.D. declares.

The monetary system engineered chiefly under the emperors Diocletian and later Constantine represents the final stage of classical coinage, although the constituent parts of this system fared differently over time. The silver coinage all but disappeared by the A.D. 310s (Abdy 2012, 590) before subsequent silver denominations were introduced by later emperors – the most important of these being the light silver coin introduced by Constantine around A.D. 325 which numismatists often called the *siliqua*. A lighter version of the *siliqua* saw regular use in Western Europe even after the fall of the Western Roman Empire. Silver *siliquae* appear in such wide ranges of weights during the second half of the fourth century that they may have simply been traded by weight rather than their nominal value (Guest 2005, 42). The gold *solidus* also endured in the west for centuries, and it formed the etymological basis for the French *sou* and English words like "soldier." Gold, both in coined and uncoiled form, became the anchor for exchange throughout much of Europe, Asian Minor, and the Near East.

At the beginning of the fourth century, small change in the form of several bronze coinage denominations appears to have spread widely. A cycle soon emerged: bronze coins were reduced in weight and subsequently replaced with higher denomination coins which would themselves be reduced in weight before the process was repeated again (Harl 1996, 162). The result must have been price inflation as well as mistrust of the coins. One solution to the problem was the use in both exchange and accounting of sealed purses of bronze coins (*folles*) in fixed numbers. As the bronze currency rotated, older coins were officially demonetized by law and were forbidden for use in transactions. By the fifth century, most bronze coins were minuscule in diameter (~9 mm) and light (~1 g), requiring high numbers of them (~2400) to make the next largest coin denomination, the gold one-third *solidus* piece known as the *tremissis*.

The area in which coinage was used contracted as the Western Empire's influence waned (Ward-Perkins 2005, 113). Numerous strategies for dealing with this problem were adopted. Coin clipping was rampant in early fifth century Britain, for example (Walton and Moorhead 2016). Imperial authorities in the fourth century promulgated a variety of laws to discourage such activities in their provinces, even making coin-melting and clipping a capital offense. Rural inhabitants seem to have primarily used coins to secure and store value rather than for monetary exchange. In areas where coinage continued to serve as money, weight standards and nominal values differed, despite the near ubiquity of the gold *solidus*. The bronze coins of Rome and Ravenna, for example, had slightly different nominal values compared to those minted in Carthage (Stahl 2012, 636). Money-use became decentralized and localized, and so it is impossible to create a single narrative of monetary transition in Late Antiquity. Coinage on nominal standards probably still dominated in cities, while the metal value of coinage may have held more value in remote and even rural areas. The military continued to requisition many goods directly rather than using coinage as a medium and land-owners practiced share-cropping, limiting the necessity of coinage (Hendy 1985, 301–303). As Late Antiquity transformed into the Middle Ages, acceptable media of exchange – whether coins, commodities or other moneys – were dictated by local customs.

Conclusion

Although money preceded coinage by many centuries, the invention of coinage brought about major changes in the Greek and Roman worlds. Because coinage first developed within the context of Greek city states, certain characteristics such as the importance of precious metals and religious and civil iconography remained foundational in ancient monetary systems. Coin-use was promoted by conquest, especially in the Greek and Hellenistic era. The Romans, however, during their ascent to Mediterranean dominance were careful to preserve existing coinage systems, preferring instead to make pragmatic alterations to promote tax flows and public order. Money also promoted the growth of markets, allowing strangers to trade with each other. Still, even with the adoption of money, markets retained a predominantly local function. Autarky was the dominant mode of living among the vast majority of ancient societies. Many of the transactions which did occur, even at the height of the monetization of the ancient world in the second century A.D., were probably performed on the basis of reciprocity or barter. Even the Roman state never completely transitioned to a fully cash-based taxation system. Money-use, therefore, remained limited, especially in rural areas. In other words, coinage probably functioned as a "general purpose" money, but only in specific areas and contexts. Meanwhile, "special purpose" money (grain, wine, bullion and other commodities) was used alongside coinage for certain transactions. The association of money with commodities among Greek and Roman populations made it impossible for states to sustain debasements of precious metal coinages for too long or without going to great lengths to hide such manipulations. Even the total political and military

dominance enjoyed by the Roman Empire did not protect its monetary system, which fragmented into a diversity of regionalized systems by the Late Antique period.

References

Abdy R (2012) Tetrarchy and the house of Constantine. In: Metcalf WE (ed) The Oxford handbook of Greek and Roman coinage. Oxford University Press, Oxford, pp 584–600

Aubert J-J (2014) For swap or sale? The Roman Law of Barter. In: Apicella C, Haack ML, Lerouxel F (eds) Les affaires de Monsieur Andreau. Economie et société du monde romain. Ausonius, Bordeaux, pp 109–121

Bland R (1996) Bland the development of gold and silver coin denominations, AD 193–253. In: King CE, Wigg DG (eds) Coin finds and coin use in the Roman world. Gebrüder Mann Verlag, Berlin, pp 64–100

Burnett A (1987) Coinage in the Roman world. Spink, London

Butcher K, Ponting M (2005) The Roman denarius under the Julio-Claudian emperors: mints, metallurgy and technology. Oxf J Archaeol 24:163–197

Corbier M (2005) Coinage and taxation: the state's point of view, A.D. 193–337. In: Bowman AK, Cameron A, Garnsey P (eds) The Cambridge ancient history, 2nd edn. Cambridge University Press, Cambridge, pp 327–392

Crawford MH (1970) Money and exchange in the Roman world. J Roman Stud 60:40–48

Crawford MH (1985) Coinage and money under the Roman republic: Italy and the Mediterranean economy. University of California Press, Berkeley

de Callataÿ F (2012) Royal Hellenistic coinages: from Alexander to Mithradates. In: Metcalf WE (ed) The Oxford handbook of Greek and Roman coinage. Oxford University Press, Oxford, pp 175–190

Jersey Pd, Haselgrove C (ed) (2006) Early Potin coinage in Britain: an update. In: Celtic coinage: new discoveries, new discussion. Archaeopress, Oxford, pp 17–28

Duncan-Jones R (1994) Money and government in the Roman Empire. Cambridge University Press, Cambridge

Elliott CP (2014) The acceptance and value of Roman silver coinage in the second and third centuries AD. Numis Chron 174:129–152

Elliott CP (2015) The crisis of A.D. 33: past and present. J Ancient Hist 3:267–281. https://doi.org/10.1515/jah-2015-0006

Guest PSW (2005) The late Roman gold and silver coins from the Hoxne treasure. British Museum Press, London

Haklai-Rotenberg M (2011) Aurelian's monetary reform. Between debasement and public trust. Chiron 41:1–39

Hall JM (2007) Polis, community, and ethnic identity. In: Shapiro HA (ed) The Cambridge companion to archaic Greece. Cambridge University Press, Cambridge, pp 40–60

Harl KW (1996) Coinage in the Roman economy, 300 B.C. to A.D. 700. Johns Hopkins University Press, Baltimore

Hendy MF (1985) Studies in the byzantine monetary economy C.300–1450. Cambridge University Press, Cambridge

Hollander DB (2005) Veterans, agriculture, and monetization in the late Roman Republic. In: Varhelyi Z, Aubert J-J (eds) A tall order: writing the social history of the ancient world. Essays in honor of William V. Harris. K. G. Saur, Munich, pp 229–239

Hollander DB (2007) Money in the late Roman Republic. Brill, Leiden

Hopkins K (1980) Taxes and Trade in the Roman Empire (200 B.C. - A.D. 400). J Roman Stud 70:101–125

Howgego C (1992) The supply and use of money in the Roman world 200 B.C. to A.D. 300. J Roman Stud 82:1–31. https://doi.org/10.2307/301282
Howgego C (1994) Coin Circulation and the integration of the Roman economy. J Roman Archaeol 7:5–21
Howgego C (1995) Ancient history from coins. Routledge, London
Kay P (2014) Rome's economic revolution. Oxford University Press, Oxford
Kurke L (1999) Coins, bodies, games, and gold: the politics of meaning in archaic Greece. Princeton University Press, Princeton
Lo Cascio E (1981) State and coinage in the late republic and early empire. J Roman Stud 71:76–86. https://doi.org/10.2307/299498
Lo Cascio E (2007) The early Roman Empire: the state and the economy. In: Scheidel W, Morris I, Saller RP (eds) The Cambridge economic history of the Greco-Roman world. Cambridge University Press, Cambridge, pp 619–647
MacMullen R (2000) Romanization in the time of Augustus. Yale University Press, New Haven
Meikle S (2002) In: Scheidel W, Von Reden S (eds) Modernism, economics and the ancient economy. Routledge, London, pp 233–250
Rathbone DW (1996) Monetisation, not price inflation, in third century A.D. Egypt? In: King CE, Wigg DG (eds) Coin finds and coin use in the Roman world. Gebrüder Mann Verlag, Berlin, pp 321–339
Rathbone DW (1997) Prices and price formation in Roman Egypt. In: Andreau J, Briant P, Descat R (eds) Économie antique. Prix et formation des prix dans les économies antiques. Musée Archéologique, Saint-Bertrand-de-Comminges, pp 183–244
Roselaar ST (2010) Public land in the Roman Republic: a social and economic history of ager publicus in Italy, 396-89 BC. Oxford University Press, London
Roymans N, Aarts J (2009) Coin use in the dynamic frontier region. Late Iron Age coinages in the lower Rhine area. J Archaeol Low Countries 1:5–26
Schaps D (2003) The invention of coinage and the monetization of ancient Greece. University of Michigan Press, Ann Arbor
Seaford R (2004) Money and the early Greek mind: homer, philosophy, tragedy. Cambridge University Press, Cambridge
Shaw BD (1981) Rural markets in North Africa and the political economy of the Roman Empire. Antiq Afr 17:37–83. https://doi.org/10.3406/antaf.1981.1072
Stahl AM (2012) The transformation of the west. In: Metcalf WE (ed) The Oxford handbook of Greek and Roman coinage. Oxford University Press, Oxford, pp 633–654
van Alfen PG (2011) Mechanisms for the imitations of Athenian coinage: Dekeleia and mercenaries reconsidered. Revue belge de numis 147:55–93
Von Reden S (2010) Money in classical antiquity. Cambridge University Press, Cambridge
Wallace-Hadrill A (1986) Image and authority in the coinage of Augustus. J Roman Stud 76:66–87
Walton P, Moorhead S (2016) Coinage and the economy. In: Millett M, Revell L, Moore A (eds) The Oxford handbook of Roman Britain. Oxford University Press, Oxford, pp 834–849
Ward-Perkins B (2005) The fall of Rome: and the end of civilization. Oxford University Press, Oxford
Wilson A (2007) The metal supply of the Roman Empire. In: Papi E, Bonifay M (eds) Supplying Rome and the Empire: the proceedings of an international seminar held at Siena-Certosa di Pontignano on May 2–4, 2004, on Rome, the provinces, production and distribution. Journal of Roman Archaeology, Portsmouth, pp 109–125

Primitive and Nonmetallic Money

4

Bill Maurer

Contents

Abstract

Feathers, beads, shells, copper bracelets, and giant stones – objects that Western observers have assumed serve the functions of money in so-called simple societies and other non-Western contexts – come in all shapes and sizes. This chapter reviews the literature on "primitive" currencies, from early ethnology to contemporary anthropology and archeology. Showing how analysts frequently misunderstood the use of such objects in context, it hones in on the social relationships and political systems those objects operated within and reflects back on the limitations of the Western imagination of currency revealed by what collectors call "odd and curious money." It also takes up the question of whether and how standards determine value and expands the social scientific vocabulary for the diversity of forms of political authority that constitute money.

B. Maurer (✉)
Department of Anthropology, School of Social Sciences, University of California, Irvine, Irvine, CA, USA
e-mail: wmmaurer@uci.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_2

Keywords
Primitive money · Origins of money · Numismatics · Bridewealth · Ethnology

Introduction

The archeological, ethnographic, and numismatic literature is replete with instances of objects of all sorts serving some or all of the classic functions of money – shells, feathers, beads, animal parts (or entire beasts sometimes), tobacco, bits of cloth, giant stones, tiny metal axes, or other miniature replicas of everyday tools, belts, bands, copper bracelets, metal crosses, bars of salt, and teeth (human, dog, whale). The list could go on and on. Indeed, listing the items that collectors group under the heading, "odd and curious money," begs the question of whether and why such items should count as money at all. And counting, in fact, may be key, but not in the way most analysts have assumed. There are more ways to count than through the decontextualization and abstraction critics have associated with modern, capitalist money and even within capitalism itself.

In this chapter I argue that the veritable wonder cabinet of odd and curious moneys opens up important questions about the nature and meaning of money and the limitations of economic and other social scientific theories of money linked to a context and consciousness formed by coin. This context and consciousness are also bound by a temporal horizon, the period between the origins of minting coin to the contemporary digitization of money. The current global use of standardized, flat, round bits of metal or pieces of paper passed hand to hand to transfer claims to value, which has shaped so much of Western thinking on the nature of money, is seen better as a brief historical interregnum. The much broader historical and geographic reach of non-coin, non-paper "oddities and curiosities" should lead us to reflect back on what is truly odd in the first place. It may be, this chapter argues, that the very moniker odd and curious is actually useful insofar as it calls attention to the misrecognition of the ubiquity of social relationships, hierarchy, and interpersonal and intergenerational ties in value formation and transfer. Coin consciousness has so limited the imagination that these social relationships are what seem "odd," rather than the apparent fixing or freezing of such relationships in tiny bits of metal that then come to be understood as having value in themselves. Ironically, physical money's heralded disappearance due to its digitization is helping open up the conversation about the true source of money's value in society at large. Indeed, one of the most interesting things about the cryptocurrency phenomenon of the late 2010s, epitomized by Bitcoin, has been its reigniting of the public debate over the true source of money's value: in fictions, in trust, in relationships, and in collective imaginings. The era of coin consciousness may be ending, even if coin and cash themselves endure.

The chapter proceeds as follows. First, it reviews the history of numismatic and theoretical engagement with the colorful entries in the compendium of "primitive" money. It reviews classic understandings of these objects as well as more recent

critical interventions seeking to situate them in light of the history of credit and debt and sociopolitical hierarchy and inequality. Second, it corrects the record on the use of some of the archetypal non-coin, non-paper objects in value transactions, by delineating the differences among exchange, understood in traditional economic terms, and substitution, drawn from the literature on the gift in anthropology. Third, it advances the academic conversation about the relationship between the stuff of money and standardization, introducing questions having to do with political authority, gender, and social complexity. Finally, it concludes by returning to the question, what counts as money, this time querying the operation of counting and its relationship to scales of value and social transformation.

Collecting the Odd and Curious

> The indexes are well done; although such curious misprints, as 'bat' for 'bar,' sends the reader off on the vain quest of a primitive people using bats for money. If dead rats are current on Easter Island, why not bats or mice? (Michell 1949, p. 255)

In 1949 two books with nearly the same title appeared in print: Alison Hingston Quiggin's *A Survey of Primitive Money* and Paul Einzig's *Primitive Money: In its Ethnological, Historical and Economic Aspects*. Their pages explode with examples of the things people around the world have used to mark or effectuate transactions with one another. Each struggled with the question of definition: what to include and what not to include; and are these items "money" or not?

Before getting into the question of classification and definition, however, consider the initial description of one such object, the so-called feather money of the Santa Cruz Islands, the southeastern-most part of the Solomon Islands, to the east of Papua New Guinea. I take this object as an archetype of the sorts of curious things that exercised early chroniclers of the so-called primitive money (Fig. 1). The passages come from Einzig, Quiggin, and Charles Opitz, a twentieth-century collector and authority on "odd and curious" money:

> The feather money of Santa Cruz islands, which was still in use in the late fifties, is regarded by ethnologists as one of the outstanding characteristics of that group. . . . this feather money consists of strip-like coils of fibre about 15 ft long and up to 2–3 ft wide, completely covered on the outer side with overlapping rows of red feathers. . . . Feather money is used to a very large extent as a store of value. It is carefully guarded and stored in a dry, warm place to preserve the colour and elasticity of the coils. Rich men sometimes build special huts for their feather money. As the feathers wear off, the coils depreciate in value. Archey claims that the feather coils are actually used as a medium of exchange. Four coils of good quality would purchase an ocean-going canoe, and a bride would cost 10 coils or more according to her looks and reputed industry. (Einzig 1949 [1966], p. 52)

> The red-feather-money coils tan, ta or tavan . . . of Santa Crux are among the more sensational of the 'curiosities of currency' in the South Seas . . . An average coil made of about 1,800 overlapping scales (lendu) is about 30 feet (10 m.) long, wound in double spirals of 6 turns each . . . on to circular drums of bark. The little honey-bird manga (*Myzomela*

Fig. 1 Feather-money "tevau" of red feathers bound to a fiber backing, coiled on a bark core and with pendant strings of seeds. (Collected by Dorota C. Starzecka, Assistant Keeper, British Museum Department of Ethnography, Santa Cruz Island, Solomon Islands, acquired 1976. © The British Museum)

> *cardinalis*) which supplies the red feathers is the size of a sparrow . . . Although this feather-money is prominent throughout the group of islands and beyond, and the whole life of the natives is centred in it, it is used more for prestige and for ostentation than for trading. . . . coils are too valuable to be expended save in transactions of high importance such as the purchase of large ocean-going canoes (in earlier days), marriage payments, and fines for fornication. On Vanikoro a good wife was worth as much as a small canoe, that is 10 coils, possibly not all of first-rate quality, but among outlying islands the price would drop to perhaps half this. (Quiggin 1949, p. 135–6)

> Each coil is two inches wide and about 30 feet long with red feathers only on one side. A considerable part of the 30 feet consists of two bark rings and twine connecting the rings with the feather strip. A coil takes about one year's labor by three specialists to make. [. . .] Until prohibited by law in the 1930s, wealthy men or groups of men on Ndendo (Santa Cruz) would pay more than 100 coils to buy concubines from the Reef Islands. The price was more than 10 times the price of a wife. The difference between the two was that all family ties were cut with the concubine, while the bride payment sealed new family relationships. (Opitz 2000, p. 142)

The writing in each passage is itself a curiosity. With a mix of encyclopedic precision, zoological detail, and cultural speculation, each also contains provocative tidbits of sociological information – in each case, too, information about gender relations and in somewhat salacious detail (with Einzig being the most prudish). This blending of natural history and the European colonial gaze is not just a reflex of imperial nostalgia, I would argue. Despite the incredible diversity of objects catalogued either by our scholars or our collector, tales of fickle or jilted lovers, blood money, and bride price recur regularly enough throughout the compendium of "primitive money" to suggest that something other than imperial pornography is in play here. Yet our authors barely grasp this (Einzig, I think, hardly at all). The stunning variability of objects simply dazzles them. Quiggin, on the very last page of

her book, comes to the conclusion that the customs of "bride price" and wergild provide opportunities for people to begin to standardize or make "conventional" their practices of payment and identifies this as the "first steps . . . in the evolution of money" (Quiggin 1949, p. 322). I shall return to this point further below.

But it is that profusion of objects used in relations Euro-American observers could see as payment that motivates these collections. Quiggin opens her book noting the narrowness of the term, "money," when defined as the Oxford English Dictionary had done as "current coin" but then worrying whether other schemes introduced to define the "heterogeneous mass of material" accumulating in museums and collections around the world were sufficient to the task (Quiggin 1949, p. 1). Einzig, similarly daunted by the "infinite variety of systems" (Einzig 1949, p. 4) encountered in the archive of "primitive money," at first defers definition due to that variety: some of the objects he discusses had a standard weight or measure, others did not; some had a standard or definite units, others did not; some were valuable items, some worthless except in their enlistment in exchanges; and some fulfilled some but not all of the classic functions of money, some of the time if not all of the time. That variability itself, to Einzig, is evidence of the contradictory and variable human nature:

> Many of our facts lend themselves to classification into conventional categories. They fit into the rules of well established monetary theories. Others defy all the known rules. Instead of supporting each other's evidence, they often tend to cancel each other out. If we are honest in presenting our material of evidence without any attempt at selecting our facts to fit certain preconceived theories, the result is apt to be a mass of apparently contradictory conclusions. (Einzig 1949, p. 6)

What follows, then, is a book structured as a series of lists. One list is organized by geographic region. So we have chapters like, "Cocoa Bean Currency of Mexico," "Coconut Standard on the Nicobars," and "Debts in Dogs' Teeth in the Admiralty Islands." The second list is organized in terms of historical periods – Ancient, Medieval, Modern – again by region, but this time limited to what he would have understood as the "great civilizations" of Greece, Rome, and the European West, China and Japan, and India. So we have both "Rings and Weighed Metal Currencies of the British Isles" in the Medieval period as well as "Rice Money of Japan" in the Modern. Quiggin and Opitz's books are much the same – lists and more lists, object and example after object and example, seemingly limitless variety.

Definitions and Discoveries

Einzig's writing is cantankerous and argumentative; yet also one can sense his own frustration – with the economists he imagined would be his main readers, with anthropologists whom by the time the 1966 edition of his book was published, he knew had rejected his cross-cultural, cross-historical approach and with the material itself. Nothing would fit into neat categories or definitions. It takes him hundreds of

examples from different regions and historical periods, and over 300 pages later, to arrive at one.

> a unit or an object conforming to a reasonable degree to some standard of uniformity, which is employed for reckoning or for making a large proportion of the payments customary in the community concerned, and which is accepted in payment largely with the intention of employing it for making payments. (p. 317, entire passage in italics in original)

Not "absolute" uniformity, he advises, only "reasonable." Not always with the intention of making payments, only "largely" or customarily. Dogmatically criticizing economists' and other ethnologists' definitions as unable to contain within them this or that particular case, Einzig opted for what he called a "broad and elastic" one and advised, further, that it never be "applied too literally" (p. 317).

Not surprisingly, then, reviewers of the book were dissatisfied. H. Michell's review expressed the frustration of readers confronted with the "bewildering array" (1949, p. 253) of examples of money units, objects, and usages across time and cultures. What we have, complained Michell, was a "mere catalogue" and one not of "sufficient importance to be taken as proving or disproving anything in particular" (Michell 1949, p. 254).

Although she covered much the same terrain as Einzig, Quiggin was less exercised than he over the question of definition. Although similar in form – a vast catalogue – Quiggin's book was written in active dialogue with the ethnographers, collectors, and curators returning from journeys abroad and in the colonies with curious objects they took to be money. She began with the collection of material artifacts at the Museum of Archaeology and Ethnology at Cambridge University, and she had as a mentor A.C. Haddon. Haddon had led the Cambridge Torres Straight Islands expedition of 1898, accompanied by W.H.R. Rivers, C.G. Seligman, and Sidney Ray. The expedition included the appropriation of hundreds of objects, some of them deemed to be primitive money. In his preface to Quiggin's book, Haddon (1949) refers to this experience of Cambridge's "field-ethnologists" who observed currency objects in use first hand and in many cases collected specimens for the university's collections. Quiggin thus had direct access to the objects, as well as the contextual information collected around them by these "field-ethnologists." Quiggin makes a point of noting that other treatments of the topic of primitive money were based on literary sources, and not the material artifacts themselves. This gives her book a different character.

Haddon (1949) recounted his intellectual and artifact exchanges with Sir William Ridgeway, whose 1892 *Origin of Metallic Currency*, though focused on the development of coinage, contains a lengthy and heavily illustrated ▶ Chap. 4, "Primitive and Nonmetallic Money". Woodpecker scalps, beaver pelts, and other items from North America; silver bullet money from Southeast Asia; and cowries, cattle, and more all appear in this Ridgeway's compendium. Ridgeway was keen to develop a comparative approach and an inductive method that would allow him to speculate on metal money's origins. Rather than define in ways an economist might appreciate – in terms of the money supply or the price mechanism, as Einzig had done – Quiggin

sought to explain money forms in terms of evolutionary history, implicitly correcting some elements of Ridgeway's account.

Ridgeway found that in almost every instance, he was able to devise tables of standard weights or measures for the objects considered money. From here, he surmised systems of value based on this or that money object, proving for him the origins of money in barter, as a medium for the commensuration of diverse values. The great diversity of forms of currency objects he then arranged in "strata" akin to geological strata: shells at the lowest layer, and in regions where the "Nature lavishly supplies plenteous stores of fruits and vegetables" (Ridgeway 1892, p. 12), leading people to value not items of necessity like furs in the far north but pretty things like shells and other adornments. Ridgeway's thesis was that when societies domesticate animals, then those animals serve as a sort of livestock standard that becomes the basis for the rise of metallic currency pegged to units of cattle. So, he writes that if Native Americans had domesticated the buffalo before the arrival of Europeans, it would surely have served as "the most general unit" of exchange in use (p. 17).

Quiggin expresses skepticism of Ridgeway's position. For one thing, she states at least twice (1949, p. 188, 322) that cattle are not money, though they may be a form of wealth, a standard, or a store of value. But they are not money because no matter what you do with them (except kill them!), they retain their status as useful objects (she borrows the term *Nutzgeld* from the German ethnologist Georg Thilenius) regardless of their enlistment as wealth, standard, or store. This then poses the problem of the dividing line between an object useful in itself or having a use value and "money." For Quiggin, the object's "social significance" has to be brought into the picture. Social significance is different from, more than, the use value of the thing in itself.

> But 'social significance' is a vague term, and in sorting out material in a museum it is difficult to discover a dividing line between the two classes. Shells are merely shells on one island, but are used in trade exchange with another, where they form the currency. Mats are used in barter, but some, acquiring dignity with age, or prestige with travel or special use, develop into a recognized currency. (p. 3)

One could attempt to define items that are easily divisible and transportable as money (and Quiggin does this, p. 188, in arguing against the idea of cattle as currency). But then one has to contend with contexts of use: "Is a string of shell-money no longer currency when you wear it round your neck? Is a sovereign no longer money when dangled on your watchchain?" (p. 3). The way collectors and curators have labeled objects in the museum did not help matters (p. 114). And, furthermore, there are the interpretations and relations between transacting parties: "the two parties in a transaction may themselves stand in different categories. The trader may consider that he is paying current money when he buys a fowl for ten lengths of brass wire; while the seller regards the exchange as 'mere barter'" (p. 2).

Quiggin's criticism of Ridgeway pointed her toward a way out of the classificatory conundrum. She did not linger over her observation that "Women were appropriately doctored for payments estimated in values of cows, mares and she-camels

according to the position of their husbands" (p. 187) until the final pages of her book. But there she made clear the signal importance of bridewealth and wergild in what she calls the evolution of money. "It is not without significance," she wrote, "that in any collection of primitive currency the majority of the items are described as 'used in bride-price'" (p. 322). Bridewealth and blood money introduce a standard, and one conventionalized in terms of a token (Thilenius's (1921) *Zeichengeld*), to measure the "price" of a person and her or his capacities (at least in Quiggin's understanding). Furthermore, Quiggin says the token must have four "essential" qualities (portability, divisibility, durability, recognizability), but it's difficult to understand why any or all of these is actually necessary: all that is needed is a conventionalized mechanism for recording debt – a debt of a human kind, that is: a debt incurred through the appropriation of an irreplaceable and unique human person.

I am intimating here that there is a very small step from Quiggin's realization of the role of bridewealth and wergild in the so-called primitive money and David Graeber's position in *Debt: The First 5,000 Years* (Graeber 2011). Like Quiggin, Graeber challenges the idea that money is developed out of barter, to solve the problem of the double coincidence of wants. Quiggin simply states, "the inconveniences of barter do not disturb simple societies" (p. 321). Like Quiggin, Graeber pays attention to the standards apparent in payments for injuries and death, which he shows were used not to establish value per se from an abstracted human being or its agencies (the price of a lost limb or an accidental death, say) but to repair social relations. Thus for Graeber "social currencies" – his term for most of the items covered by the numismatist's "odd and curious" umbrella category – help people create, maintain, sunder, and reorganize the web of social relationships that sustain them. Bridewealth and wergild (and mortuary payments, for that matter) are instances where such maintenance really matters, marking moments where there is the potential for decay or destruction of relationships. That so many social currencies are also objects of adornment, made to be seen, makes sense in this context, insofar as they mark and make memorable the contexts of the relations they de- and re-stitch together (see Graeber 1996).

Quiggin's book ends on the development of coin. Graeber pivots around it: the standardization of value in a wholly abstract token, and a token whose value and thus whose power could be hidden rather than displayed (like coins in a sack, or palmed in the hand), represents a violent separation of value from context. It permitted the total abstraction of person from context, too, in the form of slavery, which became a model for marriage. Goodbye bridewealth and beautiful and weird social currencies; hello cash nexus, patriarchy, and the violent abstractions of coined money: "marriage came more and more to resemble a simple cash transaction" (Graeber 2011, p. 180; see Maurer 2013, p. 87), and the person became just another "generic value capable of being added and subtracted and used as a means to measure debt" (Graeber 2011, p. 159).

Just to wrap Graeber's story back around to the contexts of the "discovery" of social currencies, Eagleton and Williams (2007, p. 200) note the role of Portuguese traders in providing to the West the first accounts of non-coin-based economies.

Duarte Lopez's 1541 travel account of his journey through the Congo explained the use of small shells as money. It also was the first account of the shell money that would form the basis of the slave trade.

Standards, States, and Power

> [I]f one asks what is 'primitive' about a particular money, one may come away with two answers: the money-stuff – woodpecker scalps, sea shells, goats, dog teeth – is primitive (i.e., different from our own); and the uses to which the money-stuff is sometimes put – mortuary payments, bloodwealth, bridewealth – are primitive (i.e., different from our own). (Dalton 1965, p. 44)

By the time modern anthropology got to the odd and curious collection, the actual use of social currencies in many of the contexts described by Quiggin and Einzig was going out of fashion. Colonialism and capitalist expansion had altered the terrain. Returning to our feather coils, Opitz writes that in 1962, when he visited, there were only five men who possessed the knowledge to make the object (Opitz 2000, p. 143). Yet the anthropological debate went into full force. Pitting formalists who thought the concepts and tools of economics adequate for small-scale societies just as well as for capitalist states, against the substantivists who held onto cultural particularity and diversity, the irony of the debate was that it was taking place just as the raw material for it was alternately disappearing, going underground, or creating complex syncretisms – sometimes out of temporal synch with capitalist money relations – that it took the field as a whole several decades to come to terms with. So, Paul Bohannan's (1959) classic article on the impact of Western, capitalist money on the Nigerian Tiv economy was authored just as the latter was apparently falling apart, brass rods and special cloth giving way to the pound sterling and uniscalar valuation. Yet Jane Guyer (2004) was able to show years later that the "traditional" Tiv economy could only be understood in terms of wider regional flows. And those "traditional" objects of money, at least in sub-Saharan Africa, were themselves bound up in – and even produced by – European imperial forces.

Early anthropological forays into the so-called primitive money such as those of Bronislaw Malinowski (1921) and Raymond Firth (1929) argued against the idea that the strings of shells of the Trobrianders or Solomon Islanders were money. For Firth, only those objects that served to convert between one object or service and another and thereby formed a standard of value should be called "money" (Firth 1929, p. 881). Marcel Mauss (1925), in *The Gift*, had already taken this perspective to task, laying the ground for what would develop into the "substantivist" critique of economic anthropology. According to Mauss, Malinowski had so narrowly defined money that only modern, Western capitalist money would fall under the definition.

Yet as Keith Hart has argued (2005), anthropology pretty much ignored this insight and got stuck thereafter. The formalist/substantivist debate on both sides was wedded to a conception of the market as a site of depersonalization and abstraction. For the formalists like Cook (1966), this idea of the market was not a

value judgment but simply meant that markets, pretty much the same everywhere, could be analyzed using the tools economists had applied to capitalist markets. To the substantivists, however, the capitalist market's depersonalization of relations meant that wherever the market touched, debasement and destruction were sure to follow. Thus Dalton argued that it was the capitalist market organization of Euro-American societies that determined money's uses as the classic means of exchange, measure of value, store of value, and unit of account. Other societies, organized according to other principles, have moneys not for commercial exchange but for the discharge of social obligations and debts, rituals of redistribution, or life cycle payments like bridewealth or mortuary payments. For Dalton, it is a category error to assume that primitive moneys must function like capitalist, commercial ones – each is created rather to serve different purposes. Dalton concludes, "money has no definable essence apart from the uses money objects serve, and these depend upon the transactional modes that characterize each economy" (1965, p. 62).

The substantivists' aversion to markets however led them away from Mauss's fundamental insight that money and markets are about the extension of human relationships, even if they are capitalist in nature. It also took attention away from the dynamic processes through which hierarchies of value vie with one another, intertwine, or diverge in specific contexts, most significantly for the purposes of this chapter, the colonial context.

Take the famous case of cowrie inflation. In an important paper, C.A. Gregory (1996) challenged the influential account of Hogendorn and Johnson (1986) of the cowrie shell bubble of the nineteenth century. Cowries circulated across a wide geographic range; they possess all the qualities Western observers often ascribe to money (durability, divisibility, and so on); and they are naturally occurring, non-manufactured objects – the perfect example of a nonmetallic, "primitive" money! However, these "traditional" currency objects were the subject of a substantial European trade. Europeans purchased shells from Indian merchants and used them in the West African slave trade – having learned something of how this trade could work from Duarte Lopez's travelogue, one might presume. Hogendorn and Johnson estimate that between 1700 and 1790, the equivalent of ten billion individual shells was shipped to West Africa (Gregory 1996, p. 198). After the abolition of slavery, palm oil became the commodity of choice in the cowrie trade. The bubble exploded in the middle of the nineteenth century, as the Maldivian cowrie (*Cypraea moneta*) was displaced by the Zanzibar cowrie (*C. annulus*), the money supply consequently increased faster than the number of transactions, and the price of cowries plummeted. People took what cowries they had and buried them in hoards, ready for excavation in case their value ever returned.

For Hogendorn and Johnson, this represents a classic case of the quantity theory of money and Gresham's Law. When the money supply expands faster than the quantity of transactions, prices rise (see Gregory 1996, p. 199). Further, following Gresham's Law, bad money drives out good – the cheaper Zanzibar shells crowded out the Maldivian ones. European mercantile traders, having introduced new sources of "traditional" moneys, flooded the market and created a bursting bubble. Something similar happened, they argue, with the trade in manilas (copper and brass

Fig. 2 Manillas, bronze. Made in Birmingham, UK; found in south east Nigeria. (From the collection of the Rev. J.H. Slater, acquired by the British Museum 1971. © The British Museum)

ring currencies of West Africa) and American wampum (see Gregory 1996, p. 200). In those instances, "primitive" moneys were mass produced by Euro-American capitalist commercial enterprise. Jane Guyer has related to me the vast collections of manilas in warehouses outside of Liverpool, amassed during the slave trade and manufactured in Birmingham (Fig. 2).

Gregory challenges Hogendorn and Johnson's account by noting the importance of the introduction of a new standard of value during European imperial conquest. There was not just a commercial, mercantile game taking place, with European traders seeking out new sources of "traditional" currencies in order to capture the market in slaves and palm oil. During the same period, European powers were assuming political control over vast territories and demanding payment of taxes in terms of their own token. As Gregory writes, "First comes the imperialist conquest of the kingdom. . . . New monetary standards follow. . . . [A]nd taxes are required to be paid in this new standard" (Gregory 1996, p. 208). As the state consolidates power, the old standard, the cowrie, loses value. Those seeking to purchase commodities like palm oil in the old standard had to offer more and more of it to purchase what they wanted. The rise of the level of cowrie-denominated prices thus did not occur because of a rise in the supply of cowries but a fall in demand due to the institution of a new standard (Ibid). The important corrective here is to add another variable into the story of shell money or any social currencies' inflation during the colonial period: colonial state power.

Modern money outside the colonial context, too, is a measure of state power. Gregory argues that the emphasis in other studies of money – "primitive" or otherwise – on weights and measures (e.g., as evidenced in Ridgeway or Einzig) is misplaced. It is the state the sets the standard. Therefore, the study of the transition from "primitive" currencies to "modern" ones demands a political analysis of contests over standard setting. Bohannan's Tiv history thus requires a political

analysis of imperial rule, not just distinct spheres of exchange, together with an analysis of the intercultural interfaces shaping contests over value standards (Guyer 2004).

Back to Bridewealth

But then what about places where there is or was no state? Notwithstanding the wider regional and colonial relationships in which many of the societies where social currencies were collected by Europeans during that imperial encounter, we may take a broader lesson from Gregory's insight on power and money.

Jane Collier (1988) developed a typology for understanding inequality in stateless societies. To get around the problem of contact and change, she developed ideal-typic models from the ethnographic record. Now, I am well aware of the limitations of such model building. The result may be a series of just-so stories. But either we can see them as useful for actually grasping the nature and value of social currencies in their original contexts. Or, alternately, we can use them to provide fodder for Euro-American reflections on the specificity of our own world by giving us new resources for imagining, albeit in our terms, others'.

Collier models inequality in classless, stateless societies based on differences in marriage, the site of the making of new social relationships, of advancing claims over others, and of course exchanging goods. Inequality based on gender and generation is instituted and made manifest in marriage, but differently so, according to her different models. In her bride service model, there is a division of labor by sex; subsistence is via hunting and gathering. There is a sexual division of obligations, too, with women expected to feed men. A man with a wife, then, obtains status vis-a-vis other men. But he can appear to be obligated to no one in that he "earns" a wife through his prowess. The model is useful to explain hunting and gathering or hunter-horticulturalist societies. In Collier's equal bridewealth model, in contrast, people come into the world with preexisting obligations to their elders. Junior men are dependent on elder men for the things needed to assemble a marriage gift. Junior men need to earn the respect of their elders to get the gifts that then subsequently also confer respect. Politics consist in discussions over prior gifts, which assist in determining one's obligations. The model is useful to explain small-scale societies like the Tiv or the Trobrianders where marriage gifts seem to play such an important role in maintaining social cohesion and balance, and such gifts are often made up of the exotic items filling the "odd and curious" catalogue. In her unequal bridewealth model, in contrast, people are born into hereditary statuses but rank is always unstable; gifts help people assert rank. The model is useful to explain societies like the Kiowa where outside observers have reported that "high-ranking brides 'cost' more than low-ranking ones" and marriage gifts were made in horses (Collier 1988, p. 144). Collier also identifies different political idioms for each of her models. Where "bravery" is a dominant political idiom in bride service because a man has to demonstrate his strength in order or "earn" a wife and "respect" the prevailing idiom in equal bridewealth models because a man has to demonstrate his obligations to his

elders to assemble an appropriate gift, in unequal bridewealth models, the organizing idiom is "rank."

Think back to Quiggin's discussion of bride-price setting the standard for "primitive money." She assumed that currency objects formed a kind of conventionalized standard for bridewealth. This presumed however that the people who produced such objects were all playing the same kind of game, a game of equilibrating values. Collier's models suggest not only that there might have been different games being played but that there might have been different values at stake, or different idioms, in each game. When a gift is about having earned the respect of one's elders and goods exchanged are based on mutual obligations, then it matters a great deal that the gifts can be seen by all, measured against one another, and displayed to function as memory devices about those prior obligations. Such is the case with equal bridewealth. The gift is not "exchanged for" the wife but stands in a system of relations in which the wife is embedded – "substituting" in that system (Strathern 1988, p. 183) for the person, not commensurating the value of the person according to a standard. When a gift is about asserting rank, as in unequal bridewealth, then a marriage associated with it is always unstable – gifts have to keep flowing, there needs to be a continuous circulation to keep asserting one's position in a game with others for status. Again, display matters – as a way to show rank. But this form of visibility is different in kind from that of equal bridewealth. To Western observers, however, it looks like a price is actually being set on a bride, one that goes up the higher the ranking of her family, and one at least in some contexts seemingly rendered in horses – hence Ridgeway's cattle standard.

Returning to Quiggin, echoed by Grierson and Graeber: once there is a standard for bridewealth, you start to see "money." But in bridewealth societies, gifts are not truly fungible. They are specific to specific sets of relationships. A pig is not just a pig, but *this* pig is in relation to *those* other relationships. The pig thus substitutes for the person – it is not exchanged for the person, such exchange requiring the imagination of an abstract standard "above" or outside those relations (Strathern 1988).

This is not a relativist gesture. It is not that what looks like a gift or a debt here or there has different "meanings." Rather, it is that it is produced through different practices unfolding over time, sometimes part of multiple contests over value, multiple and contending claims, and the inherent instability of relationships.

My perspective is broadly consistent with that offered by Marshall Sahlins, who argued that money-stuff tends to be found where what he called "balanced reciprocity" is taking place, that is, where there are lot of regular and regularized exchanges such as those involved in bridewealth. He does not find "primitive money" in subsistence-based bands (akin to Collier's bride service societies) nor in chiefdoms (where, he writes, "wealth tokens … tend to bear little exchange load" (Sahlins 1972, p. 227). This would make sense in terms of Collier's models: in what Sahlins calls bands, men would demonstrate their prowess through hunting, providing the wife's family with meat and skins (the pelts, perhaps, found in Einzig and Quiggin's catalogues of money). In chiefdoms the main problem people are trying to solve is how to pay tribute to their overlord; objects of wealth here take the form of prestige

items associated with rulers – who demand tax in terms of their own standard – rather than tokens used to demonstrate a claim over another person or used to facilitate a market exchange.

Collier's models help show how the analytical focus on understanding social currencies in terms of a standard of value suffers from two failures of the imagination. First, pace Gregory, while states may make standards, there may be other ways standard gets set in stateless societies (modulo contact and conquest). Collier expands the horizon of the political, allowing us to reflect on other ways of setting standards. Second, pace Quiggin, Grierson, or Graeber, the focus on standards gets stuck on number or quantification. Gregory is on the mark when he states that it is not the metric that matters but the power relations. The numbers may matter less than the game being played. And there might not be a number at all, even if it looks like people are counting. Whether balancing relationships or jockeying for position in an unstable system of rank, social currencies that can be worn or seen permit the display of status as well as provoke the memory of obligation.

Counting on Currency

I opened this chapter noting that what I call coin consciousness has limited the analytical imagination with respect to the so-called primitive money. My discussion of the political games at play in the use of any currency, and the role of political processes in shaping the idioms in which people come to understand what counts as valuable, may shed some light on the perennial association of money with number and counting. Counting and quantification in capitalist societies serve to organize different values in terms of one standard of commensuration or equivalence. Because of how we "do" money, it makes sense for us to think in terms of the equilibration of things that are different in kind or in number: I can put a price in dollars on a diamond ring, an hour of labor, a bushel of potatoes, even love, a thought, or an idea. We can easily envision these prices being counted out in coin or paper banknote, the token here standing in for the invisible, mental standard set by, in our case, the state. The classical social theorists of the nineteenth century – Marx, Weber, Simmel, and even Freud – all saw in this ability to enumerate a form of abstraction and disassociation of things and relations from their contexts and, for Marx at least, a misrecognition of the bases of value itself in human activity and consciousness through its fetishization in money's commodity form.

One difficulty with this perspective is that it is hard to specify when counting and currency came together. Is it particular to industrial capitalism? And how even to approach historical conjunctures of counting and currency when we only see the latter through the lens of our own system? Thus the barter origin story for money essentially takes capitalist market society's way of rendering exchange value and projects it back in time or into other non-Western cultural contexts. For Karl Polanyi (1944) and the substantivist economic anthropologists, the difference was between socially embedded markets that were limited in their scope and their importance for the overall society, and the self-regulating market of capitalism which appears to its

participants as socially disembedded, as following its own internal laws, and as central to the functioning of society. For Polanyi, quantification might have its place in a small-scale or tribute-based society, but it was a circumscribed place, not an overall structuring logic.

Evidence from ancient near eastern archeology helps provide another way to think about the relationship between counting and currency. And it nicely fits with the perspective offered here that the so-called primitive moneys index different political idioms – with "political" used in the expansive sense provided by Collier's analysis of marriage gifts and the relations of inequality behind them. Denise Schmandt-Besserat (1992) has provided a compelling account of the use of the small clay tokens found in great numbers in numerous sites across the ancient near east (Fig. 3). Dating from a wide range between 8000 and 3200 BCE, these tokens represented agricultural goods and livestock. They seem to have been initially passed by hand to manage agricultural production, serving as rudimentary record-keeping devices for the proto-bureaucratic organization that increasingly larger, denser settlements based on seasonal agriculture would have demanded. Later, by around 3000 BCE, tokens would be impressed on clay balls and then sealed up inside. Eventually, the tokens themselves were dispensed with, the impressions alone serving in their place. Schmandt-Besserat sees these tokens and the preliterate inscriptions people used them to produce as preliminary to both counting and writing and as "mnemonic device[s] by which to handle and store an unlimited quantity of data without risking the damages of memory failure" (Schmandt-Besserat 1995, p. 2100).

The problem the people who created these token systems were trying to solve was a memory problem and a memory problem tied to administration. Schmandt-

Fig. 3 Clay accounting tokens, Susa, Uruk period. The Louvre, Paris. Department of Oriental Antiquities, Richelieu, room 7, case 3, © Marie-Lan Nguyen/Wikimedia Commons

Besserat thus reinforces Keith Hart's (2000) longstanding argument that money is a social memory device. It also is in accord with other archeological evidence that systems of accounting emerge in the development of ancient state formation for administrative not strictly speaking "economic" reasons nor for reasons tied to trade and exchange but rather tribute and redistribution (e.g., Hudson 2004). Systems of accounting in ancient Mesopotamia did not require a circulating token to represent value in exchange. Instead, they used cuneiform inscriptions – the technological descendents, so to speak, of Schmandt-Besserat's token impressions on clay balls – and a standard of denomination. Even where people recorded such standards in terms of precious metals, those measures did not actually need to be present in any given transaction. Thus Van de Mieroop writes that money emerges in "statements that something was in the possession of someone else" and that silver, copper, or bronze could be used as measures "without having to be present" (2014, p. 20). Records of price lists and loans existed even if most precious metal was immobilized in temples and palaces.

Counting is a good thing to have and to be able to transmit to others through material, durable, extra-cognitive systems when you have to deal with the temporal cycles associated with grain or goats (and it bears reminding a modern, non-agricultural audience that goats unlike cows are seasonal breeders). The anthropologist Claude Meillassoux (1975) argued that when land becomes an instrument of production rather than an all-giving subject – when it is a tool in the production process rather than a site for hunting and gathering – and that when we look in the ethnographic record at contexts of quasi-settled horticulture instead of migratory hunting and gathering, we find people developing systems of accounting between the full season and the fallow.

In a fascinating cross-cultural analysis, Basu et al. (2009) show that record keeping like that developed in ancient Mesopotamia is a necessary but not sufficient indicator of money and credit (p. 896). External memory devices like clay tablets with cuneiform writing, or, across the world and several centuries later, the Inkan knotted string khipu accounting devices (see, e.g., Urton 2012), permit the recording of completed or planned transactions to allow for complex social and administrative functions associated with larger-scale ancient societies. I include these examples in this discussion of nonmetallic money because they present means through which people could conduct some of the interactions associated with money without specific money objects. They also point out the central role of accounting in the figuring of money – and how accounting technologies obviate the need for any transacted object at all. They thus prefigure by millennia the digital accounting by which so much contemporary money is transacted.

Conclusion

From the ethnologists' wonder cabinet to anthropological studies of small-scale societies and the archeology of ancient states, the history of nonmetallic and so-called primitive money sheds light on several core conundrums in the study of

money more generally. First, it decenters the classic focus on exchange, placing emphasis more on social payments like marriage gifts, tax, or tithe. Second, it introduces political variables – broadly defined as the processes through which people deploy power to institute and enforce inequality or hierarchy – in the creation of money and money-like institutions. Third, it demands an expansion of our vocabulary for such political processes. Whether we include terms like Sahlins' "balanced reciprocity" or Collier's "equal bridewealth," seeing these as political not just economic categories allows us to ask questions about how powerful people in any given society establish the standards through which value is configured and assessed. Fourth, it requires us to be cautious in assuming that evidence of counting or number necessarily means depersonalization, abstraction, or the kinds of equilibration we tend to assume whenever we see such quantification in our own capitalist, market societies.

This survey also reminds us to use caution in too neatly defining the boundaries of the system within which we locate such money objects. Quiggin notes that the parties to a transaction might not always understand the transaction in the same terms or be playing the same game. Guyer and Gregory spotlight the larger socio-political and economic systems within which such "primitive" moneys were used, as well as the larger zones of contact, intercultural exchange, and (mis)communication in which they assumed value. Such a caution may also lead us to reflect on our own money objects and systems, whether they are as straightforward as they generally seem in everyday use or whether they are in fact as "odd and curious" as the "primitive" moneys that so exercised the early ethnologists.

Cross-References

- ▶ Gresham's Law
- ▶ Money, Law, and Institutions
- ▶ The Role of Money in the Economies of Ancient Greece and Rome

References

Basu S, Kirk M, Waymire G (2009) Memory, transaction records, and the wealth of nations. Acc Organ Soc 34:895–917

Bohannan P (1959) The impact of money on an African subsistence economy. J Econ Hist 19:491–503

Collier JF (1988) Marriage and inequality in classless societies. Stanford University Press, Stanford

Cook S (1966) The obsolete anti-market mentality: a critique of the substantive approach to economic anthropology. Am Anthropol 68:323–345

Dalton G (1965) Primitive money. Am Anthropol 67:44–65

Eagleton C, Williams J (2007) Money: a history. British Museum Press, London

Einzig P (1949 [1966]) Primitive money: in its ethnological, historical and economic aspects, 2nd edn. Pergamon Press, Oxford

Firth R (1929) Currency, primitive. Encyclopaedia Britannica, 14th edn. New York: Encyclopaedia Britannica, Inc.

Graeber D (1996) Beads and money: notes toward a theory of wealth and power. Am Ethnol 23(1):4–24
Graeber D (2011) Debt: the first 5,000 years. Melville House, New York
Gregory CA (1996) Cowries and conquest: towards a subalternate quality theory of money. Comp Stud Soc Hist 38(2):195–217
Guyer J (2004) Marginal gains: monetary transactions in Atlantic Africa. University of Chicago Press, Chicago
Haddon AC (1949) Introduction to A. Hingston Quiggin, a survey of primitive money. Methuen, London, pp vii–viii
Hart K (2000) The memory bank: money in an unequal world. Texere Publishing, Cheshire
Hart K (2005) Notes toward an anthropology of money. Kritikos: an international and interdisciplinary journal of postmodern cultural sound, text and image, vol 2, June 2005. Available at http://intertheory.org/hart.htm. Last accessed 31 Dec 2017
Hogendorn J, Johnson M (1986) The shell money of the slave trade. Cambridge University Press, Cambridge
Hudson M (2004) Introduction: the role of accounting in civilization's economic takeoff. In: Hudson M, Wunsch C (eds) Creating economic order: record-keeping, standardization, and the development of accounting in the ancient near east, vol 4. CDL, Bethesda, pp 1–22
Malinowski B (1921) The primitive economics of the Trobriand islanders. Econ J 31:1–16
Maurer B (2013) David Graeber's Wunderkammer, debt: the first 5,000 years. Anthropol Forum 23(1):79–93
Mauss M (1925 [2016]) The gift, expanded edition. Selected, edited and translated by Jane I. Guyer. HAU Books, University of Chicago Press, Chicago
Meillassoux C (1975) Femmes, greniers, et capitaux. Editions Maspero, Paris
Michell H (1949) Review of primitive money in its ethnological, historical and economic aspects, by Paul Einzig. Can J Econ Political Sci/Rev Can d'Econ Sci Polit 15(2):253–255
Opitz CJ (2000) An ethnographic study of traditional money. First Impressions Printing, Ocala
Polanyi K (1944) The great transformation. Beacon Books, Boston
Quiggin AH (1949) A survey of primitive money: the beginnings of currency. Methuen and Company, London
Ridgeway W (1892) The origin of metallic currency and weight standards. Cambridge University Press, Cambridge
Sahlins M (1972) Stone age economics. Aldine de Gruyter, Chicago
Schmandt-Besserat D (1992) Before writing, vol. 1: from counting to cuneiform. University of Texas Press, Austin
Schmandt-Besserat D (1995) Record keeping before writing. In: Sasson J (ed) Civilizations of the ancient near east, vol 4. Scribner, New York, pp 2097–2106
Strathern M (1988) The gender of the gift. University of California Press, Berkeley
Thilenius G (1921) Primitives geld. Arch Anthropol 18:1–34
Urton G (2012) Recording values in the Inka empire. In: Papadopoulos JK, Urton G (eds) The construction of value in the ancient world. Cotsen Institute of Archaeology Press, University of California Press, Los Angeles, pp 475–496
Van de Mieroop M (2014) Silver as a financial tool in ancient Egypt and Mesopotamia. In: Bernholz P, Vaubel R (eds) Explaining monetary and financial innovation. Springer International Publishing, Cham, pp 17–29

Part II

Money, Coinage, and the State

Monetary System of the "Ancient Régime" (Third to Eighteenth Centuries)

5

Georges Depeyrot

Contents

G. Depeyrot (✉)
Centre National de la Recherche Scientifiquen, École Normale Supérieure, Cahors, France
e-mail: georges.depeyrot@orange.fr

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_6

Abstract

The monetary system of the "Ancient Régime" can designate the period in monetary history when the unit of account was not physically represented by a coin. It can mean the long period from the first coins to the appearance of marks of value, that is, the period ending more or less with the nineteenth century.

This chapter analyzes the evolution of the relation between the unit of account and currencies from the Roman empire to the end of the system. It overviews the main periods and empires, and also considers the technical evolutions which allowed the transformation of the monetary system.

Keywords

Ancient Régime · Monetary system · Roman empire · Barbaric invasions · Sasanians · Byzantine monetary system · Carolingians · Minting technology · Monetary economy · Currency

Introduction

The monetary system of the "Ancient Régime" can designate the period in monetary history when the unit of account was not physically represented by a coin. It can mean the long period from the first coins to the appearance of marks of value, that is, the period ending more or less with the nineteenth century.

The situation is not so simple.

First we have to remember that during the whole period, coinage was a benefit for the king or any person who had right to issue coins.

Some coins could be at the same time currency and unit of account. This was the case for the Roman denarius, the Carolingian denarius, etc., even if the coin never had any explicit mention or mark of its own value: its name was the unit of account. Other coins can be associated with a particular amount of units of account, according to certain characteristics: a specific typology, a lighter weight which designated them as a division of the unit, attributes on the coins that designated their value (as in the case of the Roman antoninianus whose radiate crown indicated that its value was double) or symbols that designated the weight or fineness of the coins.

The "Ancient Régime" monetary system was not characterized by the total absence of any marks of value. In some cases, the system was associated with coins with explicit marks of value in units of account and coins without any mark.

In antiquity, the dominant system was a kind of correspondence between coins and units of account. It was not explicit but implicit. Such a relationship was convenient in a period without inflation or with equilibrium between prices, demography, monetary stock, and inflation. We have called this link between coins and a certain number of units of account the "Roman system." In times of inflation, the intrinsic value of the coin would be decreased.

On the contrary, "the Alexandrian system" separated the currency from the unit of account. It is clear that the coins struck in Alexandria during the Roman period of inflation were unaffected or less affected by a reduction of fineness. This clearly underlines that the value of the coins in units of account was increased. This was possible because Egypt did not use Roman coins.

The history of this long period is, in fact, the history of the relationship between the two types of systems, "Roman" and "Alexandrian" which was in fact a kind of floating system (Depeyrot 1988, 1995a, 1996a, b).

In this chapter, I use the term "floating" for a coin not linked to a certain amount of units of account. A "semi-floating" system describes a monetary system composed of coins without fixed relationship to units of account and coins with an explicit relationship to a certain amount of units of account.

Birth and Fall of the Monetary System of the Roman Empire

After the seventh century BC and the "invention of currency" (which was only a new form of means of exchange and evaluation), the development of the monetization of society during Antiquity was mainly linked to the ability of kingdoms, states, and cities to find mines and to control the production of precious metal.

Up until the discovery of the Americas, metals could be obtained either because the mines were in the territory of a state (whatever its form was), or by trade. For example, Athens was favored by the presence of the silver mines of Laurion, as well as Thasos by the gold and silver mines of the Pangaion Hills.

Some regions were favored by the presence of important mines. It was the case for the large kingdom of Persia. The Achaemenid Empire was the most important empire at the time and included Lydia (the western part of modern Turkey) where, around 635 BC, the first coins were struck. It also included the mountains and rivers of modern Iran, Afghanistan, etc., rich in gold and silver, but gold was rarely used in monetary production and was instead kept in the royal treasury or used in jewelry. Egypt and Sudan were also very rich regions producing gold, but the pharaohs never used coinage until the arrival of the successors of Alexander the Great (BC 336–323) and gold was used as reserve of value. In Europe, mined silver and gold or alluvial gold were produced in various regions, especially in modern Romania (the gold and silver mines of Roșia Montană – Alburnus Maior in the Roman period). This was also the case in the Balkans (Andreau 1989, 1990), in Gaul (modern France) with alluvial or small mines of gold, and in Spain with gold mines (Domergue 1990).

Trade was a means of obtaining metals. The small island of Aegina acquired metal through its position as an important maritime power and its relationship with Athens. This was the case for many Greek cities in continental Greece and even abroad (the South of Italy).

During the long period from the seventh to the fourth century, the production of coins increased. Many factors explain the development of coinage, all of them can be summarized as a "social change" that was linked to a very strong demographic development. It led to the creation of colonies and new cities that developed trade between mother-cities and their colonies, professional specialization in cities (which needed currency to buy from each other), and, mainly, the development of wars and all that that involved.

These military changes were, as well as the development of trade, one of the reasons for the development of coinage. In this aspect, Homer's *Iliad* must be considered as the perfect description of a war during the premonetary period, with a long description of the walls of Troy, armies, ships, weapons, etc. Tradition dates the Trojan war to the twelfth century BC and the "redaction" of the *Iliad* to the eighth century BC; all the events took place before the "invention" of coinage. It is likely that the development of coinage was related to changes in military organization. For traditional armies composed of farmers, artisans, etc., with some rich nobles using horses and chariots, sometimes fighting in single combat (the combat between Achilles and Patroclus is an archetype) were substituted armies composed of professional hoplites, heavily armed, and grouped in phalanxes. The costs of armor, of weapons, of mercenaries, and the professionalization of these soldiers favored the use of precious coins, generally struck with metal otherwise preserved in temples. This is the reason why the coins bore the effigy of the god or goddess of the city (or the king or emperor who personified a god on earth). It was also the mark of the divine character of the origin of the metal. Only a divine metal could pay for the lives of professional soldiers.

From these origins, all western currencies inherited common characteristics: a divine representation on the obverse or a human personification of a god (king, emperor…) or religious symbols (such as the cross) or even a personification or a symbol of the State such as the Republic. They also inherited the relationship between precious metal and value: it is a major difference with Asian monetary systems based on copper or notes (Depeyrot 2013) and with pre-Columbian American or African systems, based on specific, more or less natural objects (such as glass beads, cowries) even if specific ceremonies give them their monetary functions.

Due to the circumstances of its birth, western currency also contained its weakness: the necessity to control a flow of new metal to respond to the needs of a growing population, economic expansion, and to compensate for wear and loss of coins. If the ancient mints would not have been able to find metal, the western monetary system would have been based on a different system, perhaps copper (as in Asia), perhaps a kind of fiat money.

The turning point was the conquest of the Persian Empire by Alexander the Great. According to historians, Alexander looted Persia and took the treasure of king Darius. The amount of gold and silver was, according to the tradition, 300 metric tons of gold and 3,000 metric tons of silver. A precise analysis of the number of coins issued by the mints, during and after the reign of Alexander, confirms the seizure and melting of the Persian gold and silver, its transfer to the Greek mints or to the mint producing coins for Alexander. After the invasion of Greece by the Roman legions and the looting of their treasures, the cities were obliged to pay tribute. Slowly the reserve of metals shifted from Greece to Rome and the production of the Greek mints decreased as those of Rome increased. The Roman monetary stock was created on the ruins of the Greek one. But an increased monetary stock needs a constant or even increased alimentation with new coins: the greater the number of coins in circulation, the greater the possibilities of losing them (including transformation into jewelry, coins put in tombs, etc.). The civil wars contributed to an increased need for coins. Step by step all the *imperatores* (generals in chief of the legions during the Republic) and the emperors (after the end of the Republic in 31 BC) financed their legions by looting. Pompey pillaged Armenia in 69 BC, Caesar plundered Gaul in 52 BC, Marcus Antonius used the Egyptian precious reserves of Queen Cleopatra, but after their deaths in 30 BC, Octavian-Augustus did the same (de Callataÿ et al. 1993).

Lacking new territories to invade, Nero (AD 54–68) reduced the weights of gold and silver coins and his successor (after a new civil war) Vespasian (AD 69–79) tried to restore the public finances, in the process obtaining a solid reputation of having been an "avaricious emperor." Trajan (AD 98–117) was the last emperor to have been able to conduct a campaign which ended in the plundering of a treasure. In two large and difficult campaigns, he seized the reserves of the Dacian king Decebalus (died in AD 106). According to tradition, 165 t of gold and 327 t of silver were sent to Rome where they financed a vast monetary reform with the re-melting of the old coins, the construction of the forum of Trajan, the building of Trajan's column (depicting the wars against Decebalus and – on the last scenes – the transfer of gold and silver to Rome), and the preparation of a campaign in the East (Guey 1966).

After the reign of Trajan, all the wars were defensive, even if battles took place in *Barbaricum*. Slowly, the silver content and/or the weight of the denarius decreased, as well as that of gold coins. In 215, Caracalla, in order to pay the armies (which were mainly organizing the defense of the Empire), created a new silver coin, the antoninianus, which had a silver content equal to 1.5 denarii but a legal tender of 2 denarii. An increase in the cost of defense, the loss of territories (and so loss of taxes), perhaps a climatic crisis, civil wars, all these events (and others) created a deep monetary and economic crisis.

During the third century AD, good gold and silver coins were swiftly removed from circulation. Some were hoarded inside the Empire either in hoards, or in jewelry, or as gifts to the Church; some were taken as ransom or booty by the Barbarians who invaded the Empire or asked to be paid to stay in *Barbaricum* (Bland 2013).

The crisis was dramatic: if we consider the weight of pure silver produced by all the mints and if we give an index of 100 to the production of the years 238–244, the index falls to 8 for the production of the years 276–282: the denarius (3.40 g), which contained 98% of silver at the beginning of the second century, only contained 1.25% of silver for a weight that was 2.5 g at the end of this period. The fall in production was amplified by the withdrawal of the old, good coins. The bad, new coins drove the old good ones from circulation. Prices also exploded (Depeyrot and Hollard 1987).

In the 260s, the Empire was faced with the first inflation crisis in history; the Empire was invaded regularly or divided into independent regions, and emperors were regularly assassinated. The climatic crisis favored the first pandemics.

At that time precious monetary stock nearly vanished. A huge quantity of copper coins produced by official or illegal mints was in circulation. The monetary system was in danger.

The Crisis of the Roman Empire, Coins, and Units of Account

The monetary crisis of the Roman Empire was the first and the most important. Until then, new metal acquired by looting on various campaigns was constantly being added to the monetary stock. From the first issues, the stock was regularly increased. The contribution of mines was marginal, considering the size of the requirements of such a vast empire.

In a few decades, Rome took control of territories that had been minting for centuries and using silver coins as well as bronze, with many different metrological standards. The conquest of Greece (146 BC) was followed by that of the Carthaginian territories (Carthage was destroyed in 146 BC), Spain (second century for the central part, first century for the North-Western regions), the Mediterranean regions of Gaul (121 BC), Armenia (69 BC), and Gaul (52 BC). All these regions had their own monetary systems, completely independent or in relation to a formerly dominant monetary system.

The situation was even more confused for bronze (note that in this chapter we will use the term bronze for all coins made of bronze, copper, or a copper alloy). The

Roman mint stopped its production of bronze coins in 82 BC. A multitude of mints were producing bronze coins for regional or local purposes. In the East, this was only a continuation of the traditional issues. It was the same situation in the West, in Italy, Spain, and North Africa, etc.

After the end of the civil wars that followed the assassination of Caesar and the deaths of Marcus Antonius and Cleopatra, Octavian-Augustus began to restore the administration of the Roman Republic-Empire. A new monetary system was installed with (for the first time) regular issues of gold coins. His system was based on firm relationships between the coins: 1 gold aureus = 25 silver denarii = 100 bronze sestertii = 400 asses. The unit of account was either the denarius or the sestertius. The gold and silver coins were made from high quality metal (90–95% fine) and their weights were stable (7.9–7.7 g for gold coins, 3.8–3.6 g for silver coins).

The system was excellent, provided that the metal supply met the needs of the mints and of the economy. The first alert took place in AD 64 with a reduction of the weights of gold and silver coins, but the conquest of Dacia and the seizure of the treasure of Decebalus in AD 106 gave Trajan a new period of stability.

The slow decrease in the fineness and weight of the precious metal coins in the second half of the second century AD was undetectable by the population. However, the creation of the antoninianus by Caracalla in 215 (equivalent to the value of 2 denarii but containing only the silver content of 1.5) and its huge production during the reign of Gordian III (238–244) marked the beginning of a period of over-production of this coin financed by the melting of earlier denarii. After this period, inflation increased until the almost complete disappearance of silver in the antoninianus and the disappearance of lower value coins.

The fact that the population was aware of this inflation is clearly proved by inscriptions, the earliest ones dating to the reign of Gordian III (238–244). The more the quality of the coins decreased, the more prices increased, the more the silver content of the coins decreased and so on (Mrozek 2004). To end this vicious circle, the emperors of the late third century decided to adopt what we call "the Alexandrian system" that is, to separate the currency from the unit of account, in opposition to the traditional "Roman system" which linked the coins to a certain number of units of account (Depeyrot 1995a).

This "system" was clearly a characteristic of the Egyptian mint which continued to issues coins on its own standard even during the Roman Empire, until the end of the third century. But while the quality of the Roman coins decreased, the Alexandrian mint continued to attempt to maintain the standard of its coins. This is possible only if the coin and the unit of account are separate and if the number of units of account in a coin can be increased. In this way the consequences of inflation are limited. This avoids the recasting and reissuing of the coins and limits the effects of Gresham's law, as the quality of the currency is not affected.

So, the western monetary system inherited from the Greeks:

– The shape of the coins (round, standardized and struck, not cast and of various forms as in other civilizations)

– The use of precious metals
– The idea that a coin must be a human product (and not a "natural" object, such as the cowrie, used as mean of valuation)

In these points, Greek coins fitted the Aristotelian definition of currency (means of exchange, reserve of value, evaluation of wealth), but this definition was only a description of what was familiar to Aristotle.

So, the western monetary system inherited two important experiences from the Romans:

– The attempt to create a stable monetary system linked by a rigid relationship between the coins. Even if there were no marks of value on the coins, the coins were at the same time coins and units of account.
– The separation between coins and units of account, which is a characteristic of medieval to modern coinages up until the re-introduction of the valuation of coins in units of account (even if some coins began, after the fifteenth century, to present marks in relation to their unit. The best example is the ocho de plata which had an "8" indicating that its value was 8 real de plate. But at the same time exchange books underlined the fluctuating values of currencies).

This Greco-Roman heritage was the basis of the medieval and modern coinage and economy: the way in which coins were struck, the currency and the relationship between units of account and currency stayed the same until the mechanization of the mints.

The Restoration of the Monetary System in the Late Roman Empire (Third–Fifth Centuries)

From Anarchy to the Solidus (270–309)

At the end of the third century, the situation of the Empire deteriorated. Many regions seceded, such as Gaul and the East. The loss of territories meant the loss of taxes and revenues from these regions. Many armies were more involved in civil wars than in the defense of the borders. Every general promised gifts to ensure the loyalty of his troops. Generally paid in gold, these gifts were a heavy burden on the budget of the State. Fighting in the Empire meant that the legions left the frontiers and barbarians crossed them many times, adding looting to the general crisis.

From a specifically monetary point of view, the increase in the production of bad, so-called silver coins was so great that many provincial official mints were created in order to feed the local markets with coins. These mints were generally situated close to the main military camps in order to strike both gold coins for the soldiers and debased silver-copper coins to be used in local trade for the legions and the border cities. Apart from these official mints, hundreds of local mints, some of them official others illegal, were producing millions of light coins.

As inflation continued, Diocletian was obliged either to return to a kind of "Alexandrian system" (that is to say, to declare that currencies were floating in terms of units of account) or to reduce the cost price of each coin, in order that the legal tender was lower than the cost of production.

The Floating System of Aurelian

Aurelian (270–275) tried to restore the situation of the Empire. He defeated the main competitors and took over important seceded regions. Having restored the army and the administration, he restored the monetary system, the key element for the restoration of taxes and of the budget of the state.

The new system, created in spring 274, was based on a separation between the currency and the unit of account.

The fundamental change was the abandonment of the relationship between coins and units of account. Up to this time, there had been confusion between coins, their physical characteristics (metal, weight) or other attributes (a radiate crown meaning a double value) and the unit of account. Aurelian defined the value of a new coin (4.03 g) by the silver or gold content of each coin. For example, each bronze coin bore the mark XX or XX I which specified that the coin contained 1 part of silver and 19 parts of copper (5% silver); 20 bronze coins were equivalent to 1 pure silver coin. He also created a new gold coin of 6.5 g and that new gold coin explicitly bore the mark I L which meant that each coin was a 50th part of a Roman pound (Estiot 2004).

With this reform he protected the monetary system from any trouble caused by inflation. Whatever the change in prices, the legal tender of the coins varied according to the price of gold or silver. It was a kind of floating system, where the coins were mainly a fraction of a gold and silver pound. All earlier coins were withdrawn and their circulation remained marginal.

The Semi-floating System of Diocletian (274–305)

The restoration of the Empire was also the task of Diocletian (284–305) who stabilized the borders, reformed the army, and improved the imperatorial system in creating the Tetrarchy (2 senior emperors and 2 juniors who were to be their successors). From a monetary point of view, he tried to improve the system by reforming the coinage in 294. He lightened the gold coin from 1/50th of a pound to 1/60th (5.40 g) but created a new copper-silver coin weighing 10.32 g and intended to be almost the only such coin in circulation in the Empire: all previous issues were melted by all the imperial mints. He also restored the pure silver coin that had disappeared since the beginning of the third century. This new coin was based on the standard of Nero (54–68). It seems that this coin was the equivalent of half a modius of wheat (the common unit of wheat, about 8.7 l) (Depeyrot 1995b).

However, Diocletian partially turned his back on the principles of the reform of Aurelian. Again, he clearly associated his coins (except gold) with a certain amount of units of account. It seems that the gold coin kept its "floating" value, but the silver and copper-silver coins were clearly officially valued at a certain number of units of account. The gold coins were identified as a fraction of the weight of a pound of gold, bearing the mark of weight (Σ for 1/60th of a pound) and the silver coins were clearly identified by the mark XCVI (1/96th of a pound).

In the case of the silver-bronze coins, he decided to link them to a certain amount of units of account. When inflation rose again, he had to choose between floating and linking the coins. Instead of letting the legal tender float, he doubled the value of the coins, the silver coin changing from 25 denarii of account to 50 denarii of account (Aphrodisias edict, 1st September 301). Then, confronted by a spectacular increase in prices, he published the maximum prices edict in November–December 301. According to these documents, it is clear that the value of the coins was again doubled between September 301 and November–December 301. It is not known what happened after this. One thing is clear, the silver coins quickly disappeared after the end of the reign of Diocletian (in the 310s) leaving two coins, the gold coins, with a floating value (Depeyrot 1991) and the silver-bronze coins. Keeping the relationship (the "Roman system") between coin and unit of account, the weight of this bronze coin was reduced in 307, 309, 313, etc. Each reduction of weight was associated with a huge reminting of the old coins, allowing the state the opportunity of benefiting greatly from this manipulation.

The legacy of the period was the invention of a kind of "floating" valuation of the legal tender of the coins. Clearly, the value of the gold coins was estimated as a fraction of the gold pound and had no fixed value. On the contrary, the silver-bronze coin was definitively attached to a certain amount of units of account. Tariffed several times during the reign of Diocletian (294–305), its weight was reduced by his successors.

Now the monetary circulation was dominated by gold and bronze coins. The former was floating and the latter was attached to the units of account.

The Solidus and Its Floating System

The resignation of Diocletian began a long period of civil wars that ended only in 324 with the final victory of Constantine (emperor 306–337). It is not very useful to detail the vicissitudes of the bronze coinage which maintained a form of link between the coins and the units of account. It was a continuous cycle of reduction of the weight of the coins, due to inflation, followed by the introduction of a new heavier coin, being a multiple of the last issues, and so on. These reforms were associated with the withdrawal of the earlier coins, up to the 380s, when operation became impossible due to civil wars and invasions. The monetary stock was then composed of a mixture of all available coins, even obsolete ones, probably circulating according to their diameter and weight and no longer according to any monetary scheme (Depeyrot 1996c).

The silver coin, issues of which were nearly stopped soon after the resignation of Diocletian or were very limited, was again struck in huge quantities during the second half of the century (principally during the period 350–364). However, we must note that a very significant number of hoards of these silver coins have been found in Britain, perhaps in relation to the purchase of ore from silver mines with silver coins struck on the continent, perhaps related to mercenaries fighting during this period in Gaul. Whatever the reason, the silver issues disappeared in the upheaval following the great invasion of 406.

But the legacy of the fourth century, and especially of Constantine I the Great, was the introduction of the solidus, in 309 or 310. The creation of the solidus was the result of a failure, the impossibility of maintaining the weight of the gold coin created by Diocletian, the aureus struck at 60 to 1 pound (about 5.4 g), and the decrease of the weight of the coin to 72 to 1 pound (about 4.5 g). The weight of this coin (solidus) stayed stable throughout the century however, and in fact up to the eleventh century in Constantinople. Its name was transformed in all the subsequent languages (French sol or sou; Italian soldo; etc.). Its longevity can be explained by the quantity of coins in circulation, by the fact that many taxes were based on the solidus and its importance as a unit of account. During the fourth century, the bronze coins were extremely unstable and affected by inflation, as we have noted. Silver coins were in fact quite common in circulation as a subdivision of the solidus (this is why they tended to keep their weight and purity).

The fourth–fifth centuries were a period of intense legislative activity, particularly with the redaction of the Codex Theodosianus and subsequently the redaction of the laws regulating the barbarian tribes who invaded the Empire. This was also the period of the *latifundia*, large estates owned by landlords living in Rome and Constantinople who only wanted to guarantee their revenues. Above all, it was the period of restoration of the State, of the Church (or temples), the period when soldiers wanted to be paid in gold as a guarantee of their salaries and when the state tried to secure its revenues. All of them were interested in having a stable coin that could guarantee their revenues. Gradually everything became based on the solidus and no attention was given to the bronze coinage.

But to become the standard coin for the economy, the solidus needed to be produced in large quantities. It was achieved in three steps.

After the defeat of Licinius, Constantine, who ruled only the western part of the Empire, took the control of the East, which was richer than the West. Eastern treasures were used as gifts to the Christian Church (either in Rome, but also in Palestine, where Christ lived), to build his new capital, Constantinople, inaugurated in 330, to distribute donatives to his soldiers, and to increase the gold monetary stock. A short pamphlet written in the fifth century explicitly accuses Constantine of being at the origin of a centennial trend of inflation (Ireland 1979).

The second step was the decision made in 368–369 to improve the weight and the quality of the coins. The alloy jumped from 95% to 99% and light coins were not put into circulation. At the same time, the emperors decided to restrike all the previous

issues in order to homogenize the gold stock. So the revenues from the taxes, rents, wages, etc., were at the same instant stabilized.

The third step was the civil wars at the end of the fourth century but mainly the invasions that began in the winter of 406 when the Barbarians crossed the Rhine. The mints in Gaul had been destroyed or were unable to produce large quantities of coins, and the production of gold coins was the task of the Italian mints. They had to respond to the needs of the armies, not only to pay salaries, but also to pay for everything that armies need; they had to meet the needs of the imperial administration, but also to pay for fortifications and the security of the empire. One of the main expenses was certainly the salaries of mercenaries: many tombs or hoards from Germany contain a group of newly produced gold coins and necklaces that were the distinctive sign of a chief. And greater than these salaries were the gifts given to the Barbarians in order to avoid invasion, which represented a huge number of gold coins distributed in *Barbaricum*. To these expenses, we should add the looting of the major cities and the indemnities that had to be paid to the Barbarians to abandon sieges or to stop looting and withdraw. This was the case when Rome was besieged in 408 and looted in 410. So the huge production of gold coins did not create a huge monetary stock. Expenses, looting, and gifts to the invaders diminished the stock. To these Western problems (Eastern part of the Empire was not so affected by invasions), we must add the flight of the rich inhabitants of Gaul, Spain, Italy, who left the Western Empire to live in more secure Eastern regions (Amandry et al. 1982; Morrisson et al. 1985; Depeyrot 1996c).

The Western Monetary System in the Early Middle Ages (Fourth–Seventh Centuries)

Invasion and Settlement of the Barbarians (Fourth–Fifth Centuries)

Up until the mid- fourth century, the Barbarians invasions in the Roman Empire were mainly looting raids, very quick incursions with a lot of destruction. Then the Barbarians returned to *Barbaricum* with the booty. The chariots that sank in the Rhine at Neupotz contained more than one thousand objects from, bronze containers to cut pieces of silver vases, irons for slaves (and perhaps slaves themselves), coins, weapons, etc. (Künzl 1994).

After the restoration of the Empire and the renovation of its fortifications, these raids became less easy and, possibly under pressure from deterioration of the climate, the Barbarians tried to invade the Empire to find better land and livelihoods. In exchange, the Romans gave them land and the Barbarians became soldiers or auxiliaries in the army. The first agreement (*foedus* = treaty) was signed with the installation of Visigoths in Thrace in 295, but the main settlements occurred in 374 (the Alamanni on the west bank of the Rhine), 382 (Visigoths in the Balkans and in Gaul and Spain in 418), and 435 (Vandals in Spain and North Africa).

Officially the Barbarians were settled in the Roman Empire but kept their own traditions, languages, and rules. But their kings also adopted Roman official titles, acquiring a kind of semi-Roman organization. The illusion of the Roman

Empire continued during the major part of the fifth century: Barbarians copied Rome and aped Roman traditions, but considered themselves as subordinate to the emperor. In the monetary realm, they gradually decided to produce coins. Of course they imitated gold coins and not bronze coins (there were imitations, but these were individual productions). The first issues were copies of official coins. The very first coins were struck in the 440s by the Visigoths in Gaul with only some specific marks such as a supplementary crown, a dot. Mints were also opened to strike imitations in North-Eastern Gaul, and some *tremisses* in Southern Gaul. In Spain, the Suevi produced *tremisses* with a very typical design.

However, none of the Barbarians had the technology to produce coins with a high gold content and official style. The metal was in short supply and from the first issues the blanks contained less gold than official coins.

It was only during the second part of the fifth century that the Barbarians began to organize themselves into independent states. The most typical development was the promulgation of corpora of laws concerning all those who were members of one of these tribes settled in the Empire. This was the case, for example, for the Franks and the *Salic law* (around 500), for the Visigoths and the *Breviary of Alaric* in 506, for the Burgundians who also publicized (before 516) the *Lex Burgundionum* (Depeyrot 1993).

In the very last years of the fifth century, this development also affected the coinage. Instead of more or less secret marks used to distinguish the issuers, the barbarian kings began to imitate Roman coins by adding their monograms. Apart from this, gradually, the mints began to produce coins, and they chose to produce not the solidus (4.5 g) but its third part, the *tremissis,* a smaller coin weighing 1.5 g. Although its typology was based on a uniform design, with a bust on the obverse and a Victory on the reverse, the different mints also adopted specific ways of engraving the dies. It is thus possible to recognize products of the various tribes only by looking at the general features of the coins. These imitative coins were produced during the fifth and sixth centuries (Tomasini 1964).

All the laws also included references to coins. The Visigoths, who were the first to be settled in the Roman Empire, counted in coins but also in pounds and ounces of gold (as did the Romans), the others, more recently arrived, counted in *tremisses* (Depeyrot 1992).

The Semi-floating System

The "dark ages" inherited the "semi-floating" system of the late Roman Empire: a gold coin (and sometimes silver too) that was in fact a fraction of a pound of pure metal, and bronze coins whose legal value was fixed to a certain amount of units of account, but not inscribed on the coin.

The values of the precious coins were linked to the value of a pound of metal and were therefore "floating." The values of some other coins (silver or bronze) were fixed by explicit references to a certain number of units of account. They were "fixed." A system composed of "floating" and "fixed" coins can be defined as "semi-floating."

During the period from the last decades of the fourth century to the seventh or eighth century, this division was preserved by the Byzantine Empire and by the Barbarians who entered the Western Roman Empire, even when they became completely independent from Rome.

However, not all the Barbarians preserved or were able to preserve a completely "semi-floating system." Gold coins that diminished in fineness were no longer a fraction of the gold pound, the value of the gold pound, and the value of gold coins diverged and the gold coins eventually dropped out of the value of the gold pound: the value of a pound of coins was very different from the value of a pound of pure gold.

It is important to underline that the "semi-floating" system lasted longer in Mediterranean countries than in continental regions. The main distinction was between the Barbarians who were able to keep the quality of the gold coins (the Vandals in Africa, Ostrogoths and Lombards in Italy) as it was preserved in the Byzantine Empire and those who diverged

Dichotomy of the Coinage and Within Society

The dichotomy of the monetary system after the Great Invasions (from 406) corresponded to a dichotomy within society. On one side were the landowners, the imperial administration, Senators, the main officers, clergymen, etc., who were rich enough to deal in gold coins. On the other side was the majority of the population, agricultural laborers, and artisans, who dealt with a heterogeneous bronze stock of coins. As the amenities of the towns were reduced, the roads and aqueducts not maintained, and the population decreased, there was no significant need for new bronze coins. The administration concentrated on the preservation of the quality of gold coins: coins collected via taxes were melted down and the metal kept in ingots if there was no need for new coins. The borders between coins, ingots, and even jewelry were tending to disappear.

The period was characterized by a large number of hoards composed of a mixture of precious objects, including coins (though they were often kept separately), gold or silver vases, silver spoons (mainly in Western Empire), and gold or silver plates. The latter were either in the form of a *patera* (a shallow bowl used for libations in a Roman temple) or in a flat large form (convenient for dining use) (Depeyrot 2008a; Odobescu 2008).

Apart from these, the mints were still producing large medallions. There were no medallions in Republican times (up to 31 or 27 BC) and very few in the first and second centuries. During the third century, the mints began to produce spectacular heavy coins generally in gold, weighing 6, 8, 10, 12, etc., up to 70 times and more the weight of a normal gold coin. These medallions were given by the emperors on special days, such as the anniversary of their designation as emperor, in commemoration of a victory, at the beginning of the year, etc. During the fourth and fifth centuries, a large number of these coins were produced to be given to

mercenaries or to high officers or administrators of the Empire. Many of them have been found in *Barbaricum* (Bursche 1998).

The Semi-floating System: Vandals (Fifth–Sixth Centuries)

Specific mention must be made of the Vandals. In the fourth century, they were settled in the Roman Empire, in the region around Budapest (Aquincum), but were part of the group of invaders who crossed the Rhine in 406. They crossed Gaul (407–409) entered Spain (409) and Africa (429) and took Carthage in 439. They stayed in Africa until the Byzantine conquest.

It does not seem that the Vandals imitated gold Roman coins. Perhaps trade with Italy and the selling of wheat or oil was yielded sufficient coins. On the other hand, they produced small silver coins, perhaps after 470. Gunthamund (484–496) was the first king to add his name to silver coins which also bore a mark of value, 100, 50 or 25 denarii (or nummi?).

The Vandals produced a lot of bronze coins, but without copying Roman coins. Some were without any mark except for a cross, some were marked with D (500) or N (one nummus?) which could have indicated a value (1/500th of a silver coin?). All these coins circulated in the Western Mediterranean, including a many large bronze coins with a typically Vandalic type (a representation of Carthage) with the marks of value XLII, XXI, XII, and IIII.

They countermarked a large number of Roman bronze coins issued during the first and second centuries. These countermarks were mainly indications of value to include them in the Vandal monetary system. In general the mark was XLII.

The Vandals were defeated by the byzantine armies and the conquest of the region of Carthage ended their coinage in 533. Then, a Byzantine mint was opened in the city (Wroth 1911).

The Semi-floating System: Ostrogoths (Fifth–Sixth Centuries)

The very first issues of the Ostrogoths are not identifiable. It seems that Odovacar (476–493) copied official gold coins, while issuing silver and bronze coins in his name. Odovacar or his successors up to c. 534 produced a lot of anonymous coins, but bearing the marks of value XL, XX, X, or V (40, 20, 10, or 5 nummi). His direct successor, Theodoric (493–526), issued copies of Byzantine coins except for the very famous gold medallion (15 g) with his name.

Athalaric (526–534) struck silver coins with his name and without any mark of value, while he struck bronze coins with his name and the marks of value X or V (nummi). His successors continued the same kind of issues, but added the marks XL or X on bronze.

In 555, the last Ostrogothic armies surrendered, defeated by the Byzantine armies of the general Narses. This ephemeral kingdom was another good illustration of the semi-floating system (Wroth 1911).

The Semi-floating System: Lombards (Sixth–Ninth Centuries)

The Lombard domination of Northern Italy (568–774) lasted longer than that of the Ostrogoths. Lombard issues are easily recognizable by their distinctive style and manner of striking coins. However, for a very long period, they produced imitative coins bearing only the names of the Byzantine emperors and the first issues with a Lombard name began only in 688. The coinage was composed of *tremisses* and silver coins, some of the silver coins bearing marks of value on the reverse CN or PKE (read 250 and 125). Their weight and fineness constantly decreased during the period.

The Lombard principality of Benevento was the only one to produce coins. The first issues have only the names of Byzantine emperors and King Romuald (706–731) then introduced his mark. The coinage included solidi and *tremisses* and, under the influence of the Carolingians, Grimoald (788–806) added a silver denarius. None of the coins had any marks referring to units of account. In 899 the principality disappeared.

As some coins had an indication of value in units of account, this coinage was a semi-floating coinage (Wroth 1911).

The Floating System: Visigoths

After the invasion the Visigoths moved to Aquitania where they were settled in a treaty with Honorius (416). They produced imitations of Roman coins, lighter and with a lower fineness, in their capital, Toulouse. In 507, Clovis defeated them and pushed them out of Gaul, although they kept the Mediterranean coast in their control. They moved to Spain which was already a part of their kingdom and installed a new capital in Toledo.

All the Visigothic kings produced coins, mainly gold coins, but also very rarely some silver and copper coins. The first coins (sixth century) were derived from the Roman *tremisses* with a bust and a Victory, but the bust is generally seen from the front and the reverse looks like an insect, with multiple legs.

After the move to Spain, there were a lot of mints, up to 80, but most of them are known only from a few coins. The most important mints were at the economic or administrative capitals, Toledo, Seville, or Merida.

The general features of the coins were fixed by Leovigild (572–586): a bust facing right on the obverse and the reverse up to circa 650, when the reverse type was changed to a cross. The coins are large, stamped on large blanks with a very well-engraved legend. Each mint interpreted the general type and incorporated some variants, while keeping the homogeneity of the type (Miles 1952; Tomasini 1964; Grierson and Blackburn 1986; Pliego 2009; Crusafont and Balaguer 2013).

The Floating System: Merovingians (Sixth–Eighth Centuries)

Historians traditionally choose the reign of Clovis, King of the Francs (481–511) as the creation of the kingdom of France. In fact, Clovis increased his kingdom, first by taking control of the region of Paris, then by defeating the Burgundians and the Visigoths. From a monetary point of view, the Burgundian king Gondebaud (480–516) was already by this time striking gold coins with his monogram before his defeat by of Clovis. In Mainz, the Frank Theodebertus produced a solidus in his name. On the contrary, Clovis did not produce any coins in his name. There were already many Roman and Byzantine coins in circulation in Gaul, as we can see from the discoveries in the tomb of Childeric I, father of Clovis, who died in 481.

After the death of Clovis (511), his kingdom was divided between his sons. This division of the Kingdom was frequent, with an alternating pattern of division and reunification. During the half century following the death of Clovis, the mints continued to produce coins copying Roman types. Although their style, low purity alloy, and typological changes allow the Kingdom where the coins were produced to be identified, there was no explicit mark identifying the mint or the date.

If we consider the broad period, from the beginning of the sixth century (from Clovis 481–511) to the eighth century (the last king was Childeric III, deposed in 751), Merovingian coinage shows some general features that are characteristic of the period. There were two phases: the first characterized by the production of gold coins and the second by the production of silver coins.

The main point of the first period was a global decrease in the quantity of gold used by the mints and a complete transformation of the conditions of production of the coins. During the first years, many Romano-Byzantine coins were still in circulation. Their disappearance (either through a significant export to the East or by burying or even transforming them into jewelry) forced the king to open royal mints and to let local mints transform metal into coins.

In terms of the metal, the mints almost exclusively produced *tremisses*, except for some extremely rare *solidi* perhaps linked to special events. These coins were either copies of Byzantine coins or specific royal coins. However, Merovingian coins were lighter than the coins produced in the past by Roman mints or in contemporary Byzantine mints (between 3.5 and 4 g against 4.5 g). As a consequence, the weight of the *tremisses* was reduced in proportion to 1.2–1.4 g. If the weights were more or less stabilized, the composition of the alloys shows a quasi-stability up to the 620s (around 80–90%) and then a decrease down to around 25% at the end of the gold coinage (c. 675).

We can distinguish particular phases in the gold coinage. Up to 560, the coinage was composed of imitations of Byzantine coins, some of them with royal names; during this period the coins were still heavy and the alloy was good. In a second phase (560–585), the issues bore the names of the kings, but the system of producing the coins changed with the mention of the name of a *monetarius* (moneyer) on one side and the name of a place on the other. After c. 585, the weights of the coins

were stabilized at a lower level (c. 1.25 g), and a lower purity alloy, but this alloy continued to decrease in fineness until the end of the issues.

The last brilliant period was the reign of Dagobert I (622–639) whose chief of the mints was Eligius, who signed some issues in Paris and Marseille. The seventh century was the most confused time. Hundreds of mints produced coins bearing hundreds of names of moneyers and places.

The most difficult problem is that of understanding how the Merovingian system as a whole operated.

First it is clear that there was no unification of monetary production. There are some dominant iconographies, mainly based on the bust and a cross on the reverse, but this general type is often complicated by adjuncts, such as letters that designated the mint (although other mints imitated the same marks) and symbols. In many cases, the reverse types were completely different.

Second, the decline in the purity of the alloy, which is in most periods the best way to classify issues, is useless in this case. In the sixth and seventh centuries, gold Roman or Byzantine coins were imported through trade. Some of them were collected by the royal administration, and the mints that it controlled, such as Paris and Marseille, received these coins to produce new local coins, of lighter weight and alloy. A close examination of analyses shows that the purity of the alloy of the royal mints decreased but also that local mints were producing local coins by casting and debasing the royal coins, even though they were copying their types. There was therefore a double diminution: one affecting the royal mints casting Romano-Byzantine coins and the other affecting local mints casting royal coins.

Thirdly, we need to understand which "mints" were designated on the coins. The stylistic difference between the coins is so great that it is impossible to imagine that hundreds of mints were operating in France. In some cases, coins from different "places" are both struck with the same die. It is important to underline the correlation between the lists of bishops and coin issues. We have a list of bishops for each diocese. Many of them end in the seventh century: depopulation, plagues, or other diseases left the bishopric vacant. There is a strong correlation between the fact that the bishopric was unoccupied and the end of the coin issues. So we can suggest that they key to interpreting Merovingian coinage is not that of the production of large issues of coins for military, political, or economic purposes, but in the relationship between coinage and religious activities. Were the coins struck for gifts for local churches? Were the moneyers linked to local activity relating to the episcopacy? It is impossible to say. The huge number of "moneyers" (more than 1600) prevents us from viewing these people as specialized artisans, as in Roman times.

At this time, the coinage had no relationship to any unit of account. *Tremisses* were quoted in various legal taxes or fines, but it is impossible to imagine the circulation of a variety of coins of differing fineness.

In circa 675, the mints abandoned gold coinage and issued only silver. This ended a long era of widespread circulation of gold inaugurated by the transformation of the treasure of Darius into Alexandrian staters.

Was this shift from gold to silver a deliberate decision or a consequence of a lack of gold? It is impossible to decide, but it is clear that the final issues of gold coins were very limited. The weight of the first *denarii* was equivalent to that of the last gold coins. It is very plausible to suggest that silver was a means of continuing this gold coinage, and not a substitute for it. If there had been substitution, the weight of the first silver coins would have been greater, about 10 or 12 times heavier (such was the ratio of gold to silver).

The weight and fineness of the *denarius* rapidly decreased, from 1.2 g to 1.1 g and from 92 to 70%. The types were particularly diverse, but most of them are without any inscription, and only have figures or monograms. The mints were mainly located in central France, but the most active of these was Marseille, significant for its maritime trade and its position at the mouth of the Rhone (which was the end of the trade route from central France and even Germany and England).

Merovingian *denarii* continued to be produced by the mints for the first Carolingian issues, until the reform of 754 introduced by Pepin the Short, King since 751.

As is the case for the period of gold Merovingian coinage, we have no idea of the value of these *denarii* in units of account, nor of the ratio between the various coins (Tomasini 1964; Depeyrot 1998a, b, c, d, 2001; Cochet 2008a, b).

The Floating System: England (Sixth–Ninth Centuries)

The gold coins issued by the Franks and Byzantines circulated in England, but there was no significant issue in that region. It was after the mid-seventh century that large issues of silver coins were produced. These coins were either gold (or electrum) or silver, the latter are the most common. In general, silver coins were abundant and inspired by the last Roman coins. These coins lost fineness relatively quickly and their typology diversified, with the addition of new types, such as animals, legends, and people.

At the end of the eighth century, perhaps under the influence of Carolingian coinage, the English adopted a system close to the penny, with a new technology of production, a new weight standard, and a new general appearance. These pennies were directly inspired by the issues of Pepin and Charlemagne. The issues were very limited, and apart from the large and good quality pennies of the South, some local coins were produced in Northern England.

The production of these pennies continued through the ninth century. Without going into the details of the issues, let us state that the types were more or less standardized and that, as in Europe where the Carolingian *denarius* was the only coin, the new penny became the only coin in circulation.

At the same time, the Vikings were producing coins of their own in northern England.

All these coins were, as usual, both coins and units of account. We have no information on the relationship between the various coins.

The Semi-floating Byzantine and Eastern Coinages (Third Century BC–Tenth Century)

The Semi-floating System: Parthians and Sasanians (Third Century BC–Seventh Century AD)

After the defeat of Darius and the creation of the Seleucid Empire, a new dynasty founded by Arsaces I of Parthia appeared in Persia in the middle of the third century BC, (the Arsacid dynasty). They created what we usually call the Parthian Empire. They inherited from the Seleucid Empire the general features of a coinage based on the production of silver coins, with the bust of the king on the obverse and a representation of the king or an archer seated on the reverse inside a square formed by a long legend. The system was based on a drachm of circa 4 g with multiples and fractions plus a series of bronze coins based on a chalkos of circa 2 g with multiples. During the long Parthian period, the weight of all the coins decreased. There is no information on the relationship of the coins with any unit of account. As in the Roman Empire, the drachm was at the same time both coin and unit of account and circulated as part of the silver pound. The bronze coins were perhaps linked to the drachm by decree.

The last Parthian king was killed in 224 and the region came under the control of the Sasanians. As with their predecessors, the monetary system of the Sasanians was based on a silver coin and gold coins were very rarely produced, as well as some fractions. The drachm was a very large coin (up to 34 mm), thin, in good silver, weighing c. 4 g and bearing on one side the bust of the king and on the reverse the sacred fire, alone or with two attendants. This dominance of the silver coinage explains the issue of silver coins by the Byzantines for expenses during the wars in Transcaucasia. As for the Parthian Empire, there is no mention of the value of the coins in terms of units of account. The drachm was at the same time, as was usual, both unit of account and coin, and, as usual, part of a pound of silver. The quality of the silver stayed the same for the whole period. The bronze coins were perhaps linked to the drachm by decree, but they are so rare that their significance in coin circulation was marginal. The last king was killed in 651 by Islamic troops (Göbl 1971).

The Semi-floating System: Umayyad and Abbasid Coinages (Seventh–Tenth Centuries)

At the death of Mahomed (632), Muslims ruled mainly in Saudi Arabia, but their expansion was very rapid. Twenty years later they ruled most of the Middle-East including the whole Sasanian Empire.

During this expansion, the Islamic armies conquered many countries. Their local administrations were incorporated with some changes. For example, the mints continued to produce coins with very similar types, but with a

legend referring to the Koran. Each time a mint fell under the control of Islamic troops, the traditional type was adapted with the addition of a reference to the Koran.

By the end of the seventh century, Muslims were controlling a vast region from modern Egypt to Iraq, including Saudi Arabia. The reign of the caliph Abd al-Malik (685–705) began a period of reforms and harmonization, including a monetary reform.

Without going into the details of the reform that took place mainly in AH 76–77 (695–967), the caliph created a system based on gold (4.25 g), silver (2.98 g), and copper coins (generally c. 3.25 g, though the weights varied widely between regions). Numismatists have frequently discussed the relationship between gold and silver in the Mediterranean world. The gold stock had decreased in Western Europe, either exported to the East, or preserved in hoards, jewelry or precious gifts to the Church. The two main regions where gold was in use were the Byzantine and Islamic Empires. However, all the literature on exchanges and the relationship between the two series of coins is without interest as this ignores the real relationship between the two gold coins and even the ratio of gold to silver in each Empire. The system stayed the same during the Umayyad (661–750) and Abbasid (eighth–tenth centuries) dynasties.

Once again, we have no information on the ratio between the coins themselves except very isolated ones. The precious metal content of the alloys seems to be very high. Again, these coins were a standardized weight of pure metal (Walker 1941, 1956; Grierson 1960; Broome 1985).

The Semi-floating Byzantine System (Fifth–Seventh Centuries)

The division of the Roman Empire was initiated by the foundation of Constantinople in 330. However, for historians, even during the periods when several emperors ruled the Empire from different cities (e.g., Trier, Rome, and Constantinople), the Empire stayed unified up to the death of Theodosius I (395) or up to the reign of Anastasius (491–518), after the end of the reign of the last western emperor, Romulus Augustus (476).

In numismatic terms, there is no difference between the last issues of the Roman Empire and the first ones of Anastasius. However, he introduced a major reform of the bronze coinage (498?) that justifies the choice of his reign as the first of the Byzantine Empire. This new reform created several bronze coins bearing their legal value (40, 20, 10, 5, or 1 nummus, the unit of account), while the gold coin did not bear any mark of value (there were very few silver coins). This reform was the continuation of the system of the "floating" solidus and fixed legal tender for the bronze.

The evolution of the issues shows a tendency in the sixth century to increase the number of mints, but the production of gold coins was concentrated in Byzantium, even if some (up to 12 in some periods) Western mints (Carthage, Sicily, Rome, . . .)

and Eastern mints (Antioch, Kherson, . . .) operated. If the weight of the solidus was stable, light solidi were produced. These light coins were mainly used in exchanges with Barbarians. Although the solidus does not bear any mark (of value, or weight), the light ones bear the mark OB *+*, OB +*, OB XX, designating coins of 22, 21, 20 siliquae in weight (4.1, 3.9 and 3.8 g); they were produced up to the reign of Constantine IV (668–685).

An important change was introduced by Heraclius (610–641) (if we consider the Christianization of the types as only of iconographic significance) with the introduction of a very heavy silver coin (6.8 g) in 616, the issue of which was financed by the State treasury and after 621 by the Church. This coin is mainly found in hoards on the borders, along the Danube or in modern Romania and in Transcaucasia (Georgia and Armenia). These two regions were zones of battles and conflicts (against the Bulgarians and Slavs or against Arabs), and these coins were surely linked to the salaries of soldiers and the cost of armies. They were struck up to the end of the seventh century.

The weight of the bronze coins decreased throughout the sixth–seventh centuries (Wroth 1911; Morrisson 1970a; Yannopoulos 1978).

The Floating Byzantine and Eastern Coinages (Tenth–Fourteenth Centuries)

The Floating System: Disaggregation of the Islamic Coinage (Eleventh–Twelfth Centuries)

The disaggregation of the Spanish Caliphate began in the tenth century and was marked by the capitulation of Toledo in 1085. Spain was divided into many small territories. The large Muslim Empire was divided into several more or less independent entities, in North Africa, in Egypt, and in Arabia. The phenomenon was similar in the East, with the Tahirids (821–873), Samanids (819–999), Ghaznavids (962–1187) (between Iran and Pakistan), and the Buyids (945–1055) (eastern Persia and Iraq). The Seljuqs (1037–1194) ruled most of the region between Pakistan and Greece, including Iran, Iraq, and Transcaucasia. In Egypt, the Ayyubids reigned, the most famous being Yusuf Bin Ayyub (Saladin). The organization of the states was changed after the Mongol invasion at the very beginning of the thirteenth century (mainly under Gengis Khan, 1206–1227, and Ögödei, 1227–1241).

The slow disintegration of the Caliphate was accompanied by increased pressure from the crusades. In fact, the crusades acted as a barrier between the Islamic world and Western Europe.

All of them continued to strike gold and silver coins, and of course copper. Nevertheless during the last years of the eleventh century and the first half of the twelfth century many silver mints closed, even though new mints

opened in the Middle East. The coins differed from one region to another. All the coins were accepted according to tradition and decrees (whichever form they took). There is, as far as we know, no information concerning the relationship between coins and units of account, even between Islamic and Western coins (Broome 1985).

Byzantium Moving Back to the Floating System (Eighth–Twelfth Centuries)

Slowly the monetary system began to develop in a different direction to that of the Roman Empire. Various influences were at the origin of this evolution. The creation in 695/696 of the silver dirham (c. 2.85 g) explains the creation of the new silver coin the miliaresion (2.25 g) under the reign of Leo III (717–741).

The Italians mints produced a coinage adapted to the local tradition (based more on *tremisses* than *solidi*), and they followed the decrease in the quality of the Lombard coins. The gold produced in Italy became lighter and lighter and their precious metal content decreased.

In the mid-ninth century, the system was reformed. The mark of value on the reverse was abandoned: the old copper coin struck since Anastasius, the weight of which decreased regularly, was replaced by a new coin produced in many mints. The weight of the silver coin was also modified, either becoming lighter, or heavier. The reform of the bronze coin (a new, light coin) under the reign of Theophilus (829–842) was a return to the old Roman system (first to second centuries) of coins without marks of units of account. The relationships between the coins were clarified by tradition or by decree.

During the tenth and eleventh centuries, a huge number of anonymous bronze coins were produced by many mints. At the same time, gold coins began to lose weight and fineness: a new coin was produced under Nicephorus II (963–979) to replace the old gold coin.

The Western Byzantine issues had already abandoned the pure gold solidus in the eighth century. The same development occurred in the Empire during the tenth century.

The gold fineness that was maintained in Constantinople collapsed during the 2nd part of the eleventh century, falling from nearly 99 to 33%. The silver fineness also decreased and the weight of bronze coins fell. At the beginning of the reign of Alexis I (1081–1118), the old gold coin was in fact a silver coin. The monetary system was collapsing.

The decision of Theophilus (829–842) to stop the production of coins with marks of value and the introduction of debased gold coins severed both the relationship between the bronze coinage and units of account and also the relationship between the gold coin and the pound of pure gold. After this period, the coinage became unstable; legal tender could be changed by an imperial decision.

The Byzantine Floating System (Twelfth–Fourteenth Centuries)

Alexis I introduced a reform of the currency in 1092, reintroducing a gold coin (90% fineness), and electrum, silver, billion, and bronze coins. However, the reign of Alexis corresponded to the beginning of the first crusades (the first one began in 1095), and it is difficult to separate these monetary changes from their military context: arrivals of westerners, cost of the wars, sieges of cities, etc., all these activities were expensive and needed a large monetary stock to meet these demands, especially in bronze. The monetary system stayed more or less stable during the Komnenos (last reign ending in 1185), but immediately the weight of fine metal and the weights of all the coins decreased again, in relation to revolts and the cost of armies, etc.

The pillage of Constantinople by the Venetians and warriors of the 4th Crusade on 17 July 1203 heralded the end of the Byzantine Empire. It split into feudal principalities: Nicaea, Thessalonica, and Trebizond. After a long period of destruction (there were no issues from 1203 to 1261), the Palaiologos dynasty ruled (from 1282 to its end in 1453). The debasement of the coinage continued, some issues copied Venetian issues (Michael IX, 1295–1320), and gold coins ceased to be issued between 1341 and 1391. During the fourteenth century, few gold coins were produced, with some good issues of silver coins. However, after 1203 the output of Constantinople was much reduced, the monetary stock being fed by Italian coins, especially those of Venice.

In substance, the reform of Alexis I in 1092 tried to restore a stable monetary system. It worked for 80 years, but military expenses, the division of the Empire and raids ruined Constantinople. Inflation and financial problems were at the basis of the competing debasement of monies and the revaluation of the legal tender of coins in terms of units of account.

After the 4th crusade and the pillage of Constantinople in 1203, several principalities appeared and were ruled as independent states.

The empire of Trebizond (1204–1461) was established after the pillage. The mint was very important in relation to the commercial position of the city. The mint produced huge quantities of silver, billion, or bronze coins. Their weights decreased slowly from the beginning to the end of the empire. There are a lot of marks identifying the mints, but no marks of value.

The empire of Nicaea (1204–1261) was also established after the pillage. The mint struck few gold, silver, or bronze coins.

The empire of Thessalonica was established from 1222 to 1246, also after the pillage of Constantinople in 1203. The issues were silver and bronze "gold" coins.

So, all these territories produced floating coins, without any mark of value. The fact is that coins were regularly debased: either the fineness or the weight was reduced sometimes both (Wroth 1909; Morrisson 1970b).

Carolingian Restoration (Eighth–Tenth Centuries)

The Carolingian period was the true end of the Roman Empire, one of the major breaks in European monetary history. We can consider the Merovingian period as a kind of continuation of the Empire: it had the same monetary system based on gold, with a weight more or less based on the *tremissis* (1.5 g). The typology was similar, with a bust on the obverse. Even the shift from gold to silver in c. 675 did not change the organization of the currency: the bust stayed in the same place; it was only another stage in the disintegration of the Greco-Roman monetary system and typology.

With the Carolingians, for the first time a new monetary system included a vast part of Western Europe, France, a large part of Germany, Switzerland, and a part of Italy and of Spain. The "grammar" of the coin was changed; there was no longer a bust on the obverse (except on some very rare issues) but a monogram. The technology was improved, with large enough blanks to show the complete external circle of beading, while it was absent on Merovingian dies or off the flan. And at last a new monetary system based only on silver. Typology, weight, alloy, legend, all the characteristics of a coin were the same in all the mints. After a period of really messy issues, the monetary system was again homogeneous.

We cannot separate the restoration of the monetary system from the complete renovation of the State by the Carolingians. It was part of the new foundation of the taxation system. It is impossible to separate the monetary restoration from improvements in administration, the creation of a new way of writing (Carolingian minuscule), and the huge efforts made to inventory the Empire: cartularies are the most visible evidence of this. With Charlemagne, we find the same approach as Diocletian: administration, taxation, regulation, and monetization. Both of them wanted to regulate their whole empire by decrees, the Maximum Prices Edict in the case of the former, cartularies, and *urbarium* in the case of the latter (Duby 1962; Fossier 1978).

One of the aspects of this Carolingian State revolution was to produce a significant body of legislation concerning daily life. Produced mainly by Charlemagne and, in lesser part by his direct successors, some of the edicts concerned the currency. They limited the production of coins to the royal/imperial official mints and condemned the production of forgeries. As the typology of the coins changed from time to time, the legislation included edicts concerning the demonetization of the old issues and the obligation to accept the new coins at the official value. The aim of Carolingians was to restore new money but also to control the monetary stock strictly by forbidding any addition of illegal coins or any contestation of their value. It is clear that this legislation was influenced by Roman tradition and laws. Their edicts are very close to those collected in the Codex Theodosianus or Justinianus. Some very rare coins bear an imperial bust copying the usual representation of Roman emperors whose coins could still be found in circulation or in Roman ruins. Even the choice of silver for these issues was inspired both by the abundance of the metal but also by ancient tradition.

After the long Merovingian period characterized by complete disorganization of the mints, the kings/emperors limited their number to reinforce their control. For example, about 40 mints were active during the reign of Charlemagne to Louis the Pious (768–814 and 814–840), but their number increased to 130 during the reign of Charles the Bald (840–877) and then decreased to 50 at the end of the ninth century.

A characteristic of the Carolingian issue was a production based on one coin, in silver. Its physical presentation was perfect, a large, round blank, perfectly struck, with an easily read legend and with a large monogram or a clear cross. Apart from this denarius was a light silver obol (half-denarius). A very small issue of gold coins was also produced. The Carolingian coinage was based on silver; it was a silver-standard system.

The type of the denarius was changed from time to time, in order (officially) to limit the number of forgeries. This was the case in 793/4, 840, 864, etc. The quantities issued were very great during the mid-ninth century in relation to the Viking invasions and the sums paid to them.

As the Carolingian empire included France, the borders of Spain, Italy (up to Rome and beyond), Belgium, Germany, Austria, Switzerland, and the Netherlands, and was "controlling" a large group of vassals (Serbians, Croatians, Pannonian Avars, Moravians, Czechs, etc.), and covering a large part of the territories up to the Oder and Danube, the restoration of the currency must be considered as the creation of the first European coinage and the first European monetary unification.

The Carolingian period was also when a new system of account appeared. The old monetary system had dissolved during the crises of the later Roman Empire in the West or during the various crises of the Byzantine Empire. In 793/794, Charlemagne created a new scale of values based on a pound (unit of account), composed of 20 sous (unit of account) and composed of 12 *denarii* (coin). This was the monetary system used in fiscal documents and in the *urbarium*. This system stayed in use for example up to 1971 in the United Kingdom where 1 pound = 20 shillings (modern name of the sou), 1 shilling = 12 pence (modern name of the *denarius*, the penny).

However, throughout this whole period, none of the coins had any indication in terms of units of account. We have no suggestion on the relationship between the coins and the units of account. However, the silver coin decreased in weight and fineness in the first years of the tenth century. Even though the precious metal content of the alloy seemed to stay at a high level, the weight was reduced several times after the great reform of Charlemagne. The denarius of 752–793/794 weighed c. 1.12 g, but the reform of Charlemagne in 793/794 increased the weight to 1.7 g. It decreased to 1.48 g in 822 and there was a reduction of weight in circa 910 to 1.20 g. What was the relationship with the unit of account? Clearly the reduction of weight was a consequence of a fixed relationship between the denarius and a unit of account. It is possible to think that the decline in fineness was the consequence of inflation (or a need for coins, which is the same thing, as in this case the increasing need would have increased the value of the coin).

The Carolingians recreated a currency based on a fixed relationship between units of account and currency without any indication of value on the coin itself. This conception of the coinage lasted for most of the Middle Ages.

In the ninth century and particularly in the tenth century, the capacity to control the mints decreased. Many mints either closed due to a probable lack of metal or shifted under the control of local landlords who produced, at first, coins copying the royal/imperial coinage. In some cases, the Carolingians granted the right to issue money to particular abbeys or feudal authorities.

This development was more evident in Eastern Europe (East of the Rhine) than in Western Europe. In Germany, Saxony, and Bavaria, etc., local rulers decided to produce coins with their name or their mark. The whole of Western Europe began to follow the Carolingian system and began to produce coins. This deep transformation involved all regions, in Bohemia, in Ukraine (where Byzantine coins were imitated), but also in Scandinavia and Poland, where coinage began to appear.

In England too, issues became larger and larger, and the Norse Vikings produced coins in Ireland. In c. 975, the English coinage was unified and all the coins had to be similar, with a bust on the observe and a cross on the reverse.

The contribution of Carolingian to Europe was significant. They restored a monetary system based on silver and they imposed it on most of Europe during the ninth century. The Viking raids and the tribute paid in coins by Carolingian abbeys, cities, or kings also contributed to the wide distribution of the coinage.

The disintegration and fragmentation of the Empire gave birth to a great number of principalities. Some of them quickly introduced their own local coinage. Others (in the West) first produced imitations of official coinage. Both contributed to the new monetization of the economy. However, the stock lost its homogeneity. The denarius, the coin and unit of account, became a generic word describing different coins of various weight and fineness. Each coin was to be estimated in local currency if necessary (Grierson 1976, 1991; Depeyrot 2008b).

Western Floating Re-monetization (Eleventh–Thirteenth Centuries)

If the tenth century was a period of disintegration of the royal monetary right, the eleventh century was the period of liberation of many mints from royal control and the twelfth century was a period of increase in the production of coins. The development of trade (local, regional, and international), urbanization, large fairs, and demographic expansion were the cause of the growing need for coins.

In France, the main feudal lords confiscated the right to produce coins. In general, these first issues copied traditional types, introducing some new legends or abbreviations. In general, the fineness of the coins decreased throughout this period. The situation improved during the twelfth century. The role of the royal coinage was marginal, but its importance increased along with royal power and the number of territories regained by the crown. However, during this period, some mints were so significant that they had a major role in circulation, for example, the mints of Melle, Melgueil, and Toulouse.

In England, the coinage was still organized according to uniform types that were regularly changed, each change being linked to the debasement of the coins, up until

the mid-eleventh century, when it was stabilized for at least two centuries. Royal issues continued during the twelfth century with some typological changes over time but keeping their fineness. Some barons also produced coins. The neighboring countries of Ireland and Scotland similarly produced coins for a society that was increasingly monetized.

In Germany, monetary rights shifted to the main ecclesiastical institutions, but also to lay institutions. This transfer was facilitated and increased by endless wars and concessions of monetary rights were a way to finance military operations. During the twelfth century, the production of coins increased. Some mints were producing very thin uniface coins (bracteates), while the others were producing small biface denarii. As the royal institution disappeared, these issues were mainly the result of the activities of local, feudal, or ecclesiastical mints.

In northern Europe, the plundering of Europe by the Vikings and the arrival of huge quantities of silver dirhams from Islamic countries via the Don and Volga valleys facilitated the development of a monetized economy in Scandinavia, but also in the regions dependent on the Vikings. Following this trend, mints were created and coin circulation increased in Russia and Poland. In the twelfth century, the Scandinavian kings similarly produced silver coins. Depending on the kingdom and the mint, the coins were large uniface coins or small biface ones, from 1 g to 0.25 or even 0.17.

In the hinterland, in Poland and Russia, the production of small denarii increased too, with some issues of light coins (down to 0.4 g). Between Germany and Russia, principalities also began to coin. This was the case for Hungary where small denarii and fractional coins were issued.

Silver was lacking for the monetary issues in Europe. This penury might have had several causes, perhaps a decrease in the production of mines, perhaps also a sharp increase in the need for the production of coins.

The situation also changed in the Mediterranean countries.

In Northern Spain and Italy, the coinage was under the control of the Spanish kings, or issued by feudal lords in Barcelona. The production of royal coins increased along the Reconquista. In the twelfth century, the Portuguese began to issue coins. The important trade with Muslim countries pushed both the Spanish and the Portuguese kings to strike gold coins with Arabic legends. For example, after 1180 the mint of Toledo issued gold “Islamic” coins with Christian legends in Arabic. For internal coin circulation, the mints were producing denarii following the traditional features of western coinage.

In the twelfth century, the North Italian coinage was in the hands of the Germans in Italy and became more and more abundant with many civic issues bearing the name of the German emperor. The mints of Pavia, Luca, Verona, and Milan were the most active. But two others, Genova and Venice, produced large series of coins. All these coins were very light, for example, 1 g for Pavia (40% fineness) or 0.35 g (25% fineness) in Venice.

Spanish armies took Toledo in 1085. The Southern Spanish coin circulation was under the control of Muslims who began to produce gold coins in Cordoba in 919 and the Norman coinage (Italy, Sicily) imitated Islamic monetary stock.

In 1059, the Norman Robert Guiscard received a large part of southern Italy and Sicily from the pope. The first Norman coinage (in Italy and Sicily) imitated Islamic monetary stock, but during the twelfth century the development of trade with the East and North Africa encouraged the Normans to produce a huge number of coins imitating gold and silver Islamic coinage. They also produced a large number of bronze coins, as was typical for the Mediterranean world.

The Western Mediterranean world was once more a hub for international trade. In Spain and Italy, the coin circulation consisted of a mixture of silver and gold coins, including coins of different origin. Southern mints produced gold coins to compensate for the lack of silver coins and in order to increase trade with the Muslim world, which used gold dinars. In Spain, these gold coins copying dinars were produced in Barcelona.

The situation changed in the Eastern Mediterranean world as a consequence of the crusades. In taking Jerusalem, the crusaders met a monetary stock composed of gold and bronze coins in large quantities, which had been issued for centuries. Silver was present, but unlike the situation in the West, it was not the only metal used for coins. Immediately the crusaders produced copies of gold and copper coins in the main cities under their control. These coins were produced not only in the East but also in Cyprus. Trade and the passage of crusaders returning home contributed to the popularization of gold and bronze coins in Western Europe. The first crusade (1096–1099) was followed by two others in 1147–1149 and in 1189–1192.

The eleventh–thirteenth centuries are a key-period characterized by an increase in coin production. In those regions where the tradition of coinage was not severed, the number of active or very active mints increased. This monetization now incorporated new regions where coin production had been limited. This was the case for Central Europe and Russia, etc.

All these coinages were valued according to the local economic, political, and social situation. In some countries (e.g., England), the king was powerful enough to maintain the homogeneity of coin production. But in general the coinage suffered from chronicle instability in space and time. This meant that the coins produced at the same time by different mints could have had a very different alloy and weight and that the mint could have periodically reduced the "quality" of their output.

How could such a system work? Analysis of contracts for the production of coins shows that the mint masters had precise information about the weight and the alloy of the coins. These details made it possible for anyone to evaluate the fineness of each issue, which was easy to recognize by the legend and type.

So the system worked by taking into account the coin but also its mint. These two pieces of information were enough to assess the value. There were as many units of account as there were denarii, as the value was estimated at a certain number of denarii or sous (1 sou = 12 denarii) from that mint. It was also possible to refer to another unit of account that was a pound of pure silver. In this case, the number of denarii varied according to the silver content of each coin: the lower the fineness, the higher the number of coins needed to obtain a pound of pure silver. The weight of the pound also varied from one place to another. The explosion in the number of mints implies a complete separation between the coin and the unit of account.

The relationship between the coins themselves was another reason for the generalization of the units of account. Only units of account can provide equivalence between gold, silver, and billon coins with very differing metal contents and weights. The unit of account was therefore a key element in vertical exchanges (from gold to other metals) or horizontal exchanges (between coins from various regions).

The disintegration of the Carolingian empire also marked the end of its monetary unification (Depeyrot 1987; Grierson 1976, 1991).

The Canon Quanto

The systematic debasement of the coins was a problem for the economy. If one borrows a sum in coins, the sum will be valued in units of account or even in a certain number of denarii, generally whose type is specified. A problem may arise when the borrower wants to repay the sum and if the number of units of account in each coin has been increased or if the fineness of the coins has been reduced.

Of course, in such circumstances the borrower will save some money, but the lender will lose part of his money.

Debasement was used by kings and states, etc., to reduce public debts. The issue was publicly raised in 1169 concerning a debt that was to be refunded in new debased denarii. After this issue was raised, the problem remained unsolved, but in 1196 the Emperor of Germany, Henri III, admitted that the bishops were allowed to debase their coins. To avoid too many reforms of the coinage, a letter of Pope Innocent III, dated 5 April 1199 (called canon quanto), decreed that the king (or another authority) could reform his coinage only once in his reign. This letter was included in the Decretales of Gregorius IX. All subsequent legislation was based on this decision, at least during the twelfth, thirteenth, and fourteenth centuries.

The Revival of a Monetary System (Thirteenth–Fourteenth Centuries)

The thirteenth century was clearly one of the best periods of the Middle Ages, in the continuation of the twelfth century. Improvements in agriculture, demographic growth, urban development all favored the development of monetization of the economy. To them, we must add the numerous crusades that brought many expenses, from weapons to transfers of money across the breadth of the Mediterranean Sea. New Latin states were created to be used as rear bases and trading posts: such was the case for Cyprus and the Armenian Kingdom of Cilicia.

The 4th crusade (1202–1204) was marked by the pillage of Constantinople and the transport of booty to Venice and more generally to Europe. The following crusades (5th in 1217–1221; 6th in 1228–1229, 7th in 1244–1249, 1251, and 8th in 1270) played the same role, bringing silver from Europe to the East, employing local mints to cover local expenses, and sending back precious metal to Western

Europe. We cannot forget the size of ransoms, the most famous being the ransom of Richard I of England (100,000 pounds, some 33 t of silver).

In such a context, the various small and light denarii were not convenient for all these expenses. The monetary system and medieval society both needed to develop. Slowly the mints began to produce gold coins and heavy silver coins. Both were ways of increasing the value of exchanges; the gold coins were the result of harmonization with Islamic world. The new silver coins (groats) theoretically had the value of 12 denarii, but the fineness of the denarii continued to fall, so the value of the groat increased. The groat was, as all the other coins, both coin and unit of account, as many amounts were estimated in groats; but its value in denarii was floating and changed according to the type and issue of the denarii.

This situation continued in the fourteenth century, with an increasing number of mints and types of coins. Still infrequently produced at the end of the thirteenth century, gold coins were minted in nearly all the important states of Europe. Each minted according to its own standard, introducing great heterogeneity into the stock.

The first issue of groats was produced in Venice for the 4th crusade in 1202 (the previous issue in Milan in 1190–97 had been very limited). The groat (2.1 g 95.5%) was the counterpart of 24 denarii (at this time it was only 0.35 g 25%). In few years, it became the principal coin in Italy and nearly all the Northern mints issued copies of the groat, or similar coins of the same weight. Gradually, other mints created heavier coins; for example, Rome produced 3.5 g silver coins after 1253.

Following the restoration of the silver coinage, the mints also began to issue gold coins, the production of which had ceased for centuries. In 1252, the two rival cities of Genoa and Florence began to mint gold coins (the genovino and florin, 3.5 g), and in 1284 Venice began to mint one of the most famous gold coins, the ducat.

The production of gold and silver coins continued during the fourteenth century and the number of mints increased. To these units were added divisional coins.

Southern Italy had already adopted gold coin a long time ago. This was to be expected, as Sicily was the rear base for the crusades and for trade with Islamic countries that were using gold. In 1231, Frederic II decided to mint heavier gold coins. In 1278, the bad denarii and all the gold and silver coins were replaced by new gold and silver coins minted according to the Italian standards. However, the gold coins became very rare. Silver coins, which had been rare, became common. After the political change from French to Aragonese domination (the Sicilian Vespers) the production of gold coins nearly ceased, until the fifteenth century. After 1309, the mint struck the silver gigliato (3.9 g) which was imitated by the major mints around the Mediterranean Sea until the fifteenth century, becoming one of the most popular coins.

In France, feudal coinage was dominant, with a great diversity of types, alloys, and weights. However, during the thirteenth century the territories and the power of the King increased. He gradually imposed royal coinage: all his mints were producing the same type of coins. The imitation of royal coinage was strictly forbidden and the circulation of local issues was limited to the feudal domains. In the mid-thirteenth century, the King conquered the southern part of France, opening access to the sea.

From a monetary point of view, after 1266 the King minted a heavy groat of 4.22 g (95.8%) that was an immediate success. He also decided to mint a gold coin (4.2 g) to be used for the expenses of the 8th crusade in the Easts (1270). This issue was stopped after the death of the King, but in 1296 a new coin was produced (7.09 g) and its production lasted for the whole fourteenth century. Although the King tried to maintain the good quality of gold coins, the stability of silver did not last: the thirteenth century royal denarius contained 0.335 g of fine silver, but only 0.26 g in 1295 and 0.12 g in 1303. The value in denarii of the groat increased in proportion, destabilizing the whole monetary system.

In England, the production of coins increased and was matched by a change in the types of the coins. There was a general recasting in 1205, then in 1247 and 1279. This last reform created a new denarius that was produced for two centuries and added a silver groat (5.8 g) and a gold coin (2.92 g), but both of these were abandoned very quickly. The scale of coin production and its wide distribution can be related to the development of trade and to the crusades (the ransom of Richard being the most famous episode).

These developments were the same in the Netherlands: the circulation of English coins led the mints to adopt a new standard by issuing heavy silver coins after 1269. The most important states were all producing groats at the beginning of the fourteenth century and in the 1320s; the mints also produced gold coins, first florins, then French gold coins of various types. All these coins circulated alongside a huge number of different uniface or biface denarii.

In central Europe, the Tyrol also shifted to the production of a coin of a 1.2 g silver coin. In other regions, the mints continued to strike small denarii. The German mints produced large silver coins.

The situation of Northern (e.g., Scandinavia) and Eastern Europe was not excellent. Local wars prevented any desire to transform the currency. The Scandinavian countries, Hungary, Russia, and Poland more or less continued with small biface or uniface denarii.

The principalities of central southern Europe minted imitations of Byzantine coins. Greece received coins from all the western mints and began to strike local imitations of French types.

In the East, the Armenian kingdom was one of the bases of the crusaders. They minted silver coins, some of which were bilingual, with Armenian and Arabic legends. Cyprus did the same: its first coins were imitations of the dinar and silver coins, but at the end of the century the mint began to produce silver groats.

In the Iberian Peninsula, the monetary stock was composed of traditional denarii and imitations of Islamic coins. In Spain, various issues of denarii increased the complexity of the monetary stock. Although Barcelona continued to strike good silver coins, the mints generally produced base denarii.

The thirteenth century was an important period of transition. The traditional denarii were losing their fineness. The weight of silver in each denarius decreased and the diversity of the monetary stock increased. There was no unity, each coin being locally at the same time both coin and unit of account.

This confused situation forced the states to issue a good pure silver coin which was a multiple of all the billon light denarii. The new coin was a success and progressively the main kingdoms minted a new heavy silver coin, the value of which was also to be estimated in units of account, as its value was floating in denarii, or as a fraction of the gold coins. The regions that were out of the main commercial streams stayed with the traditional denarii.

In fact, the mints not only issued silver coins, but also made gold coins. International connections, the crusades, and economic development made it essential to produce a coin similar to those in use in Islamic countries. All the countries involved in this East-West relationship adopted gold. The best example is the French King Louis IX who ordered the mint to strike gold coins in preparation for his crusade, but died in Tunis. After his death, these gold issues were suspended.

At the end of the thirteenth century, the revival of a multimetallic coinage was complete. The control of the mines in Central Europe was an essential factor.

However, the monetary stock was heterogeneous and there were no coins bearing an indication of a specified number of units of account. All estimates of coins in terms of units of account were decreed.

If the post-Carolingian period was one characterized by a slow decrease in fineness and a kind of "hidden" inflation, the development of the coinage after this period was characterized by an association between the type of the coins, their intrinsic value, and a certain amount of coins to the theoretical pound. This association between the type and a certain amount of units of account was at the base of the monetary system until the appearance of coins with marks of value (Grierson 1976, 1991).

War and the Black Death: An Opportunity for Monetization (Fourteenth Century)?

The Black Death of 1348–1349 affected the income of every kingdom. This epidemic was compounded by the Hundred Years War which begin in 1337. To the destruction caused by the war and revolts were added the cost of fortifications, armies, and mercenaries and the huge amounts of money paid for ransoms.

To finance these operations, the mints regularly changed the weights or the composition of the alloys of the coins, adding confusion to confusion.

These difficulties were aggravated by the issue of metal supply. An increase in coin output would only be possible if the mines were able to increase the production of ore, but the need for silver and gold coins was perpetually increasing: the ransom of the King of France Jean II (1360) amounted to 13 million gold coins (about 13.5 t); the output of the mint of Florence in 1344–1345 required 1.2 t of gold. The bulk of the metal came from Central Europe (Hungary) and also from Africa, but this supply of metal was decreasing and hazardous.

The difficulty of finding metal and so producing good coins favored the production of imitations. The well-known gold florin was imitated by many mints all over Europe. This was also the case for all the main coins, whether English, French, from

Belgium or the Netherlands. It is possible to identify about 150 European mints that specialized in producing copies of official coins and to this should be added all the illegal forgers fabricating unofficial forgeries.

In France, the Hundred Years War obliged the King to produce numerous issues of gold or silver coins, each one with a specific design, weight, and alloy. The successors of Jean II, Charles V (1364–80) tried to maintain the quality of the currency by limiting the number of reforms and changes to coin production. This revival was certainly inspired by Nicolas Oresme whose analysis of the monetary system concluded with the necessity of a return to stable money. This needs to stabilize the coinage led to the adoption, in the second part of the fourteenth century, of a system of secret dots on the coins: placed under the letters of the legend, their position designated the mint. Throughout the fourteenth century, the King tried to limit the importance of feudal issues. They were more and more limited to their possession. A decree of 1315 listed only about 30 mints in the kingdom. In other regions (Bretagne, the Kingdom of Arles, Savoy, etc.), feudal production continued (Dupuy 1989).

The situation in Italy was disrupted by the Black Death and particularly by the departure of the Popes to Avignon from 1309 to 1378. During this period, the silver coinage declined even though issues of gold coins continued.

In the Iberian Peninsula, the mints issued many gold coins in this way reflecting the decisions of other states. At the end of the century, large silver coins were also produced by the mints in Spain and Portugal. The ability to strike such coins was facilitated by imports of metal taken back by mercenaries or taken from the Islamic armies.

The English mint had to produce huge quantities of gold and silver to finance the Hundred Years' War. In 1351 the stabilization of the gold coin (the noble, 7.78 g) and of the silver coin (1.17 g) as a groat of 4 denarii created a stable monetary system for the complete fourteenth century. England was rich: all the European countries were buying wool. But the main reason for this wealth was the huge amount of booty and ransom taken in France during the war. Tons of gold and silver crossed the Channel to be minted in London or other British mints. This production was so considerable that it was used locally.

In the Netherlands and Germany, the mints continued to produce large silver coins and after the 1340s also produced gold coins. In all these regions, a huge number of imitations of the French groat were produced.

The development of monetization also had an impact on marginal regions, with minting of gold coins in Central Europe, Eastern Europe, Russia, and Poland, for example.

All the regions that were formerly without significant coin production, such as Hungary, Serbia, and Bulgaria, now adopted coinage.

The first half of the fourteenth century was a long continuation of the prosperous thirteenth century. In the West, growth was stopped by the Black Death and the war. These events contributed to the remonetization of huge quantities of metal that moved between states and countries. All these coins were remelted several times. The development of monetization slowly affected the whole of Europe, including all

those countries that were still excluded in the thirteenth century. In the old countries, monetization became more deeply embedded in society. Good, large silver coins had been produced since the thirteenth century, but the end of the thirteenth century and the fourteenth century were periods of large issues of gold.

However, during these centuries the basic principle was to coin currency, to issue it, and then to remelt the coins. Once again, we must underline the relationship between type and number of units of account, or rather, the number of coins of each type in a theoretical pound of silver. The monetary stock was therefore composed of an enormous diversity of coins (Grierson 1976, 1991; Stahl 2000).

Famine (Fifteenth Century)

The general feature of the fifteenth century is a lack of metal. Coin production was between the hammer and the anvil. On one hand the need for coins and currency was increasing. After the Black Death and the wars, the economy began to recover and the population increased. On the other hand, the mines had been exploited for centuries, or even since Antiquity, and were exhausted. This second point was the most problematic. Let us consider some examples. The most important mint in Italy was surely Florence. In the years 1345–1351, the mint produced 213,537 gold florins and minted 2,690 kg of silver. In 1488–97, it produced only 23,938 gold florins and 157 kg of silver. This general shortage was not similar in all countries, but all the figures underline the scale of the crisis (Day 1978).

The only way to manage the crisis was to change the value of the coins regularly. There were several possibilities. All the coins had a value in units of account. Of course, this value was not inscribed on the coins, but known from laws and decrees. During the fifteenth century, as well as before and after, there was an official relationship between the type of the coin, its weight of fine silver, and its value in units of account. One possibility was then to overstrike the coin with another type to change its correspondence in units of account. The other was to demonetize the coin and of course to issue a new type with a new value in units of account.

However, the main change in the fifteenth century was the separation between silver and bronze coins. The disaggregation of the Carolingian coinage gave birth to a multitude of denarii composed of less and less silver and more and more copper. The coinage was mainly composed of these billon coins (silver + copper). The development of silver coins of good quality (the groat in all its forms) and then the development of gold coins created a monetary stock composed of gold, good silver, and of a mass of billon coins. During the fifteenth century, a clear separation began to appear between the silver coins and new issues of copper coins, even though the monetary stock was dominated by billon coins that were produced in large quantities. The former were the continuation of earlier issues and the latter were a kind of substitute for the billon coins.

The second change occurred in the last years of the century. Technological improvements made it possible for vessels to explore new regions or at least to have more regular contact between Europe and distant regions. This facilitated the

arrival of gold from the African coast. For example, the Portuguese imported gold from various mines in modern Ghana. In Continental Europe, new mines were located and at the same time, improvements in the technology of treatment of ores made the extraction of silver from the old mines more efficient. Between 1470 and 1490, new mines were found in the Tyrol and Saxony.

So in the mid-fifteenth century, the conditions for an important change in the monetary economy were met. The most visible and spectacular aspect was the issue of new large coins bearing a bust. The first of these was the ducat of Milan, in 1462. But the main change was the introduction of a new silver coin, the teston, bearing the bust of the sovereign on the obverse and struck in Milan in 1474. This coin was directly inspired by the Roman coins, bearing a bust. Copper coins appeared later: the first coins were minted in 1472 in Venice and Naples.

Thereby, some years before the arrival of Columbus in America, all the elements were in place to prepare a monetary revolution.

In Italy, the North was dominated by the powerful cities of Venice and Milan. Venetian production of gold coins declined quickly in relation to the shortage of metal. The situation was similar with the groat, the weight of which was reduced several times until the creation of the lira, a silver coin weighing 6.5 g, in 1472. In Milan, the ducat with a bust of Francesco Sforza was, in 1450, the first gold coin with a large portrait. The type of the ducat or the double-ducat was copied everywhere in the 1460s. These Milanese coins were widely distributed and the creation in 1474 of the silver teston (9.6 g), also with a large portrait, was the model for many coins, first in Italy and later in Europe. The northern Italian mints almost immediately adopted a bust on their coins. This type was adopted before the 1490s in Central Italy and then in the 1500s in the South. After this date, the coin portrait was adopted by the main European mints. The development of coins with portraits was facilitated by the weight change in the florin and the types of the coins gradually changed.

In France the situation was dominated by the end of the Hundred Years War in 1453. During the first part of the century, the billon coinage was reformed many times, and each billon issue was gradually debased: the "florette" was 3.06 g (42.5% silver) in 1417 but only 2.04 g (2.6%) in 1422. The influence of the Italian coinage was evident during the last years of the century in coins issued with a bust.

The Iberian Peninsula was composed of several kingdoms, mainly Aragon, Castile, and Portugal. Castile minted gold, perhaps in relation to Islamic rule in the South of Spain, while Aragon mainly used silver. Portugal similarly struck gold. During the century, the three currencies tended to become similar, with gold, silver, billon, and copper issues. All of them were affected by the shortage of metal after a period of instability; the coins were stabilized or replaced by a currency similar, in some points, to the Italian coinage: a bust appeared on the coins.

In the Netherlands, the local coinage was unified with the issues of Flanders, Brabant, etc. Several issues of gold and silver coins were produced; each one superseded the previous issue. Let us cite, for example, the gold horseman coin (3.63) replaced by the gold lion and then, in 1466, by the Saint Andreas florin (3.40). Each

time, the value in units of account was changed. After 1466 and up to the end of the century, the monetary system was based on this florin and a silver coin (3.08, 92%).

In England, the monetary system was based on gold, silver, and billon. As in all the other countries, the coins were changed many times. As the blanks were composed of good metal, the intrinsic value of the English currency became greater than its legal tender, due to the shortage of metal that increased its value. The weights of the coins were reduced in 1412 and 1462. English coins also circulated in Northern Europe.

A large number of coins circulated in Central Europe and Germany, some of them of very poor quality. This led to a wide campaign of countermarking coins with symbols to make it possible to distinguish between good and bad coins. The range of currencies and the growing importance of the silver mines prompted Sigismund, Archduke of the Tyrol to produce a pfunder with a large bust in 1482, as in Italy. In Saxony, thanks to its silver mines, the mint produced the guldengroschen, which were later called thalers.

The fifteenth century was therefore a period of deep changes, mainly caused by the shortage of metal. The constant revaluation of metal affected the value of the currency. The main victim of these troubles was the billon coinage, which was regularly withdrawn and reissued with new values and types.

These difficulties led to the separation of the various metals, with the tendency for currencies of pure or quasi-pure metal to be favored. The situation favored the development of gold, silver and copper coins, even though copper coins were still very rare. The second factor that prefigured the change towards modernity was the decreasing role of eastern states. The surviving Christian states used western issues or imitations, generally of Venetian coins. The issues of Byzantium were very limited and gold issues had been more or less suspended since the mid-fourteenth century; Andronikos IV Palaiologos instead issued large silver coins (Craig 1953; Grierson 1976, 1991; Miskimin 1984; Spufford 1988).

Technological Improvements (Fifteenth–Eighteenth Centuries)

Until the end of the fifteenth century, the method of production of coins was more or less similar to the technology developed by the Greeks. Any changes were made in order to increase the speed of production. The most obvious one was a decrease in the depth of carving of the dies. The Greek mints produced silver coins with a very high relief. To produce this kind of currency, it was necessary to coin the blank several times. The Romans produced smaller coins, with a low relief. This change made it possible to produce more coins with the same pair of dies. The decrease in weight during the Middle Ages and the adoption of the denarius by the Carolingians were moves in the same direction. Coins were produced with cold blanks, struck quickly as we can see from the very numerous representations of medieval mints. We must not forget that the production of coins is an industrial question, rather than an artistic one.

During the fifteenth century, the disappearance of billon coins and the emergence of a heavy silver coinage changed working conditions in the mints. The new large blanks, thicker than the thin billon coins, needed more force to be cut and prepared; their large surface required more energy to be struck. There was a need for technological adaption of the mint.

We should not separate this question of production from the other industrial improvements that were typical of the fifteenth century. The technologies required for striking coins and printing books were very similar, and the development of each of the two manufacturing processes had consequences for the other.

It seems that the first step was the introduction of a kind of drop press using a free-falling weight, contained in a frame forming a slide way. The falling weight pressed the die and coined the blank. This rudimental machine was able to produce few coins, perhaps 10 or 15 per minute.

In the sixteenth century, some improvements were introduced to facilitate the cutting of the blanks. This kind of machine was introduced in Venice in 1528, and these devices were then adopted in the mid-1550s in Germany and England.

Screw Press

The two most important innovations were the screw press and the rolling mill. The fifteenth–seventeenth centuries were marked by competition between the corporations and the workers, who defended hammer fabrication of currencies, and the partisans of the modernization of the mints.

The screw press appeared around 1530, when a certain Cellini produced coins for Pope Clement VII. A machine was in use in 1550 in Germany, where a French ambassador saw it and sent a specimen to Paris to be tested. It provoked a strong reaction from the workers who imposed a return to the traditional hammer. Similarly, a screw press was sent to London in 1561, but in 1572 the use of the screw was condemned there too. The need for a significant output of large coins obliged the mints to adopt the screw. We can see a representation of a screw press on a stained glass in Constance dated 1624. The mint of Paris adopted it in 1643 and London in 1662. By the end of the seventeenth century, nearly all the mints were producing coins with screw presses. The introduction of these screw presses corresponded to large-scale recasting of the monetary stock.

Rolling Mill

The rolling mill was developed in the first decades of the sixteenth century because the first screw presses were unable to produce large coins. A machine was in use in Nuremberg in 1530–40 and the technology spread widely in Germany, sometimes with variants: in some cases, the rolling machine used engraved rolls with several engravings, but this required the complete roll to be changed when necessary; in others, dies were placed on the rolls and only one of them had to be changed if

required. The rolling mill was in use in the Tyrol in 1572 and a group of mint-workers were sent to Segovia in 1584 to install another rolling mill, where it produced 50 real coins of Philip IV and V. The technology was adopted in London in 1625 and then in Edinburgh. The mill was an easy way to produce coins, but the coins had to be cut out of strings and in many cases the edge was not marked. The public demanded coins with well-finished edges to prevent clipping. Only the press could produce perfectly round coins with edges bearing a legend. The mints therefore moved from rolling mills to screw presses.

Other Improvements

Stamping the edge was also major progress in the manufacture of coins. The development of the technique took place in the seventeenth century and it was constantly improved to prevent clipping. The machine used to mark the edge was invented in London in the 1660s, but we have no information until 1697 when the first report mentions its existence.

At the end of the eighteenth century, the presses used steam. Between 1780 and 1790, Matthew Boulton built steam presses and produced copper coins. After 1800, all the main mints were using steam presses (Cooper 1988; Doty 1998).

Money Changers and Their Registers

We have emphasized the diversity of the monetary stock after the Carolingian period. The disintegration of the royal government paved the way for the emergence of local issues that adopted a variety of weights and alloys. The situation worsened with each century. The monetary economy needed to liquefy this monetary stock and to offer coin users the possibility of changing currency, either by giving small coins in exchange for silver or gold coins, or by exchanging billion, or even by buying old obsolete coins in order to sell the metal to the mints. Each city had several money changers who, sometimes, were also moneylenders. The money changers had an important role until the development of the banks. They dealt with coins, and some of them were also moneyers or were involved in buying metal for the mint.

The money changers used registers with drawings of the coins and some information on the legal tender of the coins. The first registers were hand-written. The development of printing in the fifteenth century corresponded to the period when issues became more and more frequent. The diversity of the monetary stock made it necessary to have books with drawings of all the coins accompanied by technical data.

There was a similar increase in the capacity of the mint to produce coins and in the capacity of printers to produce books for money changers. Their importance decreased only when states were able to homogenize their monetary stock. From that point, the circulation of all foreign or obsolete coins was forbidden and they were to be sent to the mint to be melted down. The use of money changers' books was no longer necessary.

Placards

The definition of the legal value of coins was strictly the decision of the king or of the authority in charge of the production of currency. Generally, the fineness of the coins decreased regularly or irregularly without any visible transformation of their types, except for occasional deformation.

The restoration of various authorities (king, duke, abbey, etc.) had resulted in a proliferation of new issues characterized by new types with new values. It was important for the moneyers and the state (whatever it was) to make these new types and their values widely known. The production of new coins and possible changes in their values were published on placards.

It was during the disturbances of the sixteenth and seventeenth centuries that placards and money changers' books were most abundant. The very numerous changes and the spread of printing were, of course, the two reasons for the development of this means of communication.

The numbers of placards increased during the period up to the definitive stabilization of currency (Martin 1978).

Promissory Notes

Promissory notes appeared before 1200. They were a system of compensation between merchants who had accounts with two different money changers or bankers. This system also made it possible for a person to pay by giving a promissory note to a third person; the payer and the recipient could be in different cities, far from each other.

During the following centuries, this system became more and more important, but was mainly used by Italian bankers. The development of international trade transformed the promissory note into a method of international payment, through transfers or exchanging currency or monies from one place to another.

The development of promissory notes was linked to the development of trade and the major international fairs, like those of Champagne. In many cases, they hid a credit operation that was negotiated between the two parties, as the payment could be immediate or on a set date.

Promissory notes became endorsable in the sixteenth century and circulated all over the world.

The majority of international transactions and trade used promissory notes in the fifteenth–eighteenth centuries.

To facilitate transactions, the promissory notes made reference to a unit of account the "écu de marc," which was a universal unit of account, and the payments were referenced to this unit and transformed locally into the common currency in use on the spot. In Venice, merchants used a similar unit of account, the *ducato del banco*, created in 1619; this was the *fiorini di suggello* in Florence.

The promissory note system underlines the main characteristic of the monetary system of the "ancient regime": the complete separation between the world of

currencies and the world of units of account. People dealt in coins, while merchants, bankers, dealt in units of account, some of them even being of their own creation, as in the case of the écu de marc, the *ducato del banco,* etc. This specific unit of account was used internationally by small groups of bankers and was managed only by them, while the issue of coins was the monopoly of states or public administrations. In fact, states were concerned with the production of coins and the administration of monetary stock, while the traders were only involved with units of account, even if, ultimately, they were using currencies (de Roover 1953; Boyer-Xambeu et al. 1986; Day 1998).

Counters and Tokens

In general, numismatists make a distinction between counters and tokens. The former were used to facilitate calculations, while the latter were used for various purposes. However, we will use the term token to refer to both of them.

If the right to issue coins was the mark of a king (or similar authority), and if this right was strictly regulated, private or public individuals had the right to produce or to order tokens. These tokens were circular pieces of metal, sometimes imitating coins or with any other type. The first tokens appeared in the thirteenth century. They were used to facilitate calculations.

In the sixteenth century, the use of the tokens was widespread. Some cities were famous for their production of tokens, such as Paris, Tournai, and Nuremberg. In this period, tokens were produced as medals, or only to celebrate important events.

Apart from these copper (bronze, lead, etc.) tokens, there were silver or even gold tokens which were given to the main officers or the main administrators of the state.

There were also monetary tokens. These tokens were a kind of substitute for small change. They were produced to facilitate trade. They were produced in very large numbers and cities even ordered tokens for local use. In 1797, Matthew Boulton obtained the right to issue millions of tokens to supplement the monetary stock.

In some cases, tokens were also produced as complementary money, for example, for daily salaries, and could be exchanged against real currency from time to time, for example, at the end of the harvest. Tokens were distributed on many occasions for the crops, for the construction of buildings, canals, etc. Another category of token was distributed as tickets for the theatre, for example (Depeyrot et al. 1987; Depeyrot 1995a, 1996a, b).

Metal from the Americas (Sixteenth–Eighteenth Centuries)

The New Metal

The discovery of America by Christopher Columbus represented a major change in the monetary economy. During the fifteenth century, the shortage of precious metal, gold and silver, favored the disappearance of many billon coins and the production

of new coins, in gold and silver. The silver coins were the most spectacular, bearing the bust of the prince.

Fewer small coins, a growing proportion of large silver coins in the monetary stock and a revival of the gold currency: these points were the heritage of the Middle Ages to the modern period.

The discovery of America was at the origin of a huge influx of metal in Europe. This influx and improvements in the techniques of treatment of ores created a new monetary stock. The arrival of this metal joined imports from Africa and metal from Europe. During the sixteenth, seventeenth, and eighteenth centuries, nearly 2,700 t of gold and 72,800 t of silver arrived in Europe. The equivalent in silver of the gold and silver was 16,600 t in the sixteenth century, then 32,559, then 63,212 t during the eighteenth century.

This metal did not arrive in a steady fashion. Vessels arrived with gold very early on. The Europeans were initially looking for gold. Then, silver arrived and this influx was increased by the discovery of the silver mines, the most famous of which was the Potosi, exploited after 1545. In the eighteenth century, gold arrived again from Brazil and from the mines of Latin America. The sudden arrival of this metal caused a flood of metal into Europe, which was enduring a metallic shortage. For example, the mint of Florence was producing in c. 1503 about 41 kg of gold coins and 958 kg of silver. In 1520 the mint produced 258 kg of gold and only 204 kg of silver, but in c. 1590, its production of gold was only 19 kg while its production of silver was 5584 kg.

For the monetary system, this sudden arrival made gold more common than silver, and after some decades, silver become more abundant. While the currency was more and more plentiful, the two levels of exchange were deeply disrupted; the vertical one gold suddenly becoming common and then rare, the horizontal one (in the same metal) by the sharp increase in the output of the mints receiving the new metal.

The arrival of the metal changed not only the monetary stock, but also the whole economy. The use of coins became deeply embedded in trade. As a consequence, promissory notes became more and more common.

Thalers: A European Standard

The most important change was the first issue in the Tyrol of the guldiner, 31.5 g of good silver (93.5%), in 1484 or 1486. The monetary stock in Germany was dominated by silver coins that were more or less copies of the French groat. There were also several gold coins of unstable weight and fineness. In 1500, Saxony began to produce a thaler (29.2 g, 93.8%) copied from the Tyrolian one. Then in 1518 the Bohemian states found the mines of Joachimsthal and in 1520 the first coins began to be minted. The scale of production was enormous and from 1520 to 1528, two million Bohemian coins were produced.

The election of Charles V as Holy Roman Emperor in 1519 changed the situation in Central Europe. There were a succession of decisions to homogenize monetary production, composed of the florin and the Guldiner of 29 g (and fractions), but the

decisions of Charles or Ferdinand I and Maximillian II, his successor (1524 in Esslingen; 1551 and 1566 in Augsburg), put an end to the conflicts between the various states. Each had its own standard and the increasing value of silver in relation to the first large arrivals of gold from America made the ratio very unstable. Whatever the differences were (in terms of metal contents and weights), this period consolidated the preeminent role of the large silver coin that was now produced by nearly all the Central European mints. The wide circulation of the thaler and the need for large issues facilitated the introduction of some improvements in the identification of mints, moneyers, and date. Around the mid-sixteenth century, dates appeared on the coins. The large thaler rapidly became a means of advertising the power of various cities, dynasties, and regions. It was a way for each one to promote itself and the quality of the coins improved, giving us a large series of designs. The evolution of the silver coinage affected the billon coins still in circulation, and the development of the thaler pushed billon coins towards demonetization: silver coins became the main coinage.

Spain, succeeded in defeating the Muslims in 1492 and in 1497, Ferdinand decided to reform the coinage with new gold coins (based on the *excelente* of 3.5 g). The reform of the silver coinage is less clear, but Ferdinand created a new silver coin, the royal of silver (real de plate), weighting 3.4 g. It was also struck as a double real, quadruple, and as a multiple of eight (*real de ocho,* the *peso*, 28 g) which was similar to the large silver coins minted in Central Europe. Thanks to the American mines, these coins became the most famous coins of the modern era.

The Italian fifteenth century gave us the silver Milanese lira of 9.8 g, decorated with a large bust. Slowly, nearly all the coins minted in Milan were decorated with the bust of the sovereign. The coins produced by the mints in the surrounding region adopted this bust on the obverse. Florence, the other powerful city, produced the traditional florin until the arrival of Alexander de Medici in 1532. Venice was the most important port in the Mediterranean Sea. The ducat was one of the most famous coins and was minted in millions. In southern Italy, after attempts to reform the coinage, a pure copper coin was minted in Naples in 1472. The mints produced various coins in gold or silver.

The discovery of America heralded the end of the Venetian coinage. The ducat was still the most important gold coin in the Mediterranean. However, even though Venice was reluctant to adopt the large silver coin, the mint began to produce a silver ducat (32 g). Venice continued to issue gold coins and even very heavy gold coins (up to 69 g). Silver coins were more limited due to the fact that silver was coming through Spain. However, silver coins were so common in the Mediterranean that Venice produced such currency either in the mint for its exchange or locally. The Spanish domination of Northern Italy facilitated the introduction of the thaler. Apart from the traditional teston, Charles V introduced a ducaton of 34 g into Milan, thanks to foreign metal. After Charles V, his successors continued to produce large silver coins in Milan. Charles took Florence in 1531 and imposed Alexander de Medici as governor. Whatever the family feuds in the city, the development of the monetary system was the same; new gold coins and principally a large silver coin: in this case the piaster (32 g).

All the mints of Italy developed in the same way, reorganization of the gold coinage in relation to the arrival of metal and adoption of large silver coins. This was the case with the papal states, with a reform under Paulus III (1534–49) and later, in 1588, with the creation of the silver scudo (32 g) and even later with the adoption of the small copper coin in 1600. At the same time, the same silver coin was minted in Parma. Naples also adopted large silver coins (1586, 32 g) and copper coins (1572). Here, again, the coins bore the date of issue. Savoy, at this time independent, also adopted the thaler-style coin, with a *tallero* of 32 g.

Spain and Portugal were the first beneficiaries of the new metal. The king of Portugal created a monetary system based on the arrival of gold with large coins (up to 25 g) and smaller silver coins beside copper coins. The production of coins depended on these imports and each significant variation in the influx had serious consequences for production and the value of the currency. The Spanish mints produced the gold doubloon (6.7) and the silver *real de ocho* (27.5 g) in large quantities. They were produced in this volume either to be put into circulation in territories linked to Spain or to finance the wars against the Ottomans.

In France, the king produced many gold coins as a result of the arrival of American metal. The gold écu and the silver teston were very popular coins and their production was typical of the reign of François I, who introduced France to the Renaissance and Italian artists (amongst them Leonardo da Vinci). The production of the teston continued during that century, until it was banned in 1575. In 1551 a new machine was installed to stamp the coins. France was, in this way, following global developments in the production of coins. Some large coins of 20 g were produced first in 1564, but principally after the end of the teston, a silver franc of 14.2 g. The main reform was, after a period of trouble and speculation on the billon, the appearance of a copper coin in 1575.

In the Netherlands, the coinage was a tool in the fight between Spanish domination and the wish for independence. But the influence of the thaler was evident everywhere. In the Netherlands, copper coins were produced (after 1543), and the famous silver leeuwendaalder (coin with a lion) or rijksdaaler (30 g) that circulated in central Europe and along the Danube. They were mixed with the huge and widespread production of coins from Transylvania, Poland, Scandinavia.

In England, the situation was confused, with lots of worn coins in circulation. After the wars, Sweden minted a silver coin copied from Central European ones, weighing 29.44 g. In Poland, the mints produced gold and silver coins influenced by Central European coins.

During the sixteenth century, all these countries adopted similar systems, based on the restoration of gold coinage and issues of large silver coins. From time to time, the mints produced very large coins to demonstrate their ability and technological progress. Seville produced huge gold (338 g) and silver coins (172 g); London issued silver coins (60 and 120 g) and Transylvania gold coins (350 g).

The sixteenth-century revolution had the same role and significance as the conquests and looting of territories during the Greek and Roman periods. The armies brought back tons of gold and silver that increased the monetary stock and compensated for yearly losses due to wear or loss of currencies.

The output of the mints exploded and progressively the metal moved from one mint to another, where it was recast and struck again. Each state attempted to ban the introduction of foreign coins and to avoid exportation of its own coins. Each country adopted strict legislation to control precious metal: export was seen as a way of weakening a nation.

There was, however, no regulation, no harmonization of monetary policies in this period. The main consequence of this was irregular production and major disturbances in the value of the coins. As there was no fixed relationship between currencies and units of account, the legal tender of coins fluctuated according to decrees and the decisions of states.

The period was therefore characterized by numerous issues of new coins and demonetization of earlier issues. The role of bankers and money changers grew to allow more flexibility in the stock by changing old or foreign coins. The development of printing increased the production of placards and registers for money changers.

The issue was different for international traders who had to deal with so many currencies. The problem was solved by the development of promissory notes, generally controlled by Italians.

Slowly, several levels of circulation were established in Europe, gold, which was more and more common but limited to the upper classes, large silver coins and their fractions, and billon, in course of replacement by copper coins for daily use. If the production of coins was the monopoly of states, merchants dealt with their own creation, their unit of account, and conducted their trade with a complex system of papers.

However, change was on the way. Since the 1550s the date of issue had been inscribed on some coins and this was becoming more and more common. But the most important development was the introduction of a value in multiples of reales on Spanish coins. At the beginning of the sixteenth century, the real was the unit and multiples were marked according to their value in reales. This was the case for the 4 and 8 real pieces. The reason for this was perhaps the poor quality of the striking of the 8 reales in the large American mints, where the *plata de ocho* were primitively minted in the mints of Potosi, Bogota, Mexico, Lima, etc., on blanks that were only partially struck. This way of indicating the value slowly became usual on copper coins that were imperfectly produced.

With the introduction of indications of value on coins, the method of managing the monetary stock changed. It was impossible to proceed via administrative decisions. The revaluation or devaluation of these coins required either their recasting or the addition of a countermark giving the new value (Hamilton 1934; Spooner 1956; Clain-Stefanelli and Clain-Stefanelli 1978; Morineau 1985; Depeyrot 1995a).

Towards Stabilization

The endless influx of (mainly) silver continued in the seventeenth century. The war against the Ottomans continued and the influx of metal was used to finance armies.

All the mints produced gold and silver coins. The widespread use of presses in the mints also favored faster production of coins: Vienna obtained its rolling mill in 1680. At the same time, the moneyer decided to use this technology to strike larger and larger gold and (mainly) silver coins. The Austrian mint produced the thaler and multiples of the thaler, coins of about 60 g. The status of the small coins was not clear; the coins were debased, because either their metal content or their weight was reduced.

The mints of central Europe produced thalers, the weights of which were sometimes evaluated according to decrees. However, the general features of the monetary system remained stable: gold coins based on a unit of 3.5 g and a large silver coin based on a weight of about 30 g. The thaler of Maria Theresa was famous, especially in Africa, and the type was copied for centuries for trade in East Africa. It was the first international coin, produced by many mints, including Italian, French, and Austrian ones.

The situation was the same in the German Empire. Whatever the details of the development of the monetary system, it remained the same, as in Austria.

In France, the development of the monetary system was a consequence of an improvement introduced by Nicolas Briot and Jean Warin. The new system was based on the gold louis, 6.66 g, and a large silver coin. Both of them had multiples and fractions. The currency stayed unstable and their value changed regularly. During the last decades of the seventeenth century and the first of the eighteenth century, Louis XIV changed the value of the coins by restriking them. The physical characteristics of the coins were unaffected, but the old type was overstruck by a new one that gave the coin a new value in units of account.

Louis XV (1715–1774) tried to create a bank in 1716, under the direction of John Law. The creation of this institution was at the origin of a vast period of speculation (1718–1720) that ended in bankruptcy: the bank produced 150 million pounds in 1716–18, then 2173 million in 1719–20 before it came to an end in 1720. This was the last major event in France, except for some revaluation of coins. After 1726, the value of the currency was stabilized, which meant that the number of units of account represented by each coin was stable and no longer moved.

In Italy, the power of Venice had diminished. Atlantic trade, the role of Spain and Central Europe, and the growing importance of new ports such as Amsterdam became determining factors. The gold coin in all its forms, whether ducat or not, was no longer the main European coin. The very heavy coins struck by the mint (up to 350 g) were only prestige coins. The other major cities had their own currency or were dominated by foreign states. This was the case for Milan, which was obliged to strike coins according to the standards of Austria. The Pontifical States and Naples continued to produce coins, more or less according to the general system (gold, silver copper), even though each state had its own standard. In fact, in Italy, coinage was more a means of spreading its image and promoting an idea of art than it was intended to compete with the main currency or to create a new system. The period when Venice, also, imposed its coinage on the world was over.

As far as the Iberian Peninsula was concerned, Portugal benefitted from the arrival of gold from Brazil, where the mines in the region of Minas Gerais

were particularly rich. This made it possible to mint large gold coins and to stabilize the whole coinage. Spain around the time of the Bourbons produced gold and silver in various forms as well as in the colonies. Gold coins were produced in America, with the *plata de ocho*, the famous silver coin.

Each European country developed in more or less the same way, and it is not necessary to list here all the differences between them. There details are for numismatists and not for historians. All the mints improved their issues and were able to strike more and more coins. The production of uniform coins, in gold, silver, and copper, was no longer a problem, and the introduction of the steam press during the last years of the eighteenth century ameliorated this process. Large issues were an opportunity to develop the monetization of these societies. Copper played an important role in large areas.

The seventeenth century was also the century of copper. The increasing need for silver and the monetization of trade increased the need for metallic coins. Copper was used to compensate for the shortage of small silver coins and the disappearance of billon coins. Often these coins were stuck with a nominal value on them that caused problems when inflation made the metallic value higher than the legal tender. These coins were produced in large numbers in all countries, including France, Spain, and Italy. In Scandinavia, the copper mines exported metal to European mints or copper was struck as large plates in order to replace silver. These plates were also exported to European mints.

The seventeenth and eighteenth centuries were a period of harmonization of the currency. Of course, all the coins were different, sometimes with different weights, always with different languages, legends, and reverse types, but the main feature was an increased tendency to standardization. For a native American, or a Chinese person, whose monetary systems were so different at that time, European currency was perfectly similar and a unique way to produce currency, representing a general consensus on the shape of the means of exchange.

Stabilization

During the seventeenth century, many economists began to actively protest against the numerous reforms of the coinage. Changes to the value of currencies in units of account were an obstacle to the development of trade and loans, for example. This debate also included reflection on the fiat money, against partisan of the metallic currency. In fact there was a tendency to link several problems relating to the question of the value of coins. For many authors, in all European countries, stability was the only way to create wealth. Stability meant the absence of fiat coinage, the prevention of coin clipping, and no change in the value of the coins.

Of course, all these points were not addressed in the sixteenth and seventeenth centuries. The improvement of minting systems facilitated the detection of clipping, but the race between the metallic value of coinage and legal tender was endless and the various states were unable to stop manipulating their coinage (Depeyrot 1996b; Day 1998).

Bank Notes

The eighteenth-century coinage combined the revival of the ancient trimetallic coinage with a hitherto unimaginable potential for production of coins. Modern coinage was free from the limits imposed by manual production using the hammer.

This improvement in coin production was the result of a combination of progress in printing and in the technology of metals. They both gave birth to the press.

The eighteenth century was also characterized by the association of printing and the promissory note. It gave birth to the bank note.

The bank note can be described as a new step in the anonymity of transactions (like coins, but unlike the promissory note). The first examples appeared in the seventeenth century, either during the sieges of cities or at very specific points, for example, the production of a Canadian currency using playing cards when there was a shortage of coins. Unlike promissory notes, banknotes had a fixed value determined by the state or the institution that produced them.

The use of notes was also a consequence of the wish to avoid transporting heavy coins. Such was the case in Sweden, when silver coins disappeared in the seventeenth century and were replaced by heavy copper plates: a thaler was worth a 700 g plate and some 50 thaler plates weighed more than 35 kg. At the same time, orders of the Exchequer were printed in England and after recasting coins in 1696, the Bank of England printed notes. In France, the recasting of coins took so long that the mints delivered receipts and these receipts were accepted as notes. In 1701 they became legal tender.

The development of all forms of bank notes was a characteristic of the late seventeenth and eighteenth centuries (Grierson 1975, 1976, 1991).

The Death of the Monetary System of the "Ancient Régime": A Planned Death?

The first issues in the seventh century BC were composed of only one currency, an electrum piece of metal. There was no problem of the unit of account, as the best and only way to trade was to count the coins or, in case of fractions, to weigh the coinage.

Rapidly, the coinage became composed of several coins, some gold and some silver. The diversity of metals precluded any system of counting or weighing. To use the currency, it was necessary to use units of account as a way to compare two coins of different metals and weights.

The first monetary systems were established by linking each coin in a strong relationship of value, each one being a multiple or fraction of the unit. Each type of coin (gold, silver, or bronze) was organized around a unit. The link between the various groups of coins of each metal was decreed by law or tradition. Periods of stability were periods when the wear and loss of coins and increasing demand could be met by the acquisition of new metal either from the exploitation of mines or from looting other countries. This was the case from Antiquity to the end of the early Roman Empire (first to second centuries). In this period, units of each metal were

both currencies and units of account: there was an amalgamation of the physical coin and the unit of account.

Difficulties arose when the authorities began to debase one of the coins. In this case, the problem was the debasement of the Roman *denarius*. The traditional relationship between gold and silver coins was impossible to maintain and the value in *denarii* of the coins began to fall.

The maintenance of the purity of the gold coin led to a floating system, in which silver and bronze coins were valued in units of account in relation to gold coins. The gold coin was a fraction of a pound of pure gold and all other coins were floating, up to the point when the whole system collapsed.

The Carolingian restoration was based on one coin being again both circulating coin and unit of account. But the question of the unit of account arose again in the tenth–eleventh centuries, when the content of the *denarius* was debased by the last Carolingian kings or the first feudal landlords. In general the monetary system was composed of one gold coin and silver ones played the role of fractions. But everything fluctuated according to shortages of metal that affected the value of the coins, as coins were not fiat money but trade money without any marks of value in terms of units of account.

The arrival of metal from America had two main effects. One was the end of the shortage of metal and the possibility of issuing large number of coins, but the other was a period of disturbance in the relationship between the two metals. Apart from the units of account used by the authorities, merchants and bankers developed their own system of units of account.

With the arrival of American silver, the monetary system began to be unified, with a gold unit more or less based on the ducat or similar coins, and a silver one based on the thaler. The quantity of coins became so great that manipulations of the coinage (recasting and counter-stamping) became impossible. The increase in the sums to be paid was also so significant that banknotes became necessary to avoid transporting coins.

However, banknotes were valued in units of account and the stability of the economy was then dependent on the stability of the unit of account. Gradually the authorities stopped manipulating the coins and monetary stability became the rule. Even if their types changed, the values of the coins were stable.

The next step was the inscription of the value of the coin on the coin itself. The habit began in the sixteenth century for copper coins or coins that had been quickly or badly struck, for example, multiples of the Spanish real. To link the coin and the unit of account was the best way of linking the note, the unit of account, and the coin.

This was the end of the monetary system of the Ancient Régime.

Cross-References

▶ Bullionism
▶ Money, Trade, and Payments in Preindustrial Europe
▶ The Role of Money in the Economies of Ancient Greece and Rome

References

Amandry M, Barrandon JN, Brenot C, Callu J-P, Poirier J (1982) L'affinage des métaux monnayés au Bas-Empire: les réformes valentiniennes de 364–368. NAC, Lugano, pp 279–295

Andreau J (1989) Recherches récentes sur les mines romaines. I. Propriété et mode d'exploitation. Revue numismatique, 6e série – Tome 31, année 1989, pp 86–112. https://doi.org/10.3406/numi.1989.1940

Andreau J (1990) Recherches récentes sur les mines romaines, II. Revue numismatique, 6e série – Tome 32, année 1990, pp 85–108. https://doi.org/10.3406/numi.1990.2021

Bland R (2013) What happened to gold coinage in the 3rd c. A.D.? J Roman Archaeol 26(1):263–280

Boyer-Xambeu M-T, Deleplace G, Gillard L (1986) Monnaie privée et pouvoir des princes. L'économie des relations monétaires à la Renaissance. Éd. du CNRS : Presses de la Fondation nationale des sciences politiques, Paris, 423 p

Broome M (1985) A handbook of Islamic coins. Spink, London

Bursche A (1998) Złote medaliony rzymskie w Barbaricum. Symbolika prestiżu i władzy społeczeństw barbarzyńskich u schyłku starożytności. Światowit, Supplement series A: antiquity, vol II. Instytut Archeologii Uniwersytetu Warszawskiego, Warszawa, ss 306 + 26 tabl

Clain-Stefanelli E, Clain-Stefanelli V (1978) Monnaies européennes et les monnaies coloniales entre 1450 et 1789. Fribourg : Office du livre, Paris

Cochet J (2008a) Le tombeau de Childéric Ier (481), avec une note complémentaire de G. Depeyrot. Collection Moneta, vol 76. Moneta, Wetteren

Cochet J (2008b) Le tombeau de Childéric I^{er} (481), note de G. Depeyrot. Collection Moneta, vol 76. Moneta, Wetteren, 348 p

Cooper DR (1988) The art and craft of coinmaking, a history of minting technology. Spink, London

Craig J (1953) The mint: a history of the London mint from AD 287 to 1948. Cambridge University Press, Cambridge

Crusafont M, Balaguer AM (2013) Medieval European coinage: 6., the Iberian Peninsula. Cambridge University Press, Cambridge

Day J (1978) The great bullion famine of the fifteenth century. Past Present 79:3–54

Day J (1998) Naissance et mort des monnaies de compte (XIIIe-XVIIIe siècles). Revue numismatique, 6e série – Tome 153, année 1998, pp 335–343

de Callataÿ F, Depeyrot G, Villaronga L (1993) L'argent monnayé d'Alexandre le Grand à Auguste. Travaux du cercle d'études numismatiques, Bruxelles, vol 13. 120 pages

de Roover R (1953) L'évolution de la Lettre de Change, XIVe-XVIIIe siècles. A. Colin, Paris

Depeyrot G (1987) Trésors et émissions monétaires du Languedoc et de Gascogne (XIIe et XIIIe siècles), préface de G. Démians d'Archimbaud. Association pour la promotion de l'archéologie, Toulouse, 179 pages, in – 4

Depeyrot G (1988) Crise économique, formation des prix, et politique monétaire au troisième siècle après J. – C. Hist Mes III(2):235–247

Depeyrot G (1991) Les p. Oxy. LI 3628/36: prix et papyri au Bas – Empire, méthodes et approches. Glaux, 7/3, Ermanno A. Arslan studia dicata. Milan, pp 585–590

Depeyrot G (1992) Le système monétaire de Dioclétien à la fin de l'Empire romain. Revue Belge de numismatique, tome CXXXVIII, pp 33–106

Depeyrot G (1993) les monnaies des lois barbares (lex salica, Gombette, Wisigoths, Ripuaria, Alamanorum, Bavaroise). In: Homenatge al Dr Leandre Villaronga. Acta Numismatica, 21–22–23. Barcelona, pp 315–328

Depeyrot G (1995a) Histoire de la monnaie des origines au 18e siècle, Tome I, Introduction, de l'antiquité au treizième siècle. Moneta, Wetteren, 208 pages

Depeyrot G (1995b) Les monnaies d'or de Dioclétien à Constantin (284–337). Moneta, Wetteren, 176 pages et 21 planches

Depeyrot G (1996a) Histoire de la monnaie des origines au 18e siècle, Tome II, Du quatorzième au seizième siècle. Moneta, Wetteren, 209 pages, graphiques

Depeyrot G (1996b) Histoire de la monnaie des origines au 18e siècle, Tome III, Les dix – septième et dix – huitième siècles. Moneta, Wetteren, 240 pages, graphiques
Depeyrot G (1996c) Les monnaies d'or de Constantin II à Zénon (337–491). Moneta, Wetteren, 336 pages, 29 planches, graphiques
Depeyrot G (1998a) Le numéraire mérovingien, l'âge de l'or, I, Introduction. Collection Moneta, vol 10. Moneta, Wetteren, 200 p. ISBN 90-74623-15-8
Depeyrot G (1998b) Le numéraire mérovingien, l'âge de l'or, II, Les ateliers septentrionaux. Collection Moneta, vol 11. Moneta, Wetteren, 256 p., 55 pl., 4 pl. ISBN 90-74623-16-6
Depeyrot G (1998c) Le numéraire mérovingien, l'âge de l'or, III, Les ateliers centraux. Collection Moneta, vol 13. Moneta, Wetteren, 271 p., 67 pl., 4 pl. ISBN 90-74623-19-0
Depeyrot G (1998d) Le numéraire mérovingien, l'âge de l'or, IV, Les ateliers méridionaux. Collection Moneta, vol 14. Moneta, Wetteren, 211 p., 31 pl., 3 pl ISBN 90-74623-18-2
Depeyrot G (2001) Le numéraire mérovingien, l'âge du denier. Collection Moneta, vol 22. Moneta, Wetteren, 194 p., 30 pl. de dessins, 7 pl. ISBN 90-74623-27-1
Depeyrot G (2008a) Le numéraire carolingien, corpus des monnaies, troisième édition augmentée. Collection Moneta, vol 77. Moneta, Wetteren, 496 pages, planches
Depeyrot G (2008b) Les principaux trésors monétaires romains (3^{e} s. av.-5^{e} s. ap., J.-C.). Collection Moneta, vol 79. Moneta, Wetteren, 142 p
Depeyrot G (2013) Eastern and western monetary systems: from difference to similarity. Revue numismatique OMNI, Revista numismática 7:177–197
Depeyrot G, Hollard D (1987) Pénurie d'argent-métal et crise monétaire au IIIe siècle après J.-C. Hist Mes II(1):57–85
Depeyrot G, Hackens T, Moucharte G (eds) (1987) Rythmes de la production monétaire, de l'antiquité à nos jours, actes du colloque international organisé à Paris du 10 au 12 janvier 1986 par la Monnaie de Paris, le Centre National de la Recherche Scientifique et le Séminaire de numismatique Marcel Hoc de l'Université Catholique de Louvain. Louvain-la-Neuve. Toulouse; Association pour la promotion de l'archéologie
Domergue C (1990) Les Mines de la péninsule Ibérique dans l'Antiquité romaine, vol 127. Publications de l'École française de Rome, Rome
Doty R (1998) Soho Mint and the industrialization of money. Hardcover, London
Duby G (1962) L'économie rurale et la vie des campagnes dans l'Occident médiéval (France, Angleterre, Empire, IXe- XVe siècles). Essai de synthèse et perspectives de recherches. Aubier, Paris
Dupuy C (1989) Traité des monnaies (Nicolas Oresme) et autres écrits monétaires du XIVe (Jean Buridan, Bartole de Sassoferrato). Textes réunis et introduits par Claude Dupuy. Collection l'Histoire partagée. la Manufacture, Lyon
Estiot S (2004) Bibliothèque nationale. Catalogue des monnaies de l'Empire romain (BNCMER) XII.1. D'Aurélien à Florien. Strasbourg, Paris, pp 39–48
Fossier R (1978) Polyptyques et censiers. Brepols, Turnhout
Göbl R (1971) Sasanian numismatics. Klinkhardt and Biermann, Braunschweig
Grierson P (1960) The monetary reforms of 'Abd Al-Malik: their metrological basis and their financial repercussions. J Econ Soc Hist Orient 3(3):241–264
Grierson P (1975) Numismatics. Oxford University Press, London
Grierson P (1976) Monnaies du moyen âge. Office du livre, Fribourg, Paris
Grierson P (1991) The coins of medieval Europe. Seaby, London
Grierson P, Blackburn MAS (1986) Medieval European coinage, I, the early middle ages (5th–10th centuries), with a catalogue of the coins in the Fitzwilliam museum. Cambridge University Press, Cambridge
Guey J (1966) De l'or des Daces au livre de Sture Bolin (1958), Guerre et or; or et monnaie. In: Mélanges d'archéologie, d'épigraphie et d'histoire offerts à Jérôme Carcopino. Hachette, Paris, pp 445–475
Hamilton EJ (1934) American treasure and the price revolution in Spain, 1501–1650. Harvard University Press, Cambridge, MA

Ireland R (1979) Anonymi auctoris libellus de rebus bellicis, edited, translated and presented with commentaries on the text, language and style, by R. Ireland. BARS, vol 63. British Archaeological Reports, Oxford

Künzl E (1994) Die Alamannenbeute aus dem Rhein bei Neupotz. Monographien des Römisch-Germanisches Zentralmuseums, Forshungsinstitut für Vor-und Frühgeschichte, vol 34. Mayence, pp 1–4

Martin C (1978) Essai sur la politique monétaire de Berne, 1400–1798. Bibliothèque Historique Vaudoise, Lausanne

Miles GC (1952) The coinage of the Visigoths in Spain: Leogevind to Achila II. American Numismatic Society, New York

Miskimin H (1984) Money and power in 15th century France. Yale University Press, New Haven

Morineau M (1985) Incroyables gazettes et fabuleux métaux: les retours des trésors américains d'après les gazettes hollandaises (XVIe -XVIIIe siècles). Cambridge University Press/Paris, Éditions de la Maison des Sciences de l'Homme, London

Morrisson C (1970a) Catalogue des monnaies byzantines de la Bibliothèque Nationale, I, D'Anastase à Justinien II (491–711). Bibliothèque Nationale, Paris

Morrisson C (1970b) Catalogue des monnaies byzantines de la Bibliothèque Nationale, II, De Philippicus à Alexis III (711–1204). Bibliothèque Nationale, Paris

Morrisson C, Brenot C, Barrandon JN, Callu JP, Poirier J, Halleux R (1985) L'or monnayé I. Purification et altérations de Rome à Byzance. Éditions du Centre National de la Recherche Scientifique (Cahiers Ernest Babelon 2), Paris

Mrozek S (2004) Argent, société et épigraphie romaine (1er-3e siècles) Recueil d'études complétées. Collection Moneta, vol 38. Moneta, Wetteren, 212 p. ISBN 90-77297-06-5

Odobescu A (2008) Le trésor de Pietroasa (Roumanie) (début 5^{e} siècle), note de G. Depeyrot, Collection Monetade, vol 75. Moneta, Wetteren. 530 p

Pliego R (2009) La moneda visigoda (2 vols.). Universidad de Sevilla, Sevilla

Spooner FC (1956) L'économie mondiale et les frappes monétaires en France, 1493–1680. LIbrairie Armand Colin, Paris

Spufford P (1988) Money and its use in Mediaeval Europe. Cambridge University Press, Cambridge

Stahl A (2000) Zecca, the mint of Venice in the Middle Ages. Baltimore (Md.): Johns Hopkins university press; New York: American numismatic society, London

Tomasini WJ (1964) The barbaric tremissis in Spain and Southern France Anastasius to Leovigild. American Numismatic Society, New York

Walker J (1941) A catalogue of the Muhammadan coins in the British Museum: a catalogue of the Arab-Sassanian coins (Umaiyad governors in the East, Arab-Ephthalites, Abbâsid governors in Tabaristan and Bukhârâ). British Museum, London

Walker J (1956) A catalogue of the Muhammadan coins in the British Museum, II: a catalogue of the Arab-Byzantine and post-reform Umaiyad coins. British Museum, London

Wroth W (1909) Imperial byzantine coins in the British Museum. Printed by order of the Trustees, sold at the British Museum, London

Wroth W (1911) Western & Provincial byzantine coins of the Vandals, Ostrogoths and Lombards and of the Empire of Thessalonica, Nicaea and Trebizond in the British Museum. Printed by order of the Trustees, sold at the British Museum, London

Yannopoulos P (1978) L'Hexagramme, un monnayage byzantin en argent du VIIe siècle. Louvain-la-Neuve, [Belgique]: Institut supérieur d'archéologie et d'histoire de l'art, Séminaire de numismatique Marcel Hoc

Money, Law, and Institutions

6

David Fox

Contents

Abstract

A version of Georg Knapp's state theory of money has represented the mainstream view of money applied in the civil law and common law traditions of Western Europe since medieval times. Following the understanding of Roman law, money was identified with the payment media issued by the sovereign body in the state. Legal doctrine recognized that the right to strike coin and to give it a value in payments belonged distinctively to the sovereign. The sovereign was entitled to change the monetary standard by altering the metallic content of the coinage or by raising or lowering its valuation in monetary units. Private law doctrines on the tender of money translated the monetary valuations made by the sovereign into practical results when the courts enforced actions for the payment of debts.

D. Fox (✉)
University of Edinburgh, Edinburgh, Scotland, UK
e-mail: D.Fox@ed.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_52

Keywords
Monetary law · State theory of money · Coinage · *Ius cudendae monetae* · Legal tender · Debasement

Introduction

Georg Knapp's *Staatliche Theorie des Geldes* first published in 1905 identified the law and the state as the key institutions in the creation of money. Units of monetary value and the payment media representing them were "created by the state in its capacity as guardian and maintainer of the law" (Knapp 1905, 1924, 39). Their status as valid payment media and their value in payments depended on the fact of legal proclamation rather than their intrinsic substance. Knapp argued that it was impossible to tell from the physical form of payment media whether they were genuinely money. Since their status depended on "a certain relation to the laws," it was always necessary to "refer to the Acts and statutes, which alone can give the information" (Knapp 1905, 1924, 34).

Knapp's understanding of the role of law in the creation of money has not gone uncriticized, particularly in its application to developments of the twentieth and twenty-first centuries (Proctor 2012). Even when Knapp's state theory was first proposed, it stood in opposition to Karl Menger's societal explanation for the origin of money (Menger 1892). Menger explained the origin of money as a spontaneous social fact. Participants in economic exchange settled on one commodity as their preferred medium of exchange because it was more highly saleable than other commodities in the market. They would willingly accept it because they were confident they could exchange it for other commodities that were directly useful to their needs.

Whatever its more general merits, the state theory represents the mainstream view of money accepted in the civil law and common law traditions of Western Europe. (The civil law systems are those of continental Europe and Scotland which are descended from classical Roman law. The common law system is identified with the rules and processes developed by the courts of England.) On this view, the phenomenon of money cannot be explained solely by social recognition and use. The view is traceable to classical Roman law. The Roman emperors minted coins and legislated for criminal offenses to protect the exclusivity of their right. Lawyers since then have rarely attempted any comprehensive definition of money. But the institutions of the law – legislatures, courts, practitioners, and learned commentators – have generally identified money with the payment media issued by a sovereign body, acting to implement its exclusive powers over the monetary system. Money emerges from networks of reciprocal obligations owed between a sovereign body and the public at large. It embodies a promise of value redeemable against legal debts, including debts owed to the state (Desan 2014). Its value in units is given to it by legal enactments issued under the authority of a sovereign body. The capacity of money to discharge debts is recognized by the private law of the jurisdictions where it is issued. The

courts that determine disputes over the performance of monetary debts recognize and enforce its value.

When legal systems recognized innovations in constitutional and monetary practice, they extended their theory of money to accommodate them. In Great Britain, for example, the new constitutional settlement established in the later 17th century brought about a change in institutional practice. Money was issued under the authority of Acts of Parliament. It was no longer created through the untrammelled exercise of prerogative powers resting in the King or Queen. Monetary sovereignty came to be identified with representative institutions which, in the new constitutional order, drew their authority from the King or Queen without being wholly subject to his or her will. When banknotes began to circulate in the 17th and 18th centuries, Parliamentary legislation gave certain of them a privileged status. It put them on the same legal footing as coins.

In explaining the relationship between money, law, and institutions, this chapter concentrates on the legal history of the constituent jurisdictions of Great Britain. The English and Scots law of money proves to be fundamentally similar to that of other Western legal systems. They all exemplify – in varying degrees of directness and at different stages – a common body of principles that was first clearly articulated in the monetary law of classical and postclassical Rome. The monetary law of continental Europe during the medieval and early modern period was directly fashioned from original Roman sources (Rüfner 2016). By a continuous tradition of exegesis, European jurists worked the Roman sources into a practical system of rules that was suitable for the commercial and monetary conditions of their time (Ernst 2016). The monetary rules of English common law do not draw directly on the Roman legal tradition. But the fundamental similarities in monetary practice between England and Europe meant that the rules developed in each system are more alike than their separate doctrinal origins might first imply.

Sovereign Power over Money

The sovereign power over the monetary system was originally identified with the right to strike coin (*ius cudendae monetae*). The rule consistently recognized across Western legal systems was that the minting of coin was an exclusive prerogative of the sovereign. The rule is traceable to a rescript of the Roman Emperor Constantine dated 326 and was reproduced in the Codex of Justinian:

> If anyone mints coins through counterfeit casting (*falsa fusione*), We instruct that all his property shall be bound over to Our Treasury. For We wish that the practice of striking coins be conducted only in Our mints . . . Whoever can be discovered to be a forger of coins or is exposed as such by anyone, shall be made over, immediately, with all delay removed, to the searing flames (to be burned alive). (Justinian, Frier (ed) 2016, C.9.24.2)

Constantine's rescript criminalized the minting of false coin. The offender's crime was treated as treason (*maiestatis crimen*) since it touched upon a right and interest

of the Emperor himself. The rescript was specifically directed at the coining of money in private mints. It aimed to preserve the exclusivity of the Emperor's own mints and to suppress all competing minting activity.

In the form that it took in the Codex, the rule doubtlessly reflected a much older constitutional practice. In systems where lawmaking power was identified with the sovereign's will, what the sovereign did was an expression of his or her right in law. Coins had long been struck bearing the Roman emperor's image. The explanation in classical Roman law of the distinction between barter and sale rested on the same principle. The jurist Paulus, writing about 200 AD, identified money as an exchange object that had been struck in a prescribed form by the mint and ascribed a stable value by the state (Rüfner 2016). He wrote:

> But since it did not always and easily happen that when you had something which I wanted, I, for my part, had something that you were willing to accept, a material was selected which, being given a stable value by the state, avoided the problems of barter, by providing a constant medium of exchange. That material, struck in due form by the mint, demonstrates its utility and title not by its substance as such but by its quantity, so that no longer are things exchanged both called wares but one of them is termed the price. (Justinian, Watson (ed) 1985, D.18.1.1.pr)

We see in this definition what Georg Knapp later called a "morphic proclamatory" theory of money (Knapp 1905, 1924, 31). Roman money derived its status from the fact that it was issued by the Emperor in a form authorized by him.

With the medieval revival of Roman law in Western Europe, rules modeled on Constantine's rescript were promulgated in new codes and expounded upon in the commentaries of continental and English jurists. A version appears in the *Liber Augustalis* (1231) promulgated by Frederick, King of Sicily and Holy Roman Emperor:

> We inflict capital punishment on and confiscate the property of those who coin adulterine money or who knowingly receive it and use it. We also inflict this penalty on those who conspire with them. (Powell 1971, III, XLII, 40)

The English commentator, Henry de Bracton (c 1210–1268), explained a variety of coinage offenses punishable as lese-majesty (Bracton and Thorne 1968). He associated the coining of false money with the counterfeiting of the King's seal. The association emphasized the distinctive wrong of usurping the King's right to confer the status of money on coined metal. The power of authenticating money lay with the King alone. The falseness of money was analyzed as a legal rather than a material characteristic. The counterfeiter's crime would have been the same even if the coins had been made to the same weight or fineness as the King's own money.

Legal commentators came to identify the minting of coin with the regalia of the sovereign. In Scotland, Thomas Craig's *Jus Feudale* (c 1600) explained the King's *ius cudendae monetae* as belonging to a standard list of regalia derived from the *Liber Augustalis* of Emperor Frederick II (Craig, Dodds (ed) 2017). It belonged alongside other exclusive rights associated with the sovereign such as the levying of

taxes, the maintenance of arsenals, and the levying of fines. In the 17th and 18th centuries, the English commentators, Sir William Hale and Sir William Blackstone, used more conventional common law terminology describing the right to issue coin as the sovereign's "prerogative" (Hale 1736, 1847; Blackstone 1765). Whichever term was used, the association of *ius cudendae monetae* exclusively with the King established that the right was an inalienable incident of his sovereign power. The King held the right in the public interest. It could never be granted away. This is the significance of the dictum reported in the decision of the English Privy Council in *Gilbert v Brett* (1604): *monetandi ius principum ossibus inhaeret* ("the right of issuing money belongs in the very bones of princes"). The right to mint coin was no more separable from the King himself than were his bones from his body. Any other person minting coins was either exercising a limited permission from the King or was a treasonous usurper of the King's right. The burden was on him to prove his authority from the King, on pain of criminal penalties if he failed.

Incidents of the Sovereign Right

Legal commentary and practice recognized that *ius cudendae monetae* included other powers beyond the basic right to strike coins. The most important of these was the sovereign's right to ascribe a legal value to coins. In law, coins were more than conveniently transferable quantities of precious metal, the weight and fineness of which had been authenticated by the sovereign. They also carried the distinctive capacity to discharge debts. The parties who used coins in payments needed to know how many coins debtors had to tender to secure a valid discharge from their debts. Coins needed a value in the same monetary units as were used to denominate the debts.

The civil law analysis of the high middle ages distinguished between *moneta in obligatione* and *moneta in solutione* in payments (Ernst 2016). These terms referred, respectively, to the monetary units measuring the value of the debtor's obligation and the kinds of coins that the debtor was required to tender to get a good discharge from the debt. Like all other material commodities, coins could be given a valuation in terms of monetary units. The difference was, however, that the valuation of coins was fixed by law through the exercise of the sovereign's prerogative. The courts that enforced debt transactions were bound to take notice of these valuations. They enforced them to the exclusion of any competing market values. The value of the sovereign's money in his own jurisdiction was not a question of fact which the court was free to decide by evidence.

The Scots and English legal sources of the 17th century were explicit in identifying the right to fix the value of coin with the sovereign's prerogative. Thomas Craig's *Jus Feudale* (c 1600) described the formal valuation of money as that "official mark" with which money was "stamped" in accordance with the authority of the prince (Craig, Dodd (ed), c 1600, 2017). Craig's reference to a "mark" on the coins was a legal metaphor: the coins themselves were not stamped with numbers to show their monetary valuation. But the impressions on the coins identified them with

the physical descriptions in the legal documents by which the sovereign authorized the mint to strike a new issue of coins (these were called "warrants" in Scotland and "indentures" in England). The documents usually contained elaborate descriptions of the iconography and lettering to be stamped on each type of coin and ascribed a valuation to them.

As an adjunct to the sovereign's power to fix the legal valuation of his own coins, it also lay within his prerogative to adopt foreign coins into the local monetary system. In the trading centers of continental Europe and of Scotland, a large proportion of the coins in circulation was made up of foreign coins. They circulated at informal market rates agreed by the traders who used them. The market rates sometimes overvalued the coins relative to their intrinsic bullion content, which misled the public as to their intrinsic worth and prejudiced the circulation of the sovereign's own money. The sovereign's response was to ban the foreign coins entirely (England's situation as an island made this a feasible legal response in that country), or more usually to ascribe them a valuation in the local money of account, which would allow them to circulate on a par with the sovereign's own money. The fixing of the valuation was usually effected by a proclamation from the sovereign. The adopted foreign coins were then indistinguishable in payments from the local coins minted under the sovereign's *ius cudendae monetae*.

More controversially, the sovereign's right to fix the valuation of the coins extended to changing the monetary standard. He was entitled to alter the valuation of coins already in circulation or to make a new issue of coins with a different weight or fineness. Thomas Craig accepted as much, although he regretted the way the Scottish Kings abused their right. In his own lifetime, he had experienced a fivefold debasement of the Scottish coinage: "Little by little, by skulduggery and dishonesty, we are being cheated" (Craig, Dodd (ed) 2017, 479).

The English common law understanding of the sovereign's right to change the monetary standard was the same as Craig's. The leading case was *Gilbert v Brett* (1604), which was a dispute heard by the judges of the English Privy Council after Elizabeth I's debasement of the Irish silver coinage in 1601 (Fox 2011). Gilbert sold goods to Brett for a price of £100. Between the date of sale and the due date for payment, a proclamation by the Queen withdrew the old silver coinage and replaced it with a new coinage struck from debased silver. The proclamation decreed that the new currency had to be tendered and accepted at its nominal legal value. The proclamation made no allowance for the depreciation of the currency in terms of its silver content. Brett tendered £100 in newly minted coins when the debt fell due for payment. Gilbert rejected the tender and held out for payment in old coins or at least for enough new coins to give him an equivalent amount of silver.

The Privy Council held that Brett had made a good tender with the new debased coins and that Gilbert was in contempt of the Queen's prerogative by rejecting them. It strongly affirmed all aspects of the sovereign's prerogative over coinage. It lay in the sovereign's prerogative to make coins of whatever metallic composition she pleased and to fix their value in the money of account. It necessarily followed that she could change the monetary standard by reducing the bullion

content of the coinage (which was a debasement in the strict sense of the word) or by ascribing it a higher or lower valuation in money of account. Ultimately, she could demonetize the coins so that they reverted to bullion. This was what the Queen had done to the former Irish currency. Gilbert's insistence that the debt be paid with the pre-debasement coins was an impossible legal demand. Those coins were no longer money in legal estimation. The debtor would not have been making a valid tender that complied with the terms of his obligation to pay the price in money.

Significantly for later developments, the Privy Council held also that the sovereign's prerogative to change the monetary standard did not require the consent of Parliament. The many past occasions when the King had changed the standard of the English coinage without Parliamentary consent were taken as evidence of his unconstrained prerogative. This view would change after the constitutional struggles of the English civil war and the restoration of Charles II to the throne of Great Britain. Sovereign control over the monetary system would eventually be recognized as resting in Parliament (see section "Later Developments in the Sovereign Right over Money").

The legal process for ascribing a valuation to coin locates *ius cudendae monetae* in the larger scheme of economic regulation falling within the sovereign's powers. The striking and valuation of coin were not ends in themselves. They served more diffuse economic purposes, from facilitating domestic payment transactions to controlling the balance of international trade. Blackstone (1765) saw the connection more clearly than other writers such as Craig, who located *ius cudendae monetae* in a conventional, but eclectic, list of prerogative rights (Craig, Dodd (ed), 2017). Blackstone linked it instead with the sovereign's general responsibility as the arbiter of domestic commerce. He associated it with other sovereign powers of the King that included the standardization of weights and measures and the grant of charters to hold public markets.

The full detail of the sovereign's monetary powers is apparent from the legal instruments that implemented them rather than from the standard legal sources, such as legislation, judicial decision, or learned commentary, that authorized or explained them. The records of Acts of the English and Scottish Privy Councils throughout the 17th century are repleted with examples of the monetary prerogative in action. The sovereign would intervene by proclaiming measures intended to stimulate exports, encourage positive bullion flows into the country, and protect the local currency against damaging competition in the foreign exchange markets. It is common to find Privy Council proclamations that revalued the local currency to reflect changes in bullion prices abroad, proclamations requiring merchants to bring foreign coin to the mint so it could be reissued in the local currency, proclamations prohibiting the unauthorized melting of coin for bullion, and proclamations criminalizing the acceptance of foreign coins that had been banned from circulation (Larkin and Hughes 1973; Cochrane-Patrick 1876). The fixing of coin values was integral to this larger network of macroeconomic policies. If the local coin was undervalued against foreign currencies, then it would tend to be exported, causing a damaging drain of specie and a scarcity of coin at home.

Later Developments in the Sovereign Right over Money

In Great Britain, the later 17th and the 18th centuries saw important legal changes in the sovereign's prerogative control over the monetary system. In one sense the prerogative was curtailed. Monetary sovereignty shifted from the King and his advisers in the Privy Council to the King operating through Parliamentary legislation. In another sense, sovereign control over the monetary system expanded in response to innovations in monetary practice. Banknotes, which at first fell outside the exclusive prerogative to issue coin, were eventually encompassed within Parliament's sovereign control over the monetary system. The developments illustrate the open-ended nature of monetary sovereignty: "historical and political developments established certain facts in the first place and the related concept [of sovereignty] was developed only at a later stage, in an attempt to analyse reality as part of a legally coherent framework" (Zimmermann 2013, 9). While the sovereign prerogative over money was originally identified with the right to strike coins, that usage did not constrain the exercise of sovereign control once new kinds of money came into circulation.

The restoration of King Charles II to the English and Scottish thrones in 1660 marked a turning point in the legal history of the monetary system of Great Britain. The King's sole prerogative to control the monetary system ceased to be the constitutional norm. By the statute 18 & 19 Car II, c 5 (1666) ("An Act for encouraging of Coinage"), the English Parliament enacted that the cost of assaying, melting, and striking coins was to be met by the government out of tax income. The individuals who brought bullion to the mint no longer bore the cost of coinage, as they had under the old free minting system (Desan 2014). That system had also allowed the King to deduct a seignorage tax from each new batch of coin sold by the mint, which created an incentive to debase the coinage. The task of setting an appropriate mint price for bullion was a delicate one. All too often it was not fixed at a rate that successfully encouraged a steady flow of bullion to the mint. To remedy this problem, the statute spared the individuals who brought bullion for coining the direct costs of the minting process.

The principle that the costs of minting should be borne by taxation became the new norm in Great Britain. The costs of William III's recoinage of the English silver currency during 1696–99 were met from taxation (Li 1963). When the Scottish silver currency was recoined between 1707 and 1709 after the 1706 Articles of Union between England and Scotland, the cost was met from the "Equivalent fund" paid by the English government (Murray 1997).

The other significance of the statute of 1666 was that it was an Act of Parliament. Previously, English or Scottish legislation concerning coinage had originated in the King's Privy Council. It was publicized and implemented by official proclamations, which in England at least were direct sources of prerogative law. The change in practice became more marked after the Revolution of 1688 which deposed King James II/VII and installed King William and Queen Mary. From then onward, English legislation on coinage, and on the monetary system generally, emanated from Parliament. The most obvious instance was the legislation enacted

to govern the recoinage of England's silver currency in 1696–99. A series of Acts of Parliament provided for the demonetization of the old clipped coinage and its replacement with new coins struck to restored standards of weight and fineness (Li 1963). The change in monetary practice was consistent with the shift in constitutional norms represented by the Bill of Rights 1688 enacted by the English Parliament. The Bill made no provision for the coining of money. But the new constitutional order that prohibited the levying of taxes and the suspension of laws without Parliamentary consent was unlikely to countenance the King's untrammelled exercise of prerogative power over the monetary system.

Acts of Parliament concerning money frequently appeared on the statute book throughout the 18th century. They set the main structures of the UK monetary system on a legal foundation defined by Parliament. In 1774 the statute 14 George III, c 42 defined an upper legal tender value on silver coin of the realm:

> That no Tender in the Payment of Money made in the Silver Coin of this Realm of any Sum exceeding the Sum of Twenty five Pounds . . . shall be reputed in Law, or allowed to be a legal Tender . . . for more than according to its Value by Weight . . . and no Person to whom such Tender shall be made shall be any ways bound thereby; any Law, Statute or Usage to the contrary notwithstanding.

The effect of the statute was to enact that the United Kingdom operated on a gold standard. It had previously been on a fully bimetallic standard where gold and silver coins circulated on the same terms, with no upper limits on their legal tender status. The statute made silver a subsidiary coinage. Its legal status as money continued only so long as it was tendered in payments not exceeding £25 in value. Above that limit it reverted to bullion. The provision also demonstrates the interrelationship between the private law rules governing the discharge of debts and the definition of money in the public law of the realm. Coin carried its status as money by its capacity to compel the discharge of debts. On this view, gold coin of the realm was always money in legal estimation. A debtor could tender it to discharge a debt for any amount.

Acts of Parliament also established a new place for banknotes in the monetary system. Since the 17th century, private banks had been free to issue their own banknotes because the notes fell outside the sovereign's *ius cudendae monetae*. Seen as circulating credit instruments, banknotes were different in kind from the coins over which the sovereign had traditionally asserted an exclusive prerogative. Parliament eventually came to recognize the special status of notes issued by the Bank of England. At the onset of the Napoleonic wars, Parliament enacted the statute 37 George III c 45 ("Bank Restriction Act 1797"). It suspended the Bank's duty to pay coin on notes presented to it for payment. The statute changed the legal status of the notes. They circulated on a footing almost equivalent to that of coin. Convertibility was restored in 1821 after the end of the wars. The status of notes was again enhanced by legislation in 1833. The statute 3 & 4 William IV c 98 ("The Bank of England Act 1833") made Bank of England notes legal tender for all sums above £5. At least in private payments (although not for the purposes of redemption against the Bank), the notes were legal tender equivalent to gold coin.

This development in practice eventually earned legal recognition in the courts' rationalization of monetary sovereignty. In *Emperor of Austria v Day* (1861), the English Court of Appeal in Chancery granted an injunction to restrain the printing of banknotes by a political opponent of the Austrian Emperor and King of Hungary. The notes were intended for circulation in Hungary. The court confirmed the Emperor's authority to control the circulation of money in his own territories and that the right to issue notes followed from the *ius cudendae monetae* belonging to the supreme power in every state. It was not confined to the coining of precious metals but also included the issue of base metals or paper instruments made to represent varying amounts in value of gold and silver. That sovereign right could be enforced as exclusively for new forms of circulating money as it could for the old forms consisting in coins struck from precious metals (Proctor 2012).

The Nominal Value of Money

The sovereign's right to ascribe a money of account valuation to coins was essential to the legal analysis of monetary value and the proper discharge of monetary obligations. Until circulating paper money came into common use in the late 17th and early 18th centuries, the circulating media of most European states consisted in coins struck from precious metals. Money had a strong physical foundation. It was as much a material substance as other kinds of valuable treasure, such as plate or jewels, that were also used to store personal wealth. What distinguished coin from other kinds of treasure was the homogeneity of its physical form and the fixing of its value by a legal act of the sovereign (Blackstone 1765). Coins simultaneously carried two values: an intrinsic value based on their bullion content and an extrinsic value based on the valuation ascribed to them by the sovereign. A perennial legal question confronting jurists and courts throughout the medieval and early modern periods was how coins were to be valued for the purposes of discharging debts.

The long-standing assumption of the English common law was that monetary debts had to be discharged according to the extrinsic value of the coins tendered by the debtor on the due date for payment. The common practice among lawyers of the late medieval and early modern periods was to denominate debts as generic money of account sums. Debts were expressed, for example, in the form "xx li of lawful English money" rather than as requiring the delivery of any particular combination of coins equal to that value. The legal forms of action used for enforcing debts in court were also expressed in money of account sums (Desan 2014). The effect was to preclude any inquiry into a change in the intrinsic value of the coinage between the date when the debt was incurred and the date when the debtor paid it. This, as we shall see, was the result enforced in *Gilbert v Brett* (1604), which gave special weight to the sovereign's untrammelled prerogative to change the monetary standard. The case confirmed that English law took a nominal approach to the valuation of money and monetary obligations. The creditor bore the risk of debasement.

The writings of Viscount Stair in his *Institutions of the Law of Scotland* (1681) confirmed that the same was true for the law of Scotland (Stair, Walker (ed) 1681,

1981). In this, Stair disagreed with Thomas Craig who had argued that the legal value of money should equate to its intrinsic value as bullion and that debts should be repaid at their original intrinsic value (Craig, Dodd (ed) 2017). Stair's reasons drew upon the civil law tradition of monetary thought. They referred to the explanation of the defining characteristics of money given by Roman jurist Paulus (see section "Sovereign Power over Money"). Stair read Paulus as saying that money was a fungible token of exchange, the physical substance of which was immaterial to the discharge of the debt. Its extrinsic value imposed by the King was therefore to be respected. It did not matter whether the intrinsic value of the coins tendered by the debtor was less than the value of the debt, assessed in intrinsic terms, when the debt was first contracted.

On the whole, the civil law jurists of continental Europe were slower to reach this same nominalist conclusion than the English common law. The continental jurists' views were shaped by their analysis of the contract of loan for use in classical Roman law (*mutuum*) (Ernst 2016). *Mutuum* was contracted when a lender transferred fungible property to a borrower for the borrower's own use. The typical example was a loan of money. The ownership of the property passed to the borrower, subject to the borrower's duty to restore property of the same essential kind and quality to the lender. The question of how the debtor should repay the loan forced jurists to consider the essential quality of the coins lent by the creditor. From late in the 12th century, the orthodox view had been that the debtor was required to repay coins having the same intrinsic value as the coins originally advanced to him. The rule entailed that the risk of a debasement in the currency lay with the debtor. It rested on an assumption about the essential quality of coin. Coin was analyzed as a special kind of bullion, the weight and fineness of which was certified by the sovereign. It was as if money debts were obligations for the transfer of bullion in the form of coin.

This remained the predominant view until the late 17th and the 18th centuries. A pivotal figure in the change of view was the French jurist Charles du Moulin (1500–1566). His *Tractatus contractuum et usurarum* (1584) made a sustained argument that the extrinsic value of money was primarily relevant to its value in the discharge of debts (Dondorp 2016). Du Moulin argued that the value of money consisted in a social convention. It was established and enforced by the legal decree of the sovereign. The form and substance of money therefore lay in the sovereign's legal act of monetizing coins and decreeing the legal value at which they had to pass. It did not consist in the material substance from which the coins were struck. Debts were discharged when the money of account value of the debtor's obligation equated to the money of account value of the coins tendered by the debtor.

Du Moulin's view eventually became the norm among writers in the civil law tradition. The Dutch jurist, Johannes Voet (1647–1713), argued that when money was given on loan, it was not so much the coins that were to be considered but the legally decreed monetary amounts that they represented. He wrote:

> It follows that if the public value of coins is increased [the debtor] can obtain release by repaying fewer than he had received on loan; but if it has been lessened, he is held liable to pay back proportionately more. And if he perhaps has received coins made of purer metal, he

> can then pay back others which . . . have been made of a cheaper substance, so long as those coins, though cheaper in respect of their intrinsic goodness, have not been deprived of their currency by public authority. (Voet, Gane (ed) 1955–1958, XII.I.24(iv))

The French understanding of the 18th century was the same. The French jurist Robert Joseph Pothier (1699–1772) noted that the value of money was fixed by the sovereign's prerogative. Debts therefore had to be paid according to the monetary standard at the date of repayment rather than at the date when the debt was incurred (Pothier 1773, Part I, ch II § III).

The legal recognition of nominalism was an important step in monetary development. It assisted the development of new forms of money that derived their value from legal fiat rather than from their intrinsic content. It facilitated the circulation of paper money and token currencies from the 17th century onward and enabled their eventual recognition as legal tender.

The Homogeneity of Money as a Construct of Law

So long as money consisted of intrinsically valuable coins, the ideal of perfect homogeneity among all coins presented a technical and legal problem. To take an example from English monetary practice, all penny coins valued at 1 *d* needed to be equally acceptable in the discharge of debts expressed in penny units. Six penny coins had to be as acceptable to discharge a debt for 6 *d* as one sixpence coin which had been ascribed a money of account value of 6 *d*. As a matter of law, the value of a coin in payment was fixed in the warrant or indenture by which the sovereign authorized it to be issued. But that value depended at least partly on its intrinsic value. The value of a coin represented a fractional part of the official mint price paid by the sovereign for bullion received at the mint (Redish 2000).

The problem was that coins were not all physically homogeneous even though the legal rules of monetary valuation presumed that they had to be. The process of refining metals was subject to inevitable technological limitations (Challis 1978). Even coins of the same denomination could not all be minted to a consistent standard of fineness. The process of cutting coin blanks from sheets of assayed metal meant that some coins were unavoidably heavier than others. The problem grew worse once the coins were put into circulation. Coins lost weight by natural abrasion as they passed in circulation. They came to weigh less than they did when they were first issued. When enough heavy or light coins were sorted and gathered, the accumulated differences in their weights created possibilities for arbitrage between their extrinsic and their intrinsic values. Criminals added to the problem by coin clipping. They would pare silver from the circumference of coins before putting them back into circulation at their extrinsic value. The accumulated quantities of silver removed from the coins could then be sold as bullion.

At all stages in the life of a coin, legal regulation applied to protect the ideal of homogeneity against these physical differences. The mint indentures authorizing each new issue of coin specified so-called remedies of the assay and the shear. So

long as the fineness and weight of coins stayed within bands of permitted variability, then the coins would carry the legal status of money and pass at their legally decreed values. Coins that exceeded the remedies were not money. They remained bullion despite the imperfect technical transformation that had been wrought upon them. Stringent criminal penalties were imposed for clipping coins and for culling and selling them at rates exceeding their legal value. All offenses were treated as violations of the sovereign's *ius cudendae monetae*.

Money, Property, and Liquidity

Jurists and legal practitioners of the medieval and early modern periods rarely attempted any comprehensive legal definition of the range of things that served as money. Consistently with the analysis of the Roman jurist Paulus, it was generally sufficient to identify money with the coins issued by the sovereign, and then to explain their distinguishing feature, which was that they were the price given in exchange for an object of sale. The legal distinction between barter transactions and sale transactions was easily drawn so long as money was identified with the sovereign's coins. The distinction was less clear once legal claims to the payment of coins came to be treated as functionally equivalent to the coins themselves.

Since medieval times, merchants had commonly used written payment orders to remit money of account sums between themselves in order to avoid the hazards and inconvenience of transferring metallic coin across long distances (De Roover 1948; Geva 2011). The payment networks operating between them depended on the creation, setoff, and cancellation of debt relations rather than on the transport of specie between the transacting parties. For the most part, these transactions fell outside the conventional juristic analyses found in the civil law tradition. Indeed, when jurists did attempt to analyze them, they often found that the payment transactions made a bad fit with the received categories of legal analysis.

One legal problem, however, that jurists and courts were forced to confront in the 18th century was the way the conventional rules of property law applied to banknotes. Privately issued banknotes had been in circulation in England since the latter half of the 17th century. With the foundation of the Bank of England in 1694 and the Bank of Scotland in 1695, notes came to be issued by corporations operating under statutory authority. In legal form, banknotes were transferable instruments entitling the bearer to enforce a debt for the payment of coin. Their value in payments between private parties depended on the certainty that the bearer could present the note to the issuing bank for payment in coin. They needed the same liquidity as coins, both in being easily transferable and in giving the holder the same security of title as was enjoyed by a person in possession of coins.

Until the development of banknotes, English and Scots law had relied upon the physical homogeneity of coins to explain their liquidity. The general rule governing the transfer of ownership in property is that the transferee can have no better title than the transferor had. If property is stolen, then it still belongs to the victim of the theft. All transferees who receive the property through the thief hold it

subject to the victim's surviving ownership. But English and Scots law recognized an exception for stolen money. In this they followed the analysis of the classical Roman law. In Scotland, the reliance on the Roman analysis was explicit (Stair, Walker (ed) 1681, 1981).

The Roman rule facilitated the liquidity of money. It secured the title of the current holder against claims brought by a former holder from whom the money might have been stolen. Judging by the surviving legal texts, it seems that the specific recovery of money or its value by a vindicatory action was relatively common in Roman legal practice (Thomas and Boraine 1994). The common feature of all of these instances was that the action lay to recover money contained in a bag or purse that was traceable to the former owner. Money was commonly remitted in sealed bags to money changers for assaying or to be held on safe deposit.

But a vindicatory action for recovery did not lie if the former owner's money was no longer contained in a bag or if it had become inextricably mixed with other coins belonging to the defendant. So if a thief stole A's money and paid it to B who then mixed it, A's only action was to sue the thief for the debt arising out of the theft. A's ownership was entirely extinguished by the mixture. B's ownership of the money was complete and unchallengeable even though the money derived from the tainted transaction between A and the thief. The rule even seemed to bar A from claiming a right of co-ownership of the mixture proportionate to his contribution to it.

Until the 18th century, the English and Scots legal sources took the same approach. In the English case law sources, it was said that money was "not to be known" once it passed indistinguishably into a mixture with other money (e.g., *Banks v Whetston* (1596)). The mixture extinguished the title of any former owner and made the title of the current holder indefeasible in law. An action in detinue would not lie for recovery of the money or its value unless it stayed sealed in a bag. Detinue would only lie for property that remained identifiable among the defendant's belongings. Otherwise, defective titles to money were cleansed by the fact of mixture. Once the money was mixed, the title of any subsequent transferee became more secure. The transferee could accept the money confident in the knowledge that his title to it was practically secure against challenge by any former holder of it.

The rule followed from the physical similarity of all pieces of coined money. They were minted so as to be physically and legally homogeneous in the payment of debts. They were also treated as homogeneous when they were seen as property in the hands of a third person. The person in possession of the money was presumed to be the owner, and the burden of proving otherwise fell on the person who tried to challenge that possession. Since all coins were presumptively identical, the challenger would invariably fail to discharge the burden of proof. Stair's *Institutions of the Law of Scotland* (1681) was explicit in making the connection between the identification of coins and the commercial imperative of ensuring their liquidity:

> [I]n fungibles and all such things as are not discernible from others of that kind, possession is generally esteemed to constitute property, which is most evident in current money, which if it be not sealed, and during its remaining so, is otherwise undiscernible, it doth so far become the property of the possessor, that it passes to all singular successors without any question of

> the knowledge, fraud, or fault of the author; without which commerce could not be secured, if money, which is the common mean of it, did not pass currently without all question, whose it had been, or how it ceased to be his." (Stair, Walker (ed), 1681, 1981, II.i.34)

Banknotes presented a new legal problem, which the established rules about title to coins were ill-adapted to solve. Ordinary use among commercial parties treated banknotes as functionally equivalent to coins. As Lord Mansfield Chief Justice of the English Court of King's Bench said, banknotes were treated "as money, as cash, in the ordinary course and transaction of business, by the general consent of mankind; which gives them the credit and currency of money, to all intents and purposes" (*Miller v Race* 1758, 401). But banknotes, unlike coins, were specifically identifiable by their unique serial numbers. It was common practice for holders of notes to record these numbers as a precaution against loss or theft. If the title of a person holding the note was successfully open to challenge by a former holder who had lost the note, then their functional equivalence to coins would be lost. The issuing bank might face competing claims to payment of the note from the former and the current holders. The aim that notes would pass, as Stair said, "currently without all question, whose it had been, or how it ceased to be his" would be frustrated.

The problem was resolved by the litigation in *Crawfurd v The Royal Bank* (1749) in Scotland and by *Miller v Race* (1758) in England. In each case the result was dictated by the courts' concern to protect the liquidity of banknotes and the stability of the banking system (Reid 2016; Fox 1996), although the formal justifications drew upon the different civil law and common law traditions of each country. In Scotland the acquisition of the banknote was treated as a kind of consumption of property. Once property was consumed – whether in fact or by operation of law – the former owner's title to it was extinguished. In England, Lord Mansfield CJ formulated a common law rule adapted from the customary practice of merchants who handled banknotes and payment orders: "So, in case of money stolen, the true owner can not recover it, after it has been paid away fairly and honestly upon a valuable and bona fide consideration" (*Miller v Race* 1758, 457). In the analysis of both systems, transferees who acquired stolen banknotes in commercial transactions without any knowledge of their tainted provenance were assured of taking a secure title to them. Title no longer depended, as it did under the old rule, on the possibility that the note might be identifiable in the transferee's possession. In practice, this eliminated the need for the transferee to make any inquiries at all. Banknotes were assured of the same liquidity as coin.

Law and the Construction of the Monetary System

We have seen how the legal valuation of money depended on the private law rules on tender and discharge of debts (section "The Nominal Value of Money"). Thus the immediate question in *Gilbert v Brett* (1604) was whether the debtor had tendered money that complied with his obligation in the contract of sale. The creditor would

have been free to reject an invalid tender. The £25 limit on the monetary status of silver coin, imposed by statute in 1774, was expressed in terms of tender. A tender of silver was only valid – and in that sense "a legal tender" – if it was made in payment of debt of £25 or less. Beyond that limit, the creditor was free to reject it. The private law rules on tender, which the courts enforced by routine litigation in debt claims, translated the sovereign's public power of monetary valuation into practical results.

Legal tender rules defined the different categories of money into which the monetary system was organized. Money that enjoyed unlimited legal tender status was the primary kind of money in the system. The other forms of money were constructed from it. So until the change made in 1774, silver and gold coins were both the primary monies in the British financial system. A debtor could tender coins of either kind toward the payment of debts in any amount. Both counted as "the lawful money of England," which was the term often found in the payment clauses of contracts drafted by lawyers during the late medieval and early modern periods. The 1774 change demoted the silver coinage to a subsidiary money. It stood some way between the token coinage struck from copper and the gold coinage which, with its unlimited legal tender status, became the primary money of the system.

Coins struck under sovereign authority did not make up all the money in the system. They were supplemented by running credit arrangements, paper monetary instruments circulating from hand to hand, and eventually bank balances that were transferable by instructions given by the account holder. All these secondary forms of money consisted in legal debts. They were ultimately enforceable by a demand for payment in legal tender. Paper money and bank money derived their value from the fact that the holder or depositor might, at will, reduce the debt to payment in gold coin. Naturally, much of the convenience of issuing these secondary forms of money was that the creditor would not in fact enforce the debt in this way. Provided that a bank's credit was good, it was better for the holder of a bank account to hold his money as a debt than to reduce it to coin. Despite that practical preference, their legal form remained as debts capable of discharge by payment in coin.

When in 1833 Bank of England notes were made legal tender in England, they too become a primary kind of money alongside gold coin. There was no upper limit on the debts that could be discharged by the tender of Bank of England notes. All bank balances were reducible to payment in Bank of England notes or gold coin. The notes themselves were still enforceable against the Bank of England if the holder demanded payment in gold. The Bank could not pay the debt embodied in the note by tendering another of its notes! So although Bank of England notes were themselves legal tender, gold coin remained the base of the British monetary system throughout the 19th century.

This legal ordering of the system changed only in 1914 when the Bank of England suspended gold payments on its notes. Gold was withdrawn from general circulation. The suspension rebased the legal ordering of the system. Bank of England notes became the primary form of money, as they were the only form with unlimited legal tender status. Silver coins remained merely subsidiary. Once convertibility was suspended, the Bank of England notes derived their value simply by force of Parliamentary enactment. To be sure, when the United Kingdom went

back on gold standard in 1926, the relevant legislation went some way to rebasing the British monetary system in gold. The Gold Standard Act 1926 fixed the legal rate at which the Bank of England was bound to sell gold bullion to the public. But the effect of the Act was only to reset the value of sterling against foreign currencies which were also based on a fixed price for gold. The Act did not authorize the Bank to pay its notes in gold coin. Any duty to do so was specifically revoked. So in all domestic payment transactions, the legal position since 1914 was unchanged by the Act. Bank of England notes were legal tender. They represented debts but not ones that could be extinguished by payment of any other more basic form of money. The system had reached the final, chartal position described by Knapp is his *Staatliche Theorie des Geldes*. Money was a mere token. Its value in payments was decreed by law in the exercise of state power.

Cross-References

- ▶ International Monetary Regimes: The Gold Standard
- ▶ Origins of Money and Interest: Palatial Credit, Not Barter
- ▶ Premodern Debasement: A Messy Affair
- ▶ The Role of Money in the Economies of Ancient Greece and Rome

Bibliography

Blackstone W (1765) Commentaries on the laws of England, vol 1. Facsimile ed 1979. University of Chicago Press, Chicago/London

Bracton H, Thorne SE (1968) Bracton on the laws and customs of England. Harvard University Press, Cambridge, MA

Challis CE (1978) The Tudor coinage. Manchester University Press, Manchester

Cochrane-Patrick RW (1876) Records of the coinage of Scotland from the earliest period to the union, vol 2. Edmonston and Douglas, Edinburgh

Craig T (2017) In: Dodds L (ed) Jus Feudale Tribus Libris Comprehensum book I. 64. Stair Society annual volumes. Stair Society, Edinburgh

De Roover R (1948) Money, banking and credit in mediaeval Bruges: Italian merchant bankers, Lombards and money-changers; a study in the origins of banking. Medieval Academy of America, Cambridge, MA

Desan C (2014) Making money. Coin, currency, and the coming of capitalism. Oxford University Press, Oxford

Dondorp H (2016) The effect of debasements on pre-existing debts in early modern jurisprudence (chapter 13). In: Fox D, Ernst W (eds) Money in the western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford

Ernst W (2016) The legists' doctrines on money and the law from the eleventh to the fifteenth centuries (chapter 7). In: Fox D, Ernst W (eds) Money in the western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford

Fox D (1996) Bona fide purchase and the currency of money. Cambridge Law Journal 55(3):547–564

Fox D (2011) The structures of monetary nominalism in the pre-modern common law. *J Leg Hist* 34(2):139–169

Fox D, Velde F, Ernst W (2016) Monetary history between law and economics (chapter 1). In: Fox D, Ernst W (eds) Money in the western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford
Geva B (2011) The payment order of antiquity and the Middle Ages: a legal history. Hart Publishing, Oxford/Portland
Hale M (1736, 1847) History of the pleas of the crown., 1st American edition. RH Small, Philadephia
Justinian, Frier BW (eds) (2016) The codex of Justinian, vol 3. Cambridge University Press, Cambridge
Justinian, Watson A (eds) (1985) The digest of Justinian, vol 1. Pennsylvania University Press, Philadelphia
Knapp, G (1905, 1924), Staatliche Theorie des Geldes, published as The state theory of money. London: Macmillan and Company
Larkin JF, Hughes PL (1973) Stuart royal proclamations, vol 1. Oxford University Press, Oxford
Li M-H (1963) The Great Recoinage of 1696–1699. Weidenfeld and Nicolson, London
Menger K (1892) On the origin of money. Econ J 2(6):239–255
Murray A (1997) Sir Isaac Newton and the Scottish recoinage 1707–10. Proc Soc Antiq Scot 127:921–944
Pothier RJ (1773) Traité de droit civil. Deeure, J/Rouzeau-Montant, Paris/Orléans
Powell JM (1971) The Constitutions of Melfi or the Liber Augustalis promulgated by the Emperor Frederick II for the Kingdom of Sicily in 1231. Syracuse University Press, Syracuse
Proctor C (2012) Mann on the legal aspect of money, 7th ed. Oxford University Press, Oxford.
Redish A (2000) Bimetallism: an economic and historical analysis. Cambridge University Press, Cambridge
Reid KGC (2016) Banknotes and their vindication in eighteenth-century Scotland (chapter 25). In: Fox D, Ernst W (eds) Money in the western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford
Rüfner T (2016) Money in the Roman law texts (chapter 6). In: Fox D, Ernst W (eds) Money in the western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford
Stair, Viscount, Dalrymple, J; Walker, D (ed) (1681, 1981) The institutions of the law of Scotland. Edinburgh: Edinburgh University Press/Yale University Press
Thomas PJ, Boraine A (1994) Ownership of money and the actio Pauliana. Tydskrif vir hedendaagse Romeins-Hollandse Reg 57:678
Voet J, Gane P (ed) (1955–1958) The selective Voet: being the commentary on the pandects. Butterworths, Durban
Zimmermann CD (2013) A contemporary concept of monetary sovereignty. Oxford University Press, Oxford

Premodern Debasement: A Messy Affair

7

Oliver Volckart

Contents

Abstract

The chapter argues that in premodern Europe, the practice of debasement was far more "messy" than research has generally recognized. First, high information costs often prevented the effective control of mint officials who could exploit their resulting autonomy in order to debase coins on their own account. Second, these costs made it impossible to monitor markets closely enough to enforce regulations. Attempts by governments to debase coins by increasing their nominal value therefore "worked" only if they conformed to the market rates of these coins. Finally, high information costs prevented the creation of closed areas where the domestic currency enjoyed a monopoly. The resulting trade in coinage created incentives for governments to issue inferior copies of their neighbors' coins – a practice that had the same consequences as a debasement – and forced the affected governments to follow suit by debasing their own coinage, too.

O. Volckart (✉)
Economic History Department, London School of Economics and Political Science, London, UK
e-mail: o.j.volckart@lse.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_8

Keywords
Commodity money · Debasement · Counterfeiting · Seignorage · State capacity

Introduction

On New Year's Eve 1519, Albert of Brandenburg (1511–68), Grandmaster of the Teutonic Order that had ruled Prussia since the thirteenth century, began a war against the neighboring kingdom of Poland. In order to finance this undertaking, in particular to pay the mercenaries he engaged in Germany and Denmark, he instructed his mint to increase its output. The mint master and moneyers obliged. Churches were stripped of their ornament and the subsidies sent by Muscovy and Livonia were directly handed over to the mint to be melted and reissued as Prussian coins (Volckart 1996: 242). However, after the Holy Roman Emperor Charles V (1519–56) had mediated an armistice, trouble began. A contemporary annalist noted that when the lansquenets were disbanded and

> moved off on the Tuesday after *Misericordia Domini* [16 April 1521, OV], there was much scolding and cursing because no-one wanted to accept the coins with which they had been paid. They therefore took their pay with them and were ferried from Tolkemit across the lagoon to the Vistula Spit, and took their way from there across the mouth of the Vistula towards Pomerania. And from Danzig they were provisioned with beer and bread and other things. So the Danzigers received a great many of these coins, too; moreover, many bought them for half their value, believing to make a good deal. However, once the intrinsic value of this money had been determined, it was soon despised throughout the whole country. (Meckelburg 1865: 158 f.)

The initial acceptance of a ruler's new money, followed by disillusionment and rejection, was a common pattern under conditions such as those the Prussian annalist described. The episode thus illustrates a typical consequence of a widely practiced type of monetary policies in premodern Europe: debasement, that is, in its widest definition, an increase in the ratio between the nominal and intrinsic value of money. In the Prussian case, the ruler ordered the bullion content of his coinage to be reduced. In addition, the mint resorted to labor saving strategies, issuing square coins that were quicker to cut from the metal sheets than properly rounded ones and focusing on minting large denominations. In nominal terms, this allowed the production of huge sums of money while requiring relatively less labor than the smaller round coins that had been common until then (Volckart 1996: 243).

Reducing the bullion content of coins relative to their nominal value – i.e., replacing part of the pure silver or gold they contained with a base metal such as copper – was only one of several ways of how a debasement might be carried out. There were two other options. A ruler or government could reduce the weight of the coins that his mint issued while maintaining both the composition of the alloy of which they were made – i.e., the proportion of bullion to base metal – and their official value. Alternatively, both the weight and the fineness of the coins could be left unchanged; what the minting authority did was increase their official value. This

third option was particularly tempting because premodern coins were rarely marked with their "face" values. Thus, the French *gros tournois*, introduced as a 12-*deniers* piece under Louis IX (1226–70), did not show the numeral "XII," though its design did contain twelve *fleurs de lis*, which presumably were intended to indicate its nominal value (cf. Blanchet and Dieudonné 1916: 227). By contrast, nothing in the design of the English groat first minted at the time of Edward I (1272–1307) betrayed that it was to circulate as a 4-pence piece.

The monetary toolkit of premodern rulers and governments contained few instruments apart from debasements (and their rarely used opposite, i.e., reinforcements). This did not fundamentally change before the end of the early modern age when central banks issuing paper money became common throughout Europe. Thus, in the mid-eighteenth century, Frederick II of Prussia (1740–86) debased his coinage in much the same way and for the same reasons as his distant relative, Albert of Brandenburg, had done two and a half centuries before (with the seignorage covering about one-third of his expenditure during the Seven-Years War, Koser 1900: 359), or, for that matter, as the Dauphin, later Charles VII of France (1429–61) did in the early fifteenth century (who increased his revenues by a factor of six by debasing his coinage, Sussman 1993: 66 f.). In monetary politics, the traditional dividing line between the Middle Ages and early modernity in about 1500 is thus even less appropriate than in other fields of history (cf. Scott 2015: 19 ff.). If a line needs to be drawn, it is in the late thirteenth and fourteenth centuries when European polities began issuing complex currencies composed of multiple denominations and based on different precious metals. This is also a question of research practicalities. As chemical analyses of surviving coins have shown, debasements were frequent in the earlier Middle Ages, i.e., between roughly the sixth and the thirteenth centuries (Metcalf and Merrick 1967). However, in most places so few written sources have survived from that time that we are on shaky ground when we want to examine such policies. In many respects, therefore, the period from the fourteenth to the middle of the eighteenth century appears as a historical unit not only because monetary policies remained essentially the same but also because we are in a position to analyze them.

Until not so long ago, debasements remained poorly understood. A high-profile paper by three eminent economists (Rolnick et al. 1996) even talked of a "debasement puzzle." In fact, as shown below, there is no such puzzle; debasements were at least in principle fairly straightforward measures. A number of recent articles have done much to systematically explain how and why they occurred. Notably, Chilosi and Volckart (2017) present a statistical analysis of the relative importance of the different motives and aims pursued by authorities who debased their coinage. Summarizing core insights from a lifetime of research, Munro (2015) used one of his last papers to set out the technology and economics of debasements in the late Middle Ages and at the beginning of the early modern age. In view of these advances in recent research, the present chapter aims at presenting debasements in the wider context of contemporary conditions in society, the economy and specifically monetary policies. Its hypothesis is that while the principles underlying debasements are clear enough, the practice of debasement deviated in many places from the picture drawn by prior research. The chapter argues that this was mainly the case because of

high information costs, which not only created conditions that occasionally left authorities little choice but to debase their currency, but which also prevented them from effectively implementing such measures.

The Principles of Debasement

In order to explain how debasements worked, it is useful briefly to recapitulate the mechanisms of the money supply in the period considered here. The following sketch summarizes the standard analysis of the process (cf. Gould 1970: 7 ff.; Rolnick et al. 1996: 790; Munro 2015: 19 ff.). The fundamental assumption is that only metal sold voluntarily to the mint would be minted. In order to do so, the mint would offer a price for any given quantity of pure gold or silver it wished to buy; this was the mint price. The total nominal value of the coins made from this quantity of metal was the mint equivalent. The mint equivalent was always higher than the mint price, with the difference used to cover the rest of the production costs. After all, the mint also needed to buy base metal to produce the alloys from which the coins were struck and firewood or charcoal to melt the metal. Moreover, the upkeep of the mint buildings involved costs, the tools needed to be manufactured or purchased and the workmen to be paid. Last but not least the leading official, i.e., the mint master, needed an income. This usually consisted of what was left of the newly minted money after he had paid all the other costs and the seignorage: the share due to the political authority on whose behalf he was minting.

One implication of the mint equivalent being higher than the mint price was that the nominal value of a coin regularly exceeded its intrinsic value. Put differently, coins normally circulated at a value that was roughly equal to the sum of their production costs and seignorage (cf. Sargent and Velde 2002: 18 f.). The difference between the intrinsic and nominal values had itself a further implication that was crucially important for how debasement worked. Regardless of whether a political authority chose to reduce the weight of a coin, to reduce the proportion of pure precious metal it contained, or to increase its official value, the measure always entailed a growth of the mint equivalent. This, in turn, allowed the authority to increase its seignorage and the mint to offer a higher mint price in order to attract more bullion. Hence, debasements allowed both minting volumes and revenues from seignorage to grow. To give an example, when in the early 1540s Henry VIII (1509–47) debased the English silver coinage by reducing the purity of the metal, the mint in the Tower of London was able to pay merchants a higher nominal price for the silver they supplied. As a consequence, the output of the mint increased by a factor of nine between 1542 and 1546 (Gould 1970: 38). However, the increase in the mint price of silver meant that in relative terms, gold became cheaper. Therefore, in order to prevent merchants from buying up English gold coins and exporting them, gold had to be debased, too. This is what happened, though the purity of the gold coins remained largely unchanged: Rather, their nominal values were increased (Gould 1970: 48 f.). This episode and similar events in particular in the Burgundian

Netherlands of the late fourteenth and fifteenth centuries persuaded Munro (2015: 16 f.) that debasing silver usually forced a government to debase gold, too.

Asymmetric Information and Social Power

The "debasement puzzle" identified by Rolnick et al. (1996) concerned the causes of the increase in mint output after a debasement had been implemented. They argued that as bullion merchants were aware of the change in the ratio between nominal and intrinsic value which the debasement involved, there would have been no incentive to supply the mint with additional gold and silver that could then be used to produce a higher number of coins. It is indeed plausible that at least some merchants who sold bullion to mints knew about the fine silver or gold content of the coins they received in return. Still, the perceived puzzle has a fairly straightforward solution.

In order for a debasement to "work" in the way described above, specifically in order for mints to be able to increase their output after a debasement, the wider public – that is, not the bullion merchants – had to keep focusing on the nominal value of coins rather than on their intrinsic value at least for some time after the measure had been taken. In other words, when a ruler or government debased the coinage, bullion merchants would keep supplying the mint with raw metal as long as they expected to find consumers willing to use coins "by tale" rather than "by weight." For this to happen, at least one condition was necessary: Either it had to be possible to pass off debased money as unadulterated or to force consumers who knew of the debasement to accept the coins at a value higher than that their bullion content justified. The first condition seems frequently to have been met. For example, in Prussia in 1521, the lansquenets initially seem to have assumed that the pay they received consisted of relatively "good" coins, i.e., even if they expected some debasement, they were not aware of its extent. The issue had two aspects. First, for consumers such as them, it was hard to determine what the legal standard of their coins actually was. Governments may have published mint prices, but most kept information concerning, e.g., mint equivalents closely guarded secrets (Redish 2000: 28). Even where this information was published – as, for example, in the Holy Roman Empire where currency competition was intense and governments seem to have hoped that the public would help weed out underweight coins – it was accessible only to those able to read and understand legal texts (Volckart 2017a: LIXX f.).

Literacy – or the lack of it – therefore played an important role where debasements are concerned. To give an idea of the order of magnitude about which we are talking here: it has been estimated that over the sixteenth century, the share of the population able to read and understand complex texts such as for example currency laws grew to more than 5% in the German-speaking parts of the Holy Roman Empire (Engelsing 1973: 32). Using the number of new book editions as a lead indicator, this share must have about tripled by the mid-eighteenth century (cf. Buringh and van Zanden 2009: 421). There were obviously wide social and regional variations in literacy which moreover was always higher among men than women (Houston 2014: 142–162). Still, for most of the period discussed here and most parts of Europe, the

vast majority of consumers were illiterate in the sense of being unable to read and understand monetary ordinances, if these were published at all. It is therefore fair to assume that they had no idea of how much bullion their money was supposed to contain, and lacked any standard against which to measure its intrinsic value.

In principle, of course, knowing about the legal standard of the coinage was not necessary if one was able to determine whether the intrinsic value of the coins one handled had been changed, i.e., if one could compare the bullion content of new coins to that of older ones of the same official value. However, doing this was costly and time consuming. Most importantly, it required knowledge and tools that only a small number of experts had. To check the weight one needed a set of precision scales and weights, to check the purity at least a touchstone and a set of touch needles. The needles left lines on the touchstone whose colors one could compare with that left by the coin one was testing: A matching color signaled a matching degree of purity. This method allowed an accuracy of ±2–3% (Redish 2000: 22); more precise analyses required melting the coin, separating the components of its alloy and determining their relative weight (the method is described by Fachs 1678: 47–51). To do this, one needed metallurgical tools and in-depth knowledge and skills possessed by a very small share of the population: some merchants used to handling precious metal, professional money changers and of course mint masters. The majority of consumers – mostly illiterate peasants or craftsmen, small hucksters, local traders and certainly the lansquenets Grandmaster Albert of Brandenburg employed – were in no position to check the intrinsic value of their coins.

Under these circumstances, debasements worked because governments could rely on few people immediately becoming aware of a reduction in the weight of a coin or in its bullion content. At some point, of course, the news would spread: the informational asymmetry would be eroded. In Prussia, this took some time. It was a neighboring ruler, the bishop of Warmia, who first became suspicious and ordered a test of the money early in 1521. The result was so appalling that he notified the council of Königsberg (the Grandmaster's residence-city), whose members immediately lodged a complaint with their lord: "They were truly terribly shocked; they neither would nor could reveal this to their commune, as your grace can easily imagine what moaning and wailing would follow" (Meyer 1912: 596). The episode illustrates how different groups in society might become aware of debasements at different times: Merchants and urban councils early on, the other members of urban communes later, and crucially, the group whom the government was most interested in "fooling" – in this case the lansquenets – last. From the perspective of the Grandmaster of the Teutonic Order, his policy was a success.

There was another condition that could contribute to the success of debasements: the hierarchical structure of premodern society (Münch 1996: 65 ff.). The term hierarchy is to be taken literally. Almost any member of society was regularly confronted with someone else who was not just his or her superior in some poorly defined social sense, but who enjoyed more privileges, freedoms and often the right to exert some form of political authority. Over time, the attempts of monarchs to establish monopolies of force and to level legal distinctions among their subjects caused an erosion of this kind of hierarchy – a process that began earlier and was

quicker in some countries (e.g., England and the Dutch Republic) than in others (for example in Spain, France, Germany, or Poland) (van Zanden and Prak 2006, Hettling 2015: 124). However, over much of Europe and most of the late Middle Ages and the early modern period, social hierarchies remained strong.

Even where consumers already knew about a debasement, they were therefore frequently in no position to resist their betters when these used debased money to pay wages or goods (Rössner 2012: 585 f., 2014). Hence, informational asymmetries between producers of money and consumers were no necessary condition for debasements to be feasible. Strong social hierarchies might be sufficient. There is evidence that consumers were at least occasionally and to some extent aware of the quality of the coins with which they were paid. Thus, in mid-fifteenth-century Freiberg in Saxony, an important mining center where metallurgical information was probably easy to come by, laborers repeatedly protested against the payment of wages in debased coin (Laube 1976: 208 f.). According to Rössner (2012: 491 ff.), being forced to accept debased money contributed crucially to the decision of peasants in Upper Germany to revolt in 1525, and indeed, complaints about the circulation of debased coins were commonplace in the early sixteenth century. Some consumers, though, did not rebel but reacted in more subtle ways. For example, in late fifteenth- and early sixteenth-century Nuremberg, coins of doubtful value appeared over proportionally often in the churches' collection boxes (Groebner 1993: 48).

Recent research thus has established the principles underlying debasements and the conditions that allowed them to "work" from the ruler's or government's perspective: It has solved the "puzzle." However, it is striking that this research has so far focused on a relatively small number of examples mainly derived from England, France, Burgundy, and Italy. As will be shown below, this is problematic because these were not only economically the most advanced parts of premodern Europe, but also politically the best organized. Late medieval Burgundy and even more Tudor England stand out as polities where governmental decisions were implemented more effectively than almost anywhere else in Europe at the time. It is the core hypothesis of the present chapter that in most parts of the continent debasements – and monetary policies in general – were far more "messy" affairs than in those countries on which research has tended to focus. In order to explain why this was the case, it is necessary to extend the scope of the argument.

Information Costs, Political Capacity, and Monetary Policies

One crucial factor has already been mentioned: information. When implementing debasements, rulers and governments usually benefited from the poor state of information of the consumers using debased money. The information needed to make decisions about which coins to accept and which to reject, be it at a mint or on a market place, was difficult and costly to come by. However, while in this respect authorities benefited from high information costs, there were other ways in which the same costs affected them. Not only was literacy weak, with many branches of

administration using written records sparingly, information also spread slowly. With the exception of rare and not particularly reliable carrier pigeons, nothing moved more quickly than a horse could run. Thus, while twelfth-century express messengers could cover up to 40 miles per day, improvements in infrastructure meant that by the sixteenth century distances of up to 60 miles were reached (Denecke 1990: 217). Regular postal services – introduced first in the Empire at the end of the fifteenth century – improved information flows further. In 1500, the mail needed 30 days to cover the 400 miles between Hamburg and Augsburg; in 1615, 11 days; and in 1800, 5 days (Behringer 2005: 48). Still, while the transmission of information became more reliable, its speed did not increase as much as the travel times of mail riders suggest. Even at the end of the eighteenth century, their average speed was only about 25% higher than that achieved by express messengers 300 years before.

Conditions such as these had several implications for monetary policies. For one thing, high information costs made implementing decisions difficult. Again, the problem had several aspects. Thus, on occasion rulers even found it hard to learn what was going on in their own mints. For example, in 1527, the councillors of Albert of Brandenburg – who, after his failure in war, had turned the state of the Teutonic Order into a secular duchy and a Polish fief – found it necessary to interrogate the mint master, the leading official of the Königsberg mint. Accounts the master might have kept obviously did not exist. The result of the interview was shocking. Not only had the duke failed to engage a warden of the mint, that is, a supervisor responsible for testing newly minted money. The mint master had also never transferred any seignorage to the ducal treasurer, who himself admitted that he could not remember whether the seignorage had ever even been mentioned between the two of them. When finally asked on whose authority he was minting, the master declared "he hoped that if he had minted [. . .], this would not be to my gracious lord's [the duke's, OV] displeasure [. . .]. He had used his own silver and the residual left in the crucibles, and had only minted to meet his personal expenses" (Volckart 1996: 409 f.). In other words, the mint master acted without any supervision, running the mint like a private enterprise where he produced any amount of money he felt he needed. Nothing suggests that conditions like these were unusual. In late medieval and early modern Europe there were many hundreds of small, workshop-like mints, in the Holy Roman Empire alone c. 500 in about the year 1500 (Sprenger 2002: 81). Most of these operated on behalf of small towns, lords or abbeys who lacked the resources needed to cover the high costs of supervising officials. Moreover, rulers in many parts of Europe farmed out their mints – often for years at a time – which implied a far-reaching loss of control (Spufford 1988: 17).

High information costs also affected the ability of a ruler or government to enforce monetary regulations vis-à-vis the public. As the vast majority of consumers were unable to read and understand handwritten or printed edicts, such regulations had to be announced by word of mouth – a process prone to giving rise to all kinds of misunderstandings. Another Prussian example shows how officials tried to deal with this. A few years after the scandal about the Königsberg mint, the duchy of Prussia entered a currency union with Poland. When the new money was introduced, it

proved – as usual (Miskimin 1985/89: 148) – impossible to withdraw the old coins from circulation; they had to be allowed to continue to circulate at fixed rates expressed in new money. This caused a large amount of confusion. When taken to task by the ducal councillors, the mayor of Königsberg defended himself:

> What had happened [. . .] at the start of the annual fair was that such a crowd of common people, who had come in large numbers to visit the fair, had appeared and inquired about the coinage that he could in no way answer all their questions and correct their mistakes. However, he had ordered the town servants to walk up and down the streets to inform people about how to give and take the coins in accordance with the edict [about the new coinage, OV]. But even if there had been fifty more of them, they would not have been enough, even if each had answered a hundred and more people a day; my gracious lord's councillors might imagine when he himself might rest his mouth, and if things continued like this he would rather be a swineherd than mayor, and gave notice of much other hardship and annoyances he had suffered because of this. (Volckart 1996: 415)

On top of the high costs of informing the public about monetary policy, a second problem appeared: Once consumers (or at least some of them) had learned about a political decision, premodern governments were generally unable to monitor markets closely enough to determine whether or to what extent their regulations were observed. This is one of the factors that brought down Charles V's project to establish a common currency valid through the whole Holy Roman Empire in the years between 1551 and 1556. Like in the Prussian case a few decades earlier, the old money was supposed to continue to circulate next to the new Imperial currency, and at fixed rates expressed in new money. To give consumers an incentive to sell old coins to the mints that were to melt and reissue them as new Imperial money, old coins were officially undervalued. Most importantly, the value of Saxon *talers* in new coins was set below their intrinsic value. *Talers*, however, were hugely popular – so popular, in fact, that in the northern parts of the Empire they circulated at a premium of about 10%. Charles V's idea that it might be possible to monitor thousands of markets so closely that consumers were forced either to use *talers* at their official rate or to sell them to a mint was delusional (Volckart 2017b: 773 f.). Nothing like that was remotely possible: Again, high information costs precluded it.

In addition to the difficulties of informing the public about monetary policies and monitoring markets to make sure that regulations were observed, a further problem appeared. Premodern authorities were generally unable to effectively control the borders of their territories. In fact, in most of Europe proper external custom borders did not develop before the eighteenth century, with Bavaria creating them in 1765 and Austria 10 years later (Dipper 1991: 176). Before, trade was taxed where it was cheap to monitor: at river crossings, mountain passes, or city gates. In consequence, it was neither possible to hinder the export of domestic coins, nor to prevent "foreigners" from importing their money and spending it. A survey of coins circulating in the Empire made in preparation of Charles V's currency union project found that 97 of the 187 types of gold and 37 of the 134 types of silver coins had been minted abroad (Volckart 2017a: 318 ff.). The issue of the parallel circulation of several currencies seems to have been particularly acute in Germany, where

consumers used French, Dutch, Danish, British, and American coins in day-to-day transactions as late as the second half of the nineteenth century (Helfferich 1900: 188 f.). However, it was by no means restricted to the Holy Roman Empire. In France in 1640, for example, the crown ordered approximately 45 million *livres tournois* worth of foreign gold to be re-minted into French coins; according to modern estimates, this was about 20% of the total quantity of French money (Glassman and Redish 1985: 44 f.). The quantity of foreign silver that continued in circulation is not known.

Conditions such as those described above did not exist everywhere and at all times to the same extent. With the fall of information costs in the course of the early modern period, polities became more effective. Moreover, island governments always found it cheaper to monitor cross-border traffic and began doing so already in the Middle Ages. Thus, by the end of the thirteenth century, merchants visiting England had to declare what foreign money they brought with them and had to exchange it at politically determined rates for English coins; likewise, the export of domestic money was prohibited. There seems to have been a concerted attempt to enforce these rules, but even English consumers used foreign coins. Venetian *soldini*, for example, circulated as "galley halfpennies" (Prestwich 1979: 471, Spufford 1963, Daubney 2009). Thus, in most places and for most of the period here considered, governmental ordinances that ordered all foreign coins (and imported precious metal) to be surrendered to the mint to be converted into domestic money were rarely more worth than the paper on which they were published (cf. Munro 2015: 17 f.).

Mint Autonomy

While there is some evidence that, occasionally, premodern rulers tried to influence the composition of the coinage issued by their mints (this was sometimes contractually agreed between rulers and their mint officials. For an example see Klempin 1859: 585 f.), it was often up to the leading officials to decide which type of coins to strike – the more so, of course, if these officials were as poorly supervised as the mint master of the former Grandmaster of the Teutonic Order in 1527. What this implied for the policies practiced by the mint was described in 1532 by the former monk and metal expert Burkhard Waldis, who wrote a monetary memorandum for the Master of the remaining possessions of the Teutonic Knights in Livonia. Waldis argued that "if they [i.e., the mint masters, OV] may mint more than just one type of coin, they mint that one most which allows them to make the largest profit" (Arbusow 1910: 799). That the production of different denominations allowed profits of varying size was one of the core problems of premodern currency systems. It was primarily caused by the share of labor costs in the total production costs of the coinage. In late medieval England, for example, the labor costs involved in producing a farthing, i.e., a quarter penny, were about the same as those that had to be spent in producing a groat, whose value was four pence. As a groat contained exactly 16 times as much pure silver as a farthing, in nominal terms it was possible to produce 16 times more

money when the mint issued groats while spending largely the same amount on labor costs. As a consequence, late medieval England suffered from a chronic undersupply of small change. Most polities tried to compensate for the higher share of labor costs involved in the production of small coins by reducing the share of precious metal they contained. In this way, small change became token money that was the more overvalued relative to its intrinsic value, the smaller its nominal value was. The problem was that rulers or governments who chose this solution walked a tightrope. If they did not reduce the share of bullion in the small coins far enough, there would still be incentives to focus on the production of large coins; if they reduced it too far, the mint would issue small change only. This happened for example in early fifteenth century Prussia, where pennies were strongly overvalued relative to shillings. The result was that the country was flooded with small change (Volckart 1996: 90 ff.).

Developments in the Holy Roman Empire of the late sixteenth and early seventeenth century show which problems the ineffective supervision of mints by their rulers could cause in this context. One of the issues that Charles V's attempt to create a common Imperial currency in 1551 had thrown up was that small change (defined as all coins below a value of 6 *kreuzers*) was overvalued too strongly. It was therefore more attractive to mint smaller than larger coins. Moneyers rapidly took advantage of these incentives, occasioning frequent complaints throughout the 1550s (Volckart 2017a: 405, 410; see the next section for why this happened). Accordingly, the Augsburg Imperial Diet of 1559, which reformed Charles' currency, explicitly demanded that small coins should contain so much silver that mints would have no reason to focus on issuing them to the detriment of larger denominations (Leeb 1999: 1376). However, the new Imperial currency law published later that year again missed the mark (cf. Fig. 1). This time small change contained too much bullion; producing it cost-effectively was now impossible. As a result, while law-abiding princes such as the Elector of Saxony did not issue any small coins in the decades thereafter, many poorly supervised mints debased them illegally (Wuttke 1897: 248 f.). In particular, mints that the authorities had farmed out to private entrepreneurs, who acted without any supervision at all – so called "Heckenmünzen," literally, mints hidden behind a hedge –, were notorious for contributing to the oversupply of debased small change (Schneider 1995: 156 f.). This was the main factor that brought the Imperial currency – which in other respects was well-designed and innovative – into disrepute (Volckart 2017a: LXXIII).

Silver Debasements and the Price of Gold

As shown above, high information costs not only prevented many authorities from effectively monitoring their agents; they also made it difficult to inform the public about monetary policies and to implement these on the ground. This applied to measures on which debasements depended, too. The most striking instance of this is the manipulation of the official value of coins, which was one of the ways a debasement might be carried out. Attempts to do this worked only under very restrictive conditions.

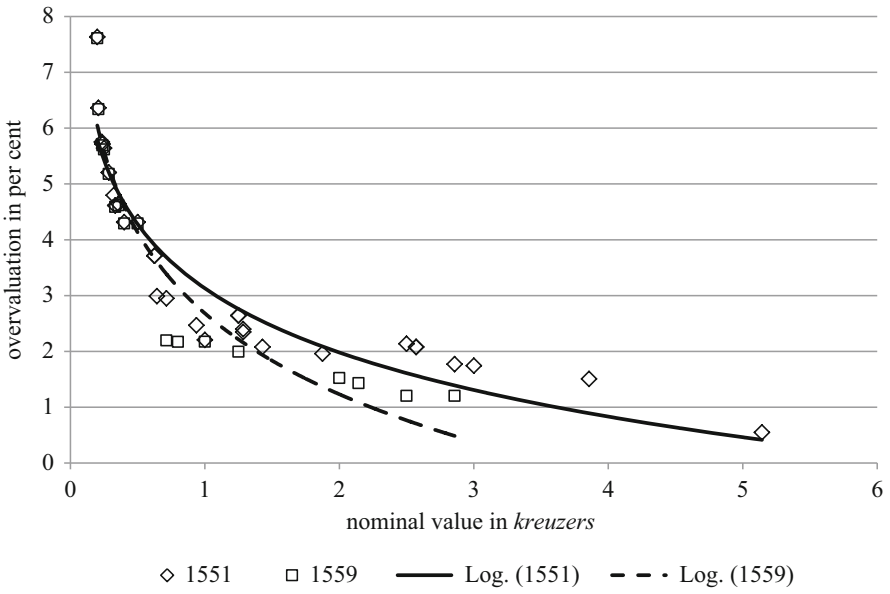


Fig. 1 Overvaluation of small change in the Holy Roman Empire, 1551 and 1559 (the graph shows the difference between the mint equivalents of small change (coins below the value of 6 *kreuzers* in 1551, 5 *kreuzers* in 1559) and larger full-bodied coins in percent of the mint equivalent of these larger coins. Both currency laws allowed the common Imperial currency to be supplemented by different types of regional small change). Sources: Volckart (2017a: 346–352), Leeb (1999: 1958–1962).

The core issue here was that in the last resort, the value of monetary units was not determined by rulers or governments but by the consumers themselves. This was the case because normally, the authorities lacked the means to force the public to use money at its official value. If they could have done this, debasements that concerned small change only would have triggered Gresham's law: Consumers would have reacted by holding back the larger, full bodied coins and would have paid goods and services with small change only. However, given that authorities lacked the resources to monitor markets closely enough, the normal consequence of a debasement of small change was that the "exchange rate" between it and larger denominations of the same currency increased (cf. Redish 2000: 30). This is what happened in Prussia in the years around 1500, where the pennies were debased and the public reacted by using the 15-penny pieces that the mint issued at a rate of 18 pennies (Volckart 1996: 230). Such developments were not necessarily a consequence of debasements, though. In the case of the coins minted according to Charles V's currency law of 1551, it was the excessive supply of small change that drove up the exchange rate of larger coins. Alternatively, a coin might be so popular that consumers were prepared to pay a premium in order to get hold of it. Thus, the Saxon *taler* that was intended to be a 21-*groschens* piece began to circulate at 24, for some time even at 25 *groschens* in the decades before 1550 (Arnold 1980: 64) – this

without any reinforcement of the *taler* or debasement of *groschens*, and without anything pointing to an oversupply of the smaller coins. Likewise, the Austrian silver gulden minted since 1521 soon increased from 60 to 64 *kreuzers* (Schalk 1881: 328). The upshot is that governments were usually incapable of fixing the value of a coin not only in relation to goods and services or to monetary units belonging to other currencies but relative to other denominations of the same currency, too. This is not to imply that there was an entirely open market where "exchange rates" were freely negotiated. Even if authorities failed to enforce politically imposed coin values, any attempt to do so would have some effect on supply and demand. Moreover, as shown above, there was no level playing field as the higher social standing of some consumers gave them more bargaining power than others. Still, it is a generally valid conclusion that governments had little influence on the value of coins, which consumers determined in a process of negotiation – skewed though this must have been – among themselves (cf. Rössner 2012: 464 f.).

This applied to gold, too – all the more so as many types of gold coins were used in long-distance trade and therefore outside the region or town ruled by the authority that had issued them. However, even domestically, the official parity between gold and silver proved normally impossible to enforce. For example, in 1468 the city of Lübeck debased its silver by almost 10%, with the bullion content of the double shilling falling from 2.82 to 2.53 g (Ropp 1890: 38, 66 f.). The gold content of the Lübeck gulden was left unchanged at 3.51 g (Jesse 1928: 218). At the same time, Lübeck's council attempted to stabilize the gulden's nominal value at 28 shillings. Consumers did not take heed. On the market, the gulden had been traded at 30 shillings even before the debasement; now it jumped to 32 shillings (Ropp 1890: 34, 41, Jesse 1928: 216, 218). This was typical. Premodern authorities, whether they were urban governments or princes, were generally unable to enforce the circulation of their gold at its nominal par value (Miskimin 1985/89: IX, 149), or to be precise, they were able to do so only if the legal par value conformed to the – in whichever way negotiated – market value. This condition was so restrictive that authorities debasing their silver usually did not concern themselves with adjusting the value of their gold. In other words, many did not even attempt to supply a bimetallic currency where there was a legally fixed ratio between gold and silver coins; rather, they supplied two distinct parallel currencies that were based on different precious metals and whose relative value was floating.

Debasements and Counterfeits

As explained above, there was a third consequence of high information costs which crucially affected debasements: the porousness of currency borders. Fiscal concerns such as those that drove, e.g., Albert of Brandenburg to debase his coinage were of course in many cases obvious. However, the literature's frequent focus on this motive obscures the importance of the *de facto* lack of borders. It was problems resulting from this lack which created incentives to pursue policies that had the same

effects as debasements; these, in turn, caused problems that many late medieval and early modern governments tried to address by debasing their own currencies, too.

By the fifteenth century at the latest, merchants had begun to exploit the porousness of borders by engaging in what policy makers called a "trade in coinage." A position paper presented at a conference convened by the princes and free cities of the Holy Roman Empire in 1549 in order to discuss Charles V's monetary policies described the phenomenon. It stated that "masses of *batzens* [popular South German 4-*kreuzers* pieces, OV] are exported from the Empire to be melted and re-minted." Conversely, "much Italian and French silver money, which matches neither the German money's weight nor its fineness, is imported into Germany; this money is being made of the raw silver and *batzens* previously exported." Rhineguldens, i.e., the gold coins most commonly used in fifteenth- and sixteenth-century long-distance trade, were also "bought up in Germany and exported to Italy where they are converted into Italian crowns and then re-imported. This has been turned into a regular trade" (Volckart 2017a: 40 f.). A similar trade flourished within the Empire. Johann Oldecop, a contemporary annalist from North Germany, pointed to the incentives it created for rulers issuing coins. According to Oldecop, in the years around 1550

> many counterfeit *talers* were issued, in part underweight and in part made of lead, some of copper, some of poor bullion. Moreover, the authorities of the lands where these invalid and false *talers* were minted struck them with fraudulent legends so that anyone who did not carefully test the *talers* would believe they were genuine and had been issued by some pious prince or town. [. . .] Devilry, wickedness and deception were at that time considered a vocation and proper mercantile business practice: Many hucksters and other merchants let coins be minted from coins [. . .]. Others travelled about, offering their goods throughout the country or in army camps where they bought up *talers* or gold with counterfeit coins. (Euling 1891: 380)

Oldecop was a cleric and theologian and his moral outrage needs to be taken with a pinch of salt. However, there is no doubt that trade in coinage was widespread. It worked and was profitable for two reasons. First, as suggested by Oldecop, underweight copies often resembled the original coins. "Replicating other authorities' designs" was considered a common malpractice in premodern coinage (cf. Volckart 2017a: 233, 364). From the perspective of the copyist, this made sense. For example, when Duke Philip the Bold (1363–1404) of Burgundy began issuing imitations of golden English Nobles in 1388, merchants flocked to the Burgundian mints to sell the English originals: unsurprisingly so, as Philip's nobles contained 7.63 g of pure gold while those minted in England contained 7.78 g (Spufford 1963: 129 f., Munro 1972: 43–63). Obviously, the merchants realized that the quality of the Burgundian coins was lower. However, they counted on being able to use them on par with the English originals, either relying on finding transaction partners who did not know the difference or on their superior bargaining power that would allow them to pressure well-informed consumers into accepting Burgundian nobles in place of English ones. The "war of the nobles" in about 1400 is a prominent example but more than 350 years later, much the same kind of policies

were still practiced. In the middle of the eighteenth century, genuine Saxon *talers* contained 17.48 g of pure silver, whereas the masses of imitations produced by Frederick II of Prussia (and exported to Saxony) during the Seven-Years War held as little as 3.24 g (Engel 1855: 53, Schrötter 1910: 509). Prussian imitations of Polish coins were equally numerous and equally poor (Koser 1900: 344–351, Hoensch 1973: 121 ff.). While here it was the government that was responsible for the decision to produce inferior imitations, this was not necessarily the case. As Burkhard Waldis pointed out in the memorandum quoted above, poorly supervised mint officials would strike underweight coins autonomously, intending to "attract their neighbours' money that they then remint in their own workshops" (Arbusow 1910: 799).

The examples demonstrate how fluid the transitions between several types of monetary policies were. What Philip the Bold did was no debasement in the strict sense of the word, as he introduced a monetary unit that was new at least within the structure of his own currency. However, given the close similarity between his own nobles and the English ones, it functioned like a debasement, attracting bullion to the Burgundian mints. Frederick II's "Saxon" *talers* and "Polish" coins had a similar effect. However, while the design of Philip the Bold's nobles showed some subtle differences to that of the English coins, which protected him against being accused of counterfeiting, the Prussian coins looked so similar to the originals that they must be considered forgeries and have been called this by contemporaries and later research (Gumowski 1948; an Austrian propaganda pamphlet rather disingenuously accused Frederick II of forging British guineas, too (Anon. (Egid Valentin Felix von Borié) 1761). Frederick depended on British subsidies for about one-third of his expenditure during the Seven-Years War; Koser 1900: 359). The examples thus go to show how blurry the line between debasements and counterfeiting was in premodern Europe.

There was a second reason for why the trade in coinage was profitable. This was linked to the use of similar units of account based on distinct actual coins. Again, the process is well attested in the Empire, where a new unit of account that facilitated trade evolved in about 1500. Towards the end of the fifteenth century, many princes and towns had tried to link their local silver currencies to gold by legally fixing their exchange rate with the rhinegulden: For example, 60 Austrian *kreuzers*, 20 Braunschweig *mariengroschens*, 21 Saxon *zinsgroschens* and 32 Brandenburg *märkische groschens* were declared to be the equivalent of 1 rhinegulden. Initially the ratio was stable: In, e.g., Vienna, the relative market price of gold and silver as expressed in the exchange rate between rhinegulden and *kreuzers* remained unchanged between 1500 and 1522 (Schalk 1881: 260 f.; moreover, by the early sixteenth century currency markets were well-integrated so that local gold-silver ratios were fairly close; Chilosi and Volckart 2011: 760). As a consequence, consumers began to call the sum of 60 *kreuzers* a gulden even when they did not handle actual gold coins. The same happened in other parts of the Empire where the gold-silver ratio was unchanged for long enough. It began generally to shift from the 1520s onwards, with the value of gold increasing in many places, but by then the public had become used to treating

the gulden as a unit of account whose value was expressed in different quantities of local silver coins. This new unit of account was still the same in many parts of the empire, but it was now divorced of the exchange rate between these silver coins and the actual golden rhineguldens and was maintained even when some princes and towns debased their silver (Volckart 2017a: XXXIII). The result was that by the 1530s, a plethora of silver coins were in circulation whose intrinsic value differed but whose ratio to the gulden of account was still fixed.

Two so-called assays – official metallurgical tests of current coins – show how large the deviations were which this could cause. One was taken in 1533; it was ordered by the council of the city of Augsburg and concerned *batzens*, the popular 4-*kreuzers* coins mentioned above, 15 of which were reckoned to be the equivalent of a gulden. *Batzens* minted by 20 different Upper-German authorities were tested (Thomann von Hagelstein 1692: 72). The other assay was the one carried out in Nuremberg in 1551 in preparation of Charles V's project to create a common Imperial currency. The mint masters gathered for the assay tested a huge variety of coins; those of interest here are the 10 types of *mariengroschens* (20 of which were a gulden of account) that had been issued by princes and towns in North-Western Germany (Volckart 2017a: 335). Table 1 shows the summary statistics of the two assays.

Despite all *batzens* and all *mariengroschens* having the same official values, the designs differed widely so that confusing the issue of one authority with that of another was hardly possible. Hence, merchants engaging in the trade in coinage could not exploit Gresham's Law in the straightforward way that the parallel circulation of the outwardly fairly similar looking English and Burgundian nobles had allowed in about 1400. However, the use of the gulden as a common unit of account still made this possible. Consumers were prepared to accept that 15 *batzens*, regardless who had minted them, were one gulden of account and thus interchangeable. *Mariengroschens* varied more widely than *batzens*, but the principle was still the same: Merchants could cull a gulden's worth of the types with the highest silver content and sell them to a mint that produced lighter equivalents. If the difference between the bullion contents was large enough to cover the costs of transporting, melting and reminting the coins they would receive more than a gulden's worth of the lighter variant, making a profit in this way. They could do the same even if no coins of the same official value were concerned. For example, from a merchant's

Table 1 Pure silver content of batzens and mariengroschens, 1533 and 1551 in grams. Sources: Thomann von Hagelstein (1692: 72); Volckart (2017a: 335)

	Batzens 1533	*Mariengroschens* 1551
N	20	10
Minimum	1.551	0.744
Maximum	1.668	0.913
Average	1.601	0.825
Standard deviation	0.032	0.058

perspective selling a gulden's worth of *batzens* to a mint that produced *mariengroschens* always made sense if the *mariengroschens* could be used in a place where *batzens* were minted. That this happened is attested e.g., from Augsburg in the late 1540s (Municipal Archive (Stadtarchiv) Augsburg, Bestand: Reichsstadt, Strafamt No. 96.1 (Strafbuch 1543–1553), fol. 156 verso. For more examples see Volckart 2017b: 758 ff.). In fact, the trade in coinage was so common that many mints would have been forced to close without it. According to Schüttenhelm's (1984: 165) estimate, more than half of the bullion used by South German mints in the first half of the sixteenth century was not bought from mines or merchants but consisted of melted foreign coins. Given the use of a common unit of account, reminting such coins amounted to a policy that for practical purposes was no better than a debasement.

The consequence of all these varieties of the trade in coinage was that in the place of origin of the heavier coins, those with a proportionally lower content of bullion would appear in circulation. This is what happened, for example, in England after Philip the Bold had begun issuing his imitation nobles. Only four years later, in 1392, the English government legislated against

> gold money of the coinage of Flanders and Brabant [. . .] as by report of a great number of credible persons it is newly come to the King's ears that natives and aliens are bringing into the realm divers such gold money much resembling English money but of less weight and value therewith buying and selling as if it were of the English coinage, which it is not, and making other payments to the deception and damage of the King and people. (Maxwell Lyte 1925: 110)

As a result of such developments, the supply of specie to the mints producing heavier coins dried up, and with it the revenues from the seignorage. More serious, from an economic point of view, was the effect on transaction costs. The more different types of coins circulated at par despite their intrinsic values being different, the less sure could the public be that the money they handled matched their expectations. In consequence, more consumers would feel the need to check the weight and bullion content of their coins, and if they were unable to do that, would be inclined to reject money or not to engage in transactions that would have been mutually advantageous if money had been uniform and familiar. The trade in coinage thus did not only harm the revenues of authorities. It impeded the functioning of the market with all consequences this had for the division of labor, specialization, and factor allocation. Short of creating a uniform common currency, the only solution was debasing the domestic coinage so far that its bullion content matched or undercut that of the foreign money that had entered the market. Chilosi and Volckart (2017: 119 ff.) have shown statistically that urban governments took this factor significantly more often into account when deciding about debasements than princes. Thus, while princes tended to debase for fiscal reasons, thereby disturbing market transactions, towns debased more readily in order to drive out underweight foreign money. In this way, they reduced insecurity and transaction costs and helped their markets to function.

Conclusion

The present chapter analyses how late medieval and early modern debasements were affected by the wider economic and social conditions of the time. It explains that debasements were possible because merchants selling bullion to a mint were able to exploit two conditions: Either an informational asymmetry existed between them and poorly informed consumers, who did not realize that the coins they accepted had been debased, or the merchants' social standing meant they were powerful enough to force even well-informed consumers to accept debased money.

The core hypothesis of the chapter is that in most of premodern Europe debasements were far "messier" measures than much of the literature, which focuses on conditions in parts of Western Europe, realizes. That high information costs allowed rulers debasing their coinage and merchants selling them bullion to "fool" poorly informed consumers was only one side of the medal. The same information costs not only prevented the effective implementation of monetary policies but often created conditions which allowed or even motivated debasements. First, high information costs prevented the effective monitoring of agents employed to carry out governmental decisions. Even where rulers had not farmed out their mints, officials therefore often enjoyed a remarkable degree of autonomy. As the decision of which type of coin to mint was usually left to them, they tended to focus on minting those coins where their profit was highest. Moreover, there is evidence that on occasion out-of-control mint officials debased the coinage without referring to the authority that employed them. Thus, the flood of debased small change that undermined the common currency of the Holy Roman Empire at the end of the sixteenth and the beginning of the seventeenth century was issued by mints operating without supervision by the rulers officially responsible for them.

Second, high costs of information impeded attempts by political authorities to enforce regulations. This affected regulations concerning the value of coins, too. Research has so far often argued that in order to prevent the export of gold, governments debasing silver were forced to debase their gold coins, too, and that they normally did this not by reducing the bullion content, but rather by increasing the nominal value of gold coins. In fact, such measures were effective only under very restrictive conditions. As a rule, premodern governments were incapable of enforcing the circulation of gold at its nominal par value. To be precise, this was possible only if the rate they set for their gold happened to conform to the market rate. The reason was that it was usually the market that determined the value of monetary units: Purchasing power, exchange rates with coins belonging to other currencies, and the rates at which other coins from the same currency circulated were all open to negotiation between consumers. In the context of the supply of gold and silver coins, this implied that few premodern governments were able successfully to manage bimetallic currencies. Rather, most issued two distinct currencies based on different metals and circulating in parallel.

The final issue analyzed is the effect of information costs on attempts to create closed areas where the domestic currency enjoyed a monopoly, with costs being usually so high that this proved impossible. Normally borders were porous to the flow of coins, the more so as by the late Middle Ages a regular trade in coinage had

developed. In part, this trade exploited the – usually intentional – similarity between coins minted by different rulers, buying those whose bullion content was higher and selling them to a mint that produced inferior copies. The ruler who issued these copies, strictly speaking, did not debase his coinage as long as the copies were new units within the structure of his own currency, but his measure still had the same effect as a debasement, allowing him to increase his mint output and seignorage. Instances of this kind therefore show how thin the line between debasement and counterfeiting was. Occasionally, the trade in coinage also exploited the use of the same unit of account in different currencies. The effect was the same, with inferior coins replacing the better ones and allowing the authority issuing them to increase its seignorage. Also, in both cases the increasing circulation of "bad" foreign coins that drove "good" domestic ones out of circulation created strong incentives for the authority suffering from this influx to debase its own money, too. Debasements were therefore by no means exclusively fiscally motivated; they could just as well be defensive.

In sum, a debasement was anything but the orderly process envisaged by much of the prior research, where a ruler decided to debase his coinage, set the new mint price and patiently waited for members of the public to voluntarily supply his mint with bullion. Rather, debasements could be measures autonomously taken by poorly supervised officials. Also, mints acquired bullion in any way they could. In particular in cases of fiscally motivated debasements, confiscated gold or silver and re-minted foreign subsidies became important sources of supply. Generally, coins minted by other authorities served as raw material if they looked similar but had a higher intrinsic value or if the public used them in conjunction with the same unit of account. Moreover, as a rule measures on which specific types of debasements – e.g., increasing a coin's legal value – depended proved impossible to enforce. Still, as the examples of the Dauphin in the early fifteenth and Frederick II of Prussia in the mid-sixteenth century show, where a ruler aimed at increasing his revenues, debasing the coinage might be a hugely successful way to achieve this end.

Cross-References

- ▶ Gresham's Law
- ▶ Monetary System of the "Ancient Régime" (Third to Eighteenth Centuries)
- ▶ Money, Trade, and Payments in Preindustrial Europe

References

Primary Sources

Anon. (Egid Valentin Felix von Borié) (1761) Staats-Betrachtungen über gegenwärtigen Preußischen Krieg in Teutschland, in wiefern solcher das allgemeine Europäische, vornehmlich aber das besondere Teutsche Interesse betrift: mit untermischten Völker-rechtlichen Bemerkungen. Kaliwoda, Vienna

Arbusow L (ed) (1910) Akten und Recesse der livländischen Ständetage, vol 3. Deubner, Riga

Euling K (ed) (1891) Chronik des Johan Oldecop (Bibliothek des Litterarischen Vereins in Stuttgart CXC). Litterarischer Verein in Stuttgart, Tübingen
Fachs M (1678) Probier-Büchlein. Christian Scholvien, Leipzig
Helfferich KT (ed) (1900) Ludwig Bamberger: Ausgewählte Reden und Aufsätze über Geld- und Bankwesen. Guttentag, Berlin
Klempin R (1859) Diplomatische Beiträge zur Geschichte Pommerns aus der Zeit Bogislafs X. A. Bath, Berlin
Leeb J (1999) Der Kurfürstentag zu Frankfurt 1558 und der Reichstag zu Augsburg 1559 (Historische Kommission bei der Bayerischen Akademie der Wissenschaften (ed) Deutsche Reichstagsakten: Reichsversammlungen 1556–1662). Vandenhoeck & Ruprecht, Göttingen
Maxwell Lyte HC (ed) (1925) Calendar of close rolls, Richard II, vol 5. His Majesty's Stationery Office, London
Meckelburg FA (ed) (1865) Die Königsberger Chroniken aus der Zeit des Herzogs Albrecht; darin die Chronik des Johannes Freiberg. W. Koch, Königsberg
Meyer S (1912) Die Beler-Platnersche Chronik. Altpreußische Monatsschrift 49:343–415, 593–663
Thomann von Hagelstein D (1692) Der Röm. Kayserl. Majestät Deß Heil. Röm. Reichs Geist- und Weltlicher Chur-Fürsten, Fürsten und Ständen Acta Publica Monetaria Schrifftliche Handlungen, Constitutiones, Bedencken und Vorschläge, Die Anordnung, Reformation und Verbesserung Deß Müntz-Wesens in Teutschland betreffende. Kroniger & Göbels Erben, Augsburg
Volckart O (ed) (2017a) Eine Währung für das Reich: Die Akten der Münztage zu Speyer 1549 und 1557 (Historische Kommission bei der Bayerischen Akademie der Wissenschaften (ed) Deutsche Handelsakten des Mittelalters und der Neuzeit XXIII). Steiner, Stuttgart
von der Ropp GF (ed) (1890) Hanserecesse von 1431–1476, vol 6 (2. Abt.). Duncker & Humblot, Leipzig

Secondary Sources

Arnold P (1980) Die sächsische Talerwährung von 1500 bis 1763. Schweizerische Numismatische Rundschau 59:50–94
Behringer W (2005) "Von der Gutenberg-Galaxis zur Taxis-Galaxis". Die Kommunikationsrevolution: Ein Konzept zum besseren Verständnis der Frühen Neuzeit. Historische Zeitschrift Beihefte 41:39–54
Blanchet J-A, Dieudonné A (1916) Manuel de numismatique Française, vol 2. Picard, Paris
Buringh E, van Zanden JL (2009) Charting the "rise of the west": manuscripts and printed books in Europe, a long-term perspective from the sixth through eighteenth centuries. J Econ Hist 69:409–445
Chilosi D, Volckart O (2011) Money, states and empire: financial integration and institutional change in Central Europe, 1400–1520. J Econ Hist 71(3):762–791
Chilosi D, Volckart O (2017) Good or bad money? Debasements, society and the state in the late middle ages. In: Allen M (ed) Money and its use in medieval Europe three decades on: essays in honour of professor Peter Spufford. The Royal Numismatic Society, London, pp 111–128
Daubney A (2009) The circulation and prohibition of Venetian Soldini in late medieval England. Br Numis J 79:186–198
Denecke D (1990) Straße und Weg als Lebensraum und Vermittler zwischen entfernten Orten. In: Herrmann B (ed) Mensch und Umwelt im Mittelalter. Fischer, Frankfurt, pp 207–223
Dipper C (1991) Deutsche Geschichte 1648–1789 (Neue historische Bibliothek 253). Suhrkamp, Frankfurt
Engel E (1855) Die Geldprägungen nach dem Leipziger Münzfuße, dem Conventionsfuße und dem 14-Thalerfuße in dem Churfürstenthume und dem Königreiche Sachsen. Zeitschrift des Statistischen Büreaus des Königl. Sächs. Ministeriums des Innern 4:49–64

Engelsing R (1973) Analphabetentum und Lektüre: Zur Sozialgeschichte des Lesens in Deutschland zwischen feudaler und industrieller Gesellschaft. Metzler, Stuttgart
Glassman D, Redish A (1985) New estimates of the money stock in France, 1493–1680. J Econ Hist 45(1):31–46
Gould JD (1970) The great debasement: currency and the economy in Mid-Tudor England. Clarendon Press, Oxford
Groebner V (1993) Ökonomie ohne Haus: Zum Wirtschaften armer Leute in Nürnberg am Ende des 15. Jahrhunderts. Vandenhoeck & Ruprecht, Göttingen
Gumowski M (1948) Fałszerstwa monetarne Fryderyka II. Instytut Zachodni, Poznań
Hettling M (2015) Bürger/Bürgerlichkeit. In: Thoma H (ed) Handbuch Europäische Aufklärung: Begriffe – Konzepte – Wirkung. J. B. Metzler, Stuttgart/Weimar, pp 123–131
Hoensch JK (1973) Friedrichs II. Währungsmanipulationen im Siebenjährigen Krieg und ihre Auswirkungen auf die polnische Münzreform von 1765/66. Jahrbuch für die Geschichte Mittel- und Ostdeutschlands 22:110–173
Houston R (2014) Literacy in early modern Europe: culture and education 1500–1800. Taylor & Francis, London/New York
Jesse W (1928) Der Wendische Münzverein (Quellen und Darstellungen zur hansischen Geschichte NF 6). Hansischer Geschichtsverein, Lübeck
Koser R (1900) Die preußischen Finanzen im siebenjährigen Kriege. Forschungen zur brandenburgischen und preußischen Geschichte 13:153–217, 329–375
Laube A (1976) Studien über den erzgebirgischen Silberbergbau von 1470 bis 1546: Seine Geschichte, seine Produktionsverhältnisse, seine Bedeutung für die gesellschaftlichen Veränderungen und Klassenkämpfe in Sachsen am Beginn der Übergangsepoche vom Feudalismus zum Kapitalismus (Forschungen zur Mittelalterlichen Geschichte 22), 2nd edn. Akademie-Verlag, Berlin
Metcalf DM, Merrick JM (1967) Studies in the composition of early medieval coins. Numis Chron 7:167–181
Miskimin HA (1985/89) The enforcement of Gresham's law. In: Harry Miskimin IX (ed) Cash, credit and crisis in Europe, 1300–1600. Variorum Reprints, London, pp 147–161
Münch P (1996) Lebensformen in der Frühen Neuzeit 1500 bis 1800. Ullstein, Frankfurt/Berlin
Munro JH (1972) Wool, cloth, and gold: the struggle for bullion in Anglo-Burgundian trade, 1340–1478. University of Toronto Press, Brussels/Toronto
Munro JH (2015) The technology and economics of coinage debasements in medieval and early modern Europe: with special reference to the Low Countries and England. In: Munro JH (ed) Money in the pre-industrial world: bullion, debasements and coin substitutes. Routledge, Abingdon, pp 15–32
Prestwich M (1979) Early fourteenth-century exchange rates. Econ Hist Rev XXXII(4):470–482
Redish A (2000) Bimetallism: an economic and historical analysis. Cambridge University Press, Cambridge/New York
Rolnick AJ, Velde FR, Weber WE (1996) The debasement puzzle: an essay on medieval monetary history. J Econ Hist 56(4):789–808
Rössner PR (2012) Deflation – Devaluation – Rebellion: Geld im Zeitalter der Reformation (Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, Beiheft 219). Steiner, Stuttgart
Rössner PR (2014) Monetary instability, lack of integration, and the curse of a commodity money standard: the German lands, c.1400–1900 A.D. Credit Cap Mark – Kredit und Kapital 47(2):297–340. https://doi.org/10.3790/ccm.47.2.297
Sargent TJ, Velde FR (2002) The big problem of small change (The Princeton economic history of the world). Princeton University Press, Princeton/Oxford
Schalk C (1881) Der Wiener Münzverkehr im 16. Jahrhundert. Numismatische Zeitschrift 13: 243–329
Schneider K (1995) Heckenmünze. In: North M (ed) Von Aktie bis Zoll: Ein historisches Lexikon des Geldes. Beck, Munich, pp 156–157
Schüttenhelm J (1984) Zur Münzprägung und Silberversorgung süddeutscher Münzstätten im frühen 16. Jahrhundert. In: Kroker W (ed) Der Anschnitt: Zeischrift für Kunst und Kultur im

Bergbau. Beiheft 2: Montanwirtschaft Mitteleuropas vom 12. bis 17. Jahrhundert. Stand, Wege und Aufgaben der Forschung. Deutsches Bergbau-Museum, Bochum, pp 159–169
Scott HM (2015) The Oxford handbook of early modern European history, 1350–1750. Vol. 1: Peoples and place. Oxford University Press, Oxford
Sprenger B (2002) Das Geld der Deutschen: Geldgeschichte Deutschlands von den Anfängen bis zur Gegenwart, 3rd edn. Schöningh, Paderborn/Munich/Vienna/Zürich
Spufford P (1963) Continental coins in late medieval England. Br Numis J 32:127–139
Spufford P (1988) Mint organisation in late medieval Europe. In: Spufford P, Mayhew NJ (eds) Later medieval mints: organisation, administration and techniques. The eighth Oxford symposium on coinage and monetary history. B.A.R. Publications, Oxford, pp 7–29
Sussman N (1993) Debasements, royal revenues, and inflation in France during the Hundred Years' War, 1415–1422. J Econ Hist 53(1):44–70
van Zanden JL, Prak M (2006) Towards an economic interpretation of citizenship: the Dutch Republic between medieval communes and modern nation-states. Eur Rev Econ Hist 10(2):111–145
Volckart O (1996) Die Münzpolitik im Deutschordensland und Herzogtum Preußen von 1370 bis 1550. Harrassowitz, Wiesbaden
Volckart O (2017b) Power politics and princely debts: why Germany's common currency failed, 1549–1556. Econ Hist Rev 70(3):758–778
von Schrötter F (1910) Das Preussische Münzwesen im 18. Jahrhundert: Münzgeschichtlicher Teil. Vol. 3: Das Geld des siebenjährigen Krieges und die Münzreform nach dem Frieden: 1755–1765. Parey, Berlin
Wuttke R (1897) Die Probationsregister des Obersächsischen Kreises. Numismatische Zeitschrift 29:237–302

Gresham's Law

8

George Selgin

Contents

G. Selgin (✉)
Center for Monetary and Financial Alternatives, The Cato Institute, Washington, DC, USA
e-mail: gselgin@cato.org

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_9

Abstract

"Gresham's Law" – the tendency for bad money to drive good money out of circulation – is one of the most well-known propositions of economics. Although it is far from universally valid, provided it is appropriately qualified, by noting the particular conditions that can make it operate, it helps to account for many important episodes in the history of money. This chapter reviews early statements of Gresham's Law and the circumstances that informed them. It then considers the determinants of currency selection in different market settings, starting with the case of unhindered market, so as to arrive at a more precise understanding of the factors that can cause Gresham's Law to operate. Next it shows how Gresham's Law can account for some past episodes involving irredeemable paper money and bimetallism, and why it cannot predict the consequences of private coinage, or explain the replacement of coins by *redeemable* paper money. The chapter ends by critically assessing Arthur Rolnick and Warren Weber's claim that Gresham's Law, even in its more carefully stated versions, is a "fallacy."

Keywords

Gresham's Law · Debasement · Bimetallism · Legal tender

Introduction

Among the better-known propositions of economics, "Gresham's Law" is often put bluntly as "bad money drives good money out of circulation." Stated so it is, like many sweeping assertion, as likely to be at odds with experience as consistent with it. Indeed, as Robert Mundell (1998) has observed, were there nothing more than this to Gresham's Law, instead of having a "claim to our attention as a serious proposition of economics," it would be just one more "completely false generalization."

Yet Gresham's Law does deserve our attention, for provided it is stated carefully, with due attention to the special – but historically far from uncommon – circumstances to which it applies, it can account for, and shed light upon many important episodes in the history of money.

In this chapter I plan, first of all, to briefly review early statements of Gresham's Law in order to consider the circumstances that informed them. I'll then review the determinants of currency selection in different market settings, starting with the case of unhindered market, so as to arrive at a more precise understanding of the factors that may cause Gresham's Law to operate. Next I'll discuss how the law can be properly appealed to shed light on certain episodes involving irredeemable paper money and bimetallism, and how it has been *improperly* invoked to describe the likely outcome of private coinage and to explain the supplanting of coins by *redeemable* paper money. Finally I'll respond to Arthur Rolnick and Warren Weber's (1986) claim that Gresham's Law, even in its more carefully stated versions, is a "fallacy."

Early Statements of Gresham's Law

Sir Thomas Gresham

To gain a proper understanding of Gresham's Law, it is essential to consider just what its original exponents meant by it.

The expression "Gresham's Law" itself dates back only to 1858, when British economist Henry Dunning Macleod (1858, pp. 476–478) decided to name the frequent tendency for bad money to drive good money out of circulation after Sir Thomas Gresham (1519–1579), the English merchant and financier who founded the Royal Exchange and served as advisor to four British monarchs, beginning with Henry VIII and ending with Elizabeth I.

Writing to Queen Elizabeth on the occasion of her accession, Gresham observed "that good and bad coin cannot circulate together," and that this fact accounted for the "unexampled state of badness" to which England's coinage had fallen in consequence of the "Great Debasements" of Henry VIII and Edward VI. Those debasements reduced the silver content of English silver coins to a small fraction of the standard that had prevailed during the reign of Henry VII; and it was thanks to them, Gresham told the Queen, that "all your ffine goold was convayd ought of this your realm."

Copernicus and Oresme

In his brief remarks, Gresham, rather than explicitly appealing to a general principle, refers to a single episode only. Nicole (or Nicolas) Oresme and Nicolaus Copernicus had, as Macleod himself (1896, p. 38) eventually discovered, "fully explained the matter [of Gresham's Law] 160 and 32 years respectively previous to Gresham."

In his 1526 tract *De monetae cudendae ratio* ("On the principle of coining money"), Copernicus, in advising Poland's King Sigismund I of the steps needed to rehabilitate Prussia's debased currency, observed that previous debasements had driven away the better coins "so to speak, by main force," and that to restore the old standard the debased coins would themselves have to be cried down. Otherwise, he explained, introducing new, full-bodied coins would prove futile, for then the better coins would either be diminished themselves or would disappear.

Copernicus in turn is supposed to have been anticipated by Oresme, who (according to one authority's summary) had written over a century before that

> whenever the public currency was altered or tampered with in such a way as to bring into circulation two moneys, bearing the same designation but in reality having two different values, the money of lower value inevitably drove the money of higher value out of circulation. For the merchants found it to their advantage either to melt down the pieces of money that contained the higher amount of metal and to sell the bullion by weight or else to export the high weight coins to other lands. (Balch 1908, p. 23)

Although the main passage upon which the above summary rests has since been shown to have been written, not by Oresme himself, but by his French translator (Bridrey 1906, pp. 264–296), Oresme still came closer to explicitly stating Gresham's law than Gresham himself did. In one indisputably authentic passage of his treatise he notes, for example, how "in spite of all vigilance and prohibition that may be taken," mutations of the currency end up inspiring merchants and others to convey precious metal to "the places where they know these have a greater value" (quoted in Balch 1908, p. 23). It is, surely, but a small step from that observation to the conclusion that full-bodied coins will disappear from circulation, leaving their reduced substitutes behind.

Aristophanes

Finally we come to the earliest known reference to what would come to be known as Gresham's Law. In his comedy *The Frogs*, Aristophanes bemoans the fact that "the full-bodied coins that are the pride of Athens are never used while the mean brass coins pass hand to hand." He then attributes the prevalence of bad politicians to similar forces:

> Oftentimes have we reflected on a similar abuse
> In the choice of men for office, and of coins for common use;
> For your old and standard pieces, valued and approved and tried
> Here among the Grecian nations, and in all the world beside,
> Recognized in every realm for trusty stamp and pure assay,
> Are rejected and abandoned for the trash of yesterday;
> For vile, adulterate issue, drossy, counterfeit and base,
> Which the traffic of the city passes current in their place.
> (Frere's translation, as quoted in Balch 1908, p. 18)

The Context of Gresham's Law

Each of these early references to what we now call Gresham's Law took place against the background of deliberate reductions, through debasement or otherwise, of the metallic, though not the declared, value of official coins. Aristophanes' play, for example, was written in 405 B.C., or just before the end of the Peloponnesian War. When that war began, Athenian silver coins had, thanks to their high and consistent quality, become the standard currency of the Ancient world. However, when, in 413 B.C., the Spartans seized Decelia, depriving Athens of access to its Laurion silver mines, the Athenian government was eventually compelled to strike copper ("brass") coins for the first time, in 406 B.C. These "were made legally, for the time being, equivalent to their counterparts in silver" (Head 1888, p. xxviii). Oresme's *De Moneta* was, according to Charles Johnson (1956, p. x), similarly "provoked by the successive debasements of the coinage by Philip VI and John II": between 1351 and 1360, Balch (1908, pp. 22–23) tells us, the metallic content of the

livre tournois changed "no less than seventy-one times." As we have seen, Copernicus was also concerned with problems connected with debasement, and the difficulties involved in restoring "good" coin. Gresham's remarks, finally, were inspired by Tudor England's Great Debasement.

In each of these episodes, "Gresham's Law" referred to the tendency of coins possessing a relatively low metallic content ("bad" money) to be employed in routine payments, while coins made of the same metal, but to a higher standard ("good money") disappeared into hoards, or melting pots, or were themselves reduced through clipping or sweating to an intrinsic value no greater than that possessed by their inferior counterparts.

Currency Selection in Hindered and Unhindered Markets

An Unhindered Currency Market

To gain a more complete understanding of the conditions needed to put Gresham's Law in motion, it is helpful to first consider the case of a free domestic currency market, meaning one in which there is neither government interference with the free selection and pricing of alternative currencies nor imperfect information regarding the different currencies' qualities, as well as a world market for precious metals. It is easy to show that, in such an environment, Gresham's Law does *not* apply.

Suppose, then, that there are no legal tender laws to compel the acceptance of coins by tale, that is, according to their officially declared value. Suppose as well that everyone possesses complete information regarding coins' metallic composition. Finally, suppose that it is no more costly for coins to pass by weight, that is, according to their metallic values, than by tale. In such a setting, it should make no difference to merchants and their customers which coins they accept or proffer in exchange; coins would command values reflecting their metallic worth, and there is no reason why some would be preferred to others. Instead, better and worse coins of any given nominal value might co-circulate as "parallel" currencies. Goods might be priced in terms of both sorts of money, or one sort alone might serve as the medium of account, with the other trading at a discount (if "bad") or a premium (if "good").

Costly Nonpar Exchange

Evidently, to put Gresham's Law in motion we must relax one or more of the assumptions made above. Let's start by allowing it to be far more convenient for merchants to set prices and keep accounts in terms of a single medium of account only, and also for coins to pass by tale than by weight. Suppose as well that both "good" and "bad" coins are made available, where both have a face value of "one dollar," but "good" coins contain substantially more precious metal. Which coins will be favored?

The intuitive answer is that market forces will favor the coins that are merchants' chosen medium of account, and that therefore trade at par. But which coins will play

this role? The answer is not clear: under laissez-faire, sellers might post prices in terms of either sort of coin. If they post prices in terms of "good" coins, those will trade at par, and "bad" coins, if they circulate at all, will do so at a discount, reflecting their lower metal content as well as the extra transactions costs merchants bear in dealing with them. If, on the other hand, sellers post prices in terms of "bad" coins, those will circulate at par, while "good" coins, if they circulate, will command a premium, albeit one that is less than proportional to their superior metallic worth. If nonpar exchange is sufficiently costly to merchants, nonpar money will not circulate. In that case, par money may be said to drive nonpar money out of circulation.

This last-mentioned outcome I dub "Rolnick and Weber's Law," after Rolnick and Weber (1986), who propose it as a possible alternative to Gresham's Law, which they hold to be fallacious. I address their dismissal of Gresham's law later in this essay. For now it will suffice to note that Rolnick and Weber's Law is distinct from Gresham's, because it suggests that either "good" or "bad" coins might disappear from circulation, depending on the type that trades at par, or, put differently, the type that serves as the "medium of account."

The case just described can be further understood as one in which buyers and sellers play a monetary selection game. Suppose, for example, that sellers and buyers all expect a payoff of 4 from a particular exchange conducted at par. In the case of nonpar exchange, the net payoff is 3. (Costs of nonpar exchange are assumed here to be borne equally by seller and buyer.) Sellers must choose the medium of account in which to post prices, while buyers must choose which medium of exchange to offer. Clearly, if buyers are expected to offer "good" money, it will make sense for sellers to post prices accordingly. Alternatively, if sellers expect buyers to offer "bad" coins, they will find it convenient to price their wares in terms of "bad" coins. Buyers will in turn prefer to employ whatever money sellers treat as the medium of account. If "good" and "bad" coins are equally likely to be chosen as the medium of exchange and account, buyers and sellers may be said to be playing the pure coordination game shown in Table 1.

An Asymmetrical Monetary Selection Game

To arrive at the Gresham's Law outcome, we must move beyond the pure coordination game to one in which circumstances "nudge" players' payoffs so as to favor a

Table 1 A pure-coordination monetary selection game

	Buyer	
Seller	Offer "Good" MoE	Offer "Bad" MoE
Employ "Good" MoA	4,4[a]	3,3
Employ "Bad" MofA	3,3	4,4[a]

Notes: [a]Denotes equilibrium. *MofA* = Medium of account, *MoE* = Medium of exchange. Left-hand payoff is seller's

particular equilibrium. It is not difficult to imagine such circumstances. Suppose, for example, that one kind of money is more prominent than the other. Sellers may then expect to minimize transactions costs by posting prices in terms of that money (White 1984, pp. 177–178). Returning to the previous example, if the nominal outstanding quantity of "good" coins is known to be twice that of "bad"coins, sellers are likely to post prices in terms of "good" coins, openly signaling a preference for them. Buyers and sellers could then be expected to converge upon the "good" money equilibrium. Alternatively, suppose that, although they contain more metal, "good" coins are considered cumbersomely large, and therefore awkward to carry. In that case, sellers may anticipate having more customers arrive with "bad" coins and will post prices accordingly. In these instances the monetary selection game has a unique equilibrium. However, because the equilibrium in these cases may still favor either "good" or "bad" money, the conditions considered so far still do not suffice to give effect to Gresham's Law.

Legal Tender Laws

To account for a systematic tendency for "bad" money to displace "good" money, it is necessary to appeal to other currency market imperfections. According to Richard Dutu (2004), two distinct sorts of market imperfections served the purpose in premodern currency markets. These were (1) legal tender laws and (2) imperfect information.

Legal tender laws are here understood simply as laws that compel people to receive particular monies at their officially declared ("face") values. As Macleod (1899, p. 59) explains,

> Mediaeval princes conceived that it was part of their inalienable Divine Right to alter the weight and name, and debase the purity of their Coin as much as they pleased, and to compel their subjects to receive the diminished and degraded and debased Coin at the same value as good, full-weighted Coin. This was termed *Morbus numericus*.

In her classic study *Legal Tender*, Sophonisba Preston Breckinridge (1903, p. 17) likewise reminds us, with specific reference to medieval England, that

> freedom of contract and of commerce did not exist in the sense in which we understand these terms. Government monopolized the function of coinage and enforced its monopoly by imposing penalties for the offense of refusing the king's coins at the values set upon them by the king, and by prohibiting the currency of coins whose circulation would interfere with the coins issued from the king's mints.

Legal tender laws might, of course, be more or less strictly enforced, and it was only relatively strict laws that served to give effect to Gresham's Law. Yet such laws were not uncommon in premodern times; and they have occasionally been resorted to more recently.

Although they aspire to enforce the equal treatment of intrinsically distinct monies, legal tender laws influence currency selection, not (as some writings on

Gresham's Law suggest) by actually establishing an operational, fixed exchange rate between distinct currencies, but by making it costly for persons to openly assign distinct values to them, A legal-tender law might, for example, punish sellers who refuse or place a discount on "bad" money, while rewarding buyers who report such discrimination.

During the reign of Henry VII, for example, Parliament declared (19 Henry VII, c. 5) that all coins bearing the King's stamp "should go and be current in payment through all this his realm for the sum that they were coined for," and that all silver pence "having the print of the king's coin, shall have course and be current for payment, as well to him in all his receipts, as to all his receivers, and to all other lords spiritual and temporal and their receivers, and to all other within this his realm, without any manner refusal or contradiction. . . Any person refusing to take coins in payment for the values aforesaid, to be liable to punishment at the decision of a justice." Penalties for disobeying the law included fines and imprisonment as well as the forfeiting of any unlawfully exchanged sums.

From Pure Coordination to Prisoner's Dilemma

In the presence of legal tender laws, a seller posting prices in terms of "good" coins would risk being penalized for refusing to accept "bad" ones at their face value. Such laws can therefore put Gresham's Law in motion by turning the pure monetary coordination game described previously into a Prisoner's Dilemma game.

Suppose, for example, that instead of receiving the payoffs shown in the above pure-coordination game, sellers who set prices in terms of "good" money, but are offered "bad" coins, receive a reduced net payoff of 1, because they must either accept such coins at their face value or risk being penalized. Buyers who offer "bad" money in turn earn an increased payoff of 5. If, on the other hand, buyers offer "good" money to sellers who set prices in terms of "bad" money, the payoffs are reversed. The exchange game now has a unique, "bad" coin equilibrium. Observe that this outcome does *not* depend on laws explicitly favoring "bad" money or making "good" money illegal.

Gresham's Law can prevail, moreover, notwithstanding market considerations that might favor the use of "good" money were it is possible for merchants to openly discriminate in its favor. Suppose, for example, that "bad" money is more bulky and less convenient than "good" money, so that its employment at par yields a lower payoff, say, 3, than the par employment of "good" money. "Bad" money will still prevail, making the game a genuine Prisoner's Dilemma, as shown in Table 2. This example also makes clear that legal-tender laws do not merely amplify market-based costs of nonpar change. Were that the case, legal-tender laws would as often as not favor "good" money.

What becomes of "good" coins in the presence of strict legal-tender laws? If they are not to be traded at par for goods priced in terms of "bad" money, "good" coins may either be hoarded (in anticipation, perhaps, of a future change in the law) or

Table 2 The Gresham's Law game

	Buyer	
Seller	Offer "Good" MoE	Offer "Bad" MoE
Employ "Good" MoA	4,4	1,5
Employ "Bad" MofA	5,1	3,3[a]

Notes: Same as Table 1

exported: even strict legal-tender laws are difficult to enforce outside of domestic political boundaries (Miskimin 1989, p. 149). Alternatively holders of "good" coins may be encouraged to surrender them to the mint in return for a nominal profit to the mint: in order to supply "bad" coins, and thereby gain seignorage, a mint must secure a ready supply of precious metal, which it can do by sharing some of its nominal coining profits with those who supply it with needed raw material (Hicks 1969, pp. 90–91; Breckinridge 1903, pp. 34–36; compare Rolnick et al. 1996, who claim that "debasements provide no additional inducements to bring coins to the mint.").

Here again, Tudor England supplies a convenient illustration. As John Chown (1994, pp. 46–47) explains, dramatic increases in the mint equivalent of silver (the face value of silver coins made from a troy pound of silver) during the Great Debasement were accompanied by corresponding increases in the mint price of silver (the face value of coins offered by the mint in return for a troy pound of silver). Just prior to April 1544, for example,

> a citizen bringing a pound of silver to the mint would receive 584 good quality pennies. After that date he would have received 629 slightly lighter ones. If he assumed that the quality of silver was unchanged there was little real change in the deal he was being offered. In fact the silver content had been reduced from 0.925 to 0.75 fine. The mint equivalent was 768 pence. 139 pence, about 22 per cent, had been kept by the mint. This was nearly all profit, but a modest skim compared to what was to come. (ibid.)

Besides allowing their citizens to profit, if only in nominal terms, by bringing precious metal to the mint, medieval governments generally prohibited them from either retaining or dealing in bullion, meaning precious metal considered as a commodity. According to John Munro (2012, p. 317), "almost everywhere it was illegal to trade or to make transactions in bullion. For the law in most medieval countries or principalities stipulated that all precious metals deemed to be bullion (billon) – excluding metals for licensed goldsmiths – had to be surrendered to the prince's mint for coinage."

Imperfect Information

Richard Dutu (2004, p. 557, see also Sargent and Smith 1997, and Velde et al. 1999) has proposed, as "an alternative to the legal tender explanation" of Gresham's Law, the possibility that "private information held by money experts played a crucial role in activating the driving out of the good coins":

> The great variety of coins, the imperfect coinage technique, frequent mutations, wear, the poor communication network, all these factors made it difficult for one type of coin to have a single price and a stable intrinsic content. Our thesis is that this imperfect information on coins was the source of large profits for people who invested in gathering knowledge on money – the moneychangers – and that their activity most of the time led to the driving out of the undervalued currencies.

Moneychangers engaged in both "billonage" – "comparing the intrinsic content of two supposedly identical coins [and] keeping the heavier ones" – and arbitrage – disposing of the heavier coins in markets where they command more than at home (ibid.). "A few people holding some private information" could therefore suffice to see to it that undervalued currencies were "driven out, exported, or melted down" (ibid. p. 569). Copernicus himself recognized this in observing how

> those specialized in precious metals. . .sort out ancient coins, melt them again and then sell them to the silverer always receiving from inexperienced persons more money with the same amount of money. When older coins have almost disappeared, they choose the best from the rest and just leave the worst currencies. (quoted in ibid, p. 563)

Thus the inability of ordinary persons to afford to trouble themselves with determining the true metallic worth of different equally rated coins, particularly when relatively small amounts are involved, combined with the fact that others can profit by specializing in making just such a determination, may lead to be systematic disappearance of better coins even were strict legal tender laws are lacking.

Nor need the "specialists" responsible for this outcome be ones who deal in money and bullion exclusively: ordinary merchants might also find it worthwhile to start weighing or otherwise assessing coins, "putting aside those worth the most, melting them down and getting profit from them" (Dutu 2004, p. 564). But note that, whether merchants or other specialists do the culling, it must be the case both that coins pass by tale *and* that lighter coins serve as the de facto medium of account. Otherwise moneychangers and merchants would suffer losses to the extent that they received bad coins at par. To take advantage of the general public's imperfect information, they must price their merchandise, whether it consists of other coins or of goods, as if they expected to be paid mainly in "bad" money.

To offer yet another illustration from Tudor England, Li (2009, p. 11) observes that, although in theory anyone there might have calculated coins' intrinsic value, in practice

> the touchstone test – the common technology used to identify the fineness of coins in the early modern period – was accurate at best only to within two or three percentage points and required a great deal of specialized knowledge and instrument [sic]. This implies that for many consumers, the transaction costs of determining the intrinsic value of coins must have been prohibitive. Therefore, the general public likely handled coins by tale. On the other hand, merchants involved with large transactions especially in international trade and the government valued coins on their intrinsic value. (that is, by weight)

Munro (2012, pp. 317–318) likewise observes that in medieval times "even if it had been legal to make transactions in bullion, doing so would not have been economically feasible in terms of the required transaction costs: the cost of weighing the bullion, assaying it for fineness, and determining its market or exchange value." Legal sanctions against nonpar exchange and the transactions costs of valuing coins according to their bullion content thus tended to reinforce one another in driving "good" coin, first into the hands of expert money changers, thence into melting posts, and finally (as bullion) either to foreign markets or back to the domestic mint in return for a nominal profit.

Extensions of Gresham's Law

Although Gresham's Law originally referred mainly to circumstances in which people were confronted with both "good" and "bad" coins bearing identical face values but containing different quantities of the same precious metal, the law, as William Stanley Jevons observed (1875, p. 84), may also be applicable to "the relations of all kinds of money, in the same circulation. Gold compared with silver, or silver with copper, or paper compared with gold."

Irredeemable Paper Money

Just as a monetary authority might declare debased coins legally equivalent to full-bodied ones, so might it attempt to make irredeemable paper notes legally equivalent to coin. It seems only reasonable, then, to assume that Gresham's Law might in this case cause coins to disappear from circulation.

There is, however, an important difference, to wit: that there's no question of even less – sophisticated market participants being unable to appreciate the difference between paper money and coins, and of their tending for that reason to tender or receive either sort of money indiscriminately on that account. Strict legal tender laws are therefore likely to be of particular importance in causing "good" coins to be driven out of circulation, not by "bad" coins, but by irredeemable paper notes.

Continentals

To give an example, no sooner did the Continental Congress discover that its Continental bills, first authorized in 1775, had fallen to a discount relative to Spanish milled dollars, than it resolved that anyone who refused to receive them at par would be "deemed and treated as an enemy of his country, and be precluded from intercourse with its inhabitants." Some states adopted similar resolutions. Thus the Pennsylvania Council of Safety resolved, on December 27, 1776, that anyone who

> shall refuse to take Continental Currency in payment of any Debt or Contract whatsoever, or for any Goods, Merchandize or Commodity offered for sale, or shall ask a greater Price for any such Commodity in Continental Currency than in any other kind of money or specie. . .. shall for the first offense be considered a dangerous Member of Society, and forfeit the Goods offered for sale or Bargained for, or debt Contracted, to the person to whom the Goods were offered for Sale or by whom they were bargained for, or for whom such Debt is due, and shall moreover pay a fine of five pounds to the State. . .; and every person so offending, shall for the second offence be subject to the aforementioned penalties, and be banished from this State, to such a place and in such manner, as this Council shall direct. (State of Pennsylvania 1852, pp. 70–71)

Evidently it is not always true, as Arthur Rolnick and Warren Weber (1986, p. 193) claim, that placing a premium on "good" money "would not be in violation of . . . legal-tender laws."

According to Pelatiah Webster (1791, p. 129), resolutions like Pennsylvania's were "executed with a relentless severity," leading to the ruin of "Many thousand families of full and easy fortune." Yet as the value of Continentals became increasingly doubtful, still harsher penalties were imposed on those who discriminated against them. By December 1780, Webster says, the penalty for refusing a dollar of Continental currency was greater than that for stealing ten times that amount (ibid., p. 137). William Graham Sumner (1892, pp. 48–52) refers to many specific prosecutions of persons accused of violating the laws in question.

Their severity notwithstanding, such laws were of course unable to make Continentals command the same value as their precious-metal counterparts. But they did achieve the unintended, Gresham's Law outcome of driving specie out of open circulation. In fact, specie was seldom seen after its initial disbursement by newly arrived English and French troops, except among certain merchants who engaged in a regular, though clandestine, trade in it (Bezanson 1951, p. 320). Specie came out of hiding only after March 16, 1781, when Congress officially recognized its free-market value relative to paper money (ibid.). Although Charles Calomiris (1988, p. 697, n. 6) has called the subsequent re-emergence of specie, and concurrent disappearance of Continentals, "a clear contradiction of Gresham's law," it is hardly surprising that Gresham's Law ceased to operate once Continentals were no longer assigned official specie values enforceable by law.

Assignats

If the punishments meted-out by the Continental Congress were Draconian, those that France's National Convention employed on behalf of paper assignats were still more so. According to Andrew Dickson White (1933, pp. 42–43), the Convention first

> decreed any person selling gold or silver coin, or making any difference in any transaction between paper and specie, should be imprisoned in irons for 6 years: – that any one who refused to accept a payment in assignats, or accepted assignats at a discount, should pay a

> fine of 3,000 francs; and that any one committing this crime a second time should pay a fine of 6,000 francs and suffer imprisonment 20 years in irons. Later, on the 8th of September, 1793, the penalty for such offences was made death, with confiscation of the criminal's property, and a reward was offered to any person informing the authorities regarding any such criminal transaction. To reach the climax of ferocity, the Convention decreed, in May, 1794, that the death penalty should be inflicted on any person convicted of "having asked, before a bargain was concluded, in what money payment was to be made."

Nor were the government's threats empty. A dozen men were in fact sent to the guillotine for the crime of hoarding specie (Harris 1930, p. 183). Small wonder that, under these circumstances, paper money alone circulated openly!

Bimetallism

Although Robert Giffen (1891, pp. 304–305) has said that Thomas Gresham was "only responsible for the suggestion that bad coins . . . drive good ones of the same metal out of circulation," and that he made no mention of either "the analogous case of inconvertible paper" or bimetallism, we have seen that in his famous letter to Elizabeth Gresham clearly alludes to bimetallism in attributing the disappearance of *gold* coins to the debasement of silver. It was, nevertheless, only in modern times, when debasement had largely become a thing of the past, that Gresham's Law first gained prominence, no longer as an argument against princely abuses of the coinage prerogative, but as one against bimetallism.

When bimetallic legislation allows for both silver and gold coins representing a common monetary unit, the "mint equivalents" of the two metals (that is, the nominal value of coins struck from a given weight of either metal) imply a corresponding "mint ratio," meaning the ratio of the metals' official values. If coinage is gratuitous, or if coining charges are proportionally the same for both metals, the mint ratio is the same as the ratio of the metals' "mint prices," that is, the value of coins returned to those who surrender bullion to the mint for coining.

So long as the mint ratio coincides with the ratio of prices the metals command in the world bullion market, dealers in both metals will be willing to deliver them to the mint for coining, and full-bodied coins made of either metal can circulate simultaneously. If, however, the two ratios diverge substantially, one metal will be overvalued by the mint, while the other is undervalued by it. The undervalued metal will then cease to be coined, while extant coins made from it will tend to be treated as bullion rather than as money. If we then regard coins made from the overrated metal as "bad," and those made from the undervalued metal as "good," the tendency of de jure bimetallism to give way to de facto monometallism can be considered an instance of Gresham's Law.

The tendency in question will, however, be limited by the fact that coins of different metals are unlikely to be equally useful in different transactions. In particular, gold coins will generally be of larger denominations, so that they

cannot themselves satisfy the need for smaller change (cf. Sargent and Velde 2002). Consequently, even if gold is legally overvalued, and silver is no longer brought to the mint for coining, silver coins may not altogether disappear. Instead, they may circulate at a premium or (if that is too inconvenient) they may be clipped or sweated by private agents until their metallic values no longer exceeds their face value in gold, as happened in Britain during the eighteenth century. In the latter case, Gresham's Law may still be said to hold in the strict sense that *full-bodied* coins of the undervalued metal cease to circulate.

England's Switch to Gold

A famous instance of Gresham's Law's operating in a bimetallic setting is the one that led to Great Britain's inadvertent shift from a de facto silver standard to a de facto gold standard. Although the English pound sterling had, as its name suggests, traditionally been a silver unit, when gold guineas were introduced in 1663, they were officially rated at one pound sterling, or 20 (silver) shillings. Bimetallism was thus formally established. However, because the implied mint ratio undervalued silver, which was then in great demand in the Far East, instead of bringing silver to the mint merchants would exchange it for gold, for which they could obtain a greater nominal quantity of guineas. The inevitable result was that full-bodied silver coins, instead of remaining in circulation, were clipped or melted down. Although in 1717 guineas were rated at 21 shillings, at that price gold was still overvalued. Consequently although it did not officially embrace gold monometallism until 1816, England had stumbled into a gold standard, without realizing it, more than a century before.

Bimetallism in the Antebellum United States

In 1792, the United States' first Coinage Act provided for the free and gratuitous coinage of dollar coins, containing 371.25 grains of pure silver, as well as 10-dollar "eagles" containing 247.5 grains of pure gold, thereby establishing a mint ratio of 15 to 1, which at the time roughly corresponded to the ratio of the two metals' market prices. But within a few years, the relative price of gold rose substantially, making it more profitable to exchange gold for silver in the open market than to deliver it to the US Mint for coining. The situation was, in other words, precisely opposite that which England first encountered several decades earlier. Consequently, although the United States remained officially committed to bimetallism, in practice it found itself on a silver standard.

Following the Appalachian gold discoveries, the market price of gold declined, though not enough to keep it from being legally undervalued. The new gold mining interests, however, pressured Congress to reduce the gold content of the eagle to 232 grains (with corresponding reductions in the content of other gold coins), establishing a new mint ratio of 16 to 1, which was almost as far above the market

ratio as the old mint ratio had been below it. The result was a switch from de facto silver monometallism to de facto gold monometallism, which was to be further reinforced by a quadrupling of gold output in consequence of the Californian and Australian finds of 1848 and 1851.

Whether bimetallism need never fall victim to Gresham's Law provided that enough governments subscribe to identical bimetallic ratios, has been a subject of considerable debate. The details cannot occupy us here; but those interested in them should consult the writings on this topic by Milton Friedman (1992), Marc Flandreau (2004), and Christopher Meissner (2015). For a thoroughgoing general survey of bimetallism, see Angela Redish (2000).

Misapplications of Gresham's Law

Although Gresham's Law has many historical applications, it is far from being universally valid. On the contrary, history is also well-supplied with instances in which "good" currencies, instead of being driven out by "bad" ones, have prevailed over them.

Robert Mundell (1998) notes that this anti-Gresham's Law tendency has been especially evident in the realm of international exchange.

> Over the span of several millennia, strong currencies have dominated and driven out weak in international competition. The Persian daric, the Greek tetradrachma, the Macedonian stater, and the Roman denarius did not become dominant currencies of the ancient world because they were "bad" or "weak." The florins, ducats and sequins of the Italian city-states did not become the "dollars of the Middle Ages" because they were bad coins; they were among the best coins ever made. The pound sterling in the 19th century and the dollar in the 20th century did not become the dominant currencies of their time because they were weak. Consistency, stability and high quality have been the attributes of great currencies that have won the competition for use as international money.

Nor is it surprising that Gresham's Law fails to predict the outcome, even in ancient times, of competition among alternative international monies. As we have seen, the law, properly understood, depends on the presence of strictly enforced tender laws, imperfectly informed agents, or both. In the realm of international exchange, where sophisticated traders are able to price "foreign" coins without regard to their official values, these conditions are absent.

Redeemable Banknotes

But Gresham's Law is also incapable of accounting for many instances of currency selection within domestic markets. For reasons we have already considered, it does not predict currency selection outcomes in cases in which different currencies are readily distinguishable from one another and in which people remain free to value those currencies as they please, or to refuse them altogether.

Yet Gresham's Law is sometimes improperly invoked to explain or predict currency selection outcomes in unhindered market settings. For example, Mundell himself (ibid., p. 10) appeals to it to account for the gradual, voluntary substitution, during the eighteenth and nineteenth centuries, of commercial banknotes for gold and silver coin.

Mundell begins by quoting David Hume and Adam Smith on the advantages Great Britain enjoyed relative to France thanks to its having allowed "paper credit" to take the place of gold and silver coin. "The substitution of paper in the room of gold and silver money," Smith observed, "replaces a very expensive instrument of commerce with one much less costly, and sometimes equally convenient. Circulation comes to be carried on by a new wheel, which it costs less both to erect and to maintain than the old one" (ibid.)

According to Mundell, this substitution of redeemable banknotes for coin was just another instance of Gresham's Law at work, for what that law really implies, in his view, is that "cheap money drives out dear, if they exchange for the same price." Irving Fisher (1894, p. 527n2) took a similar view, identifying Gresham's law with the proposition that, when different kinds of money are available, "the cheaper will be substituted for the dearer."

Important as it has been historically, the tendency for less efficient monies to give way to more efficient ones is hardly what concerned Gresham, Oresme, Copernicus, or Aristophanes! Those writers were addressing, not monetary selection outcomes in competitive and otherwise unhampered markets, but the peculiar selection processes at work in markets characterized by state coinage monopolies, legal tender laws, and such. In those markets, monetary selection favored, not more "efficient" monies but genuinely "bad" ones, meaning ones which, though officially just as valuable as others, were held by the public to be less valuable in fact.

The commercial banknotes to which Hume and Smith refer were, in contrast, regarded by their users to be, not only as "good" as, but *better* than, the coins they replaced; and that opinion rested, not on the presence of legal tender laws but on the combination of banknotes' relative convenience and their issuers' readiness to redeem them in specie on demand. In fact, commercial banknotes generally were not legal tender: although the notes of some state-favored banks were sometimes legal tender (Bank of England notes, for example, were made legal tender in 1833), the notes issued by ordinary commercial banks, including the Scottish banks to which Hume and Smith's remarks directly refer, were not.

As Charles Kindleberger (1989, pp. 43) quite correctly observes in his Raffaele Mattioli lecture on Gresham's Law, "Convertibility of one money into another. . .is the touchstone. When such convertibility is maintained, Gresham's Law is held at bay." More generally, to treat any "good" currency-selection outcome as an instance of Gresham's Law is to strip that law of the paradoxical quality that is its very essence.

Private Coinage

Another example of an inappropriate appeal to Gresham's Law occurs in *Money and the Mechanism of Exchange* (1875), William Stanley Jevons's popular textbook on

money. Here Jevons appealed to Gresham, not to explain why certain products of the same official mint prevail over others but to counter Herbert Spencer's defense, in *Social Statics* (1851, pp. 396–402), of *private* coining.

Spencer maintained, according to Jevons, "that just as people go by preference to the grocer who sells good tea, and to the baker whose loaves are sound and of full weight, so the honest and successful coiner would gain possession of the market, and his money would drive out inferior productions." In reply Jevons asserted that "Gresham's Law alone furnishes a sufficient refutation of Mr. Herbert Spencer's doctrine":

> People who want furniture, or books, or clothes, may be trusted to select the best which they can afford, because they are going to keep and use these articles; but with money it is just the opposite. Money is made to go. They want coin, not to keep it in their own pockets, but to pass it off into their neighbour's pockets; and the worse the money which they can get their neighbours to accept, the greater the profit to themselves. Thus there is a natural tendency to the depreciation of the metallic currency, which can only be prevented by the constant supervision of the state.

Although Jevons was an outstanding economist, and Spencer was not, Spencer's controversial view was in fact more consistent both with a proper understanding of Gresham's Law and with the results of actual private coinage episodes. Thus when, during Great Britain's Industrial Revolution, the Royal Mint stopped producing small change, a score of private firms began striking custom-made private coins, called "tradesman's tokens," to take official coins' place. Instead of being shoddy, many of these tokens were of higher quality than their official (poorly engraved, badly worn, and easily counterfeited) counterparts. For that reason the tokens quickly became Great Britain's preferred currency for wage payments and retail sales, remaining so until 1821, when the government suppressed them (Selgin 2008).

California's gold rush supplies further proof of the inapplicability of Gresham's law to a private and competitive coinage system. Apart from its headquarters in Philadelphia, the US Mint at the time had only one branch, in Charlotte North Carolina. So California gold miners who wished to avoid the delay and risk involved in sending gold east for coining patronized local, private mints instead. Between 1849 and 1855, no fewer than 15 such mints catered to them; and while a few of these firms produced "bad" coins, meaning ones with bullion contents that fell short of their declared values, their products quickly fell out of favor with local merchants, while those of several more reputable mints, which had as much, if not more, gold in them than their US Mint counterparts, rapidly gained market share (Summers 1976).

Why did gold miners not patronize mints that made the worst, rather than the best, coins, so as to be able to pass those coins "off into their neighbour's pockets" and thereby profit more themselves? The simple answer is, first, that no legal tender laws were present to prop-up demand for "bad" coins, and, second, that while imperfect information may for a time have encouraged the circulation of inferior coins (thereby allowing some bad mints to at least get off the ground), over time, thanks to the publication of assay results, the quality of different mints' products became more widely known. The threat of failure on one hand and the prospect of success on the other supplied a powerful motive for competitive mints to look after their

reputations – a motive that official mints generally lacked, except to the extent that they supplied coins for international use. Merchants' eventual refusal to accept coins from less reputable mints in turn encouraged gold miners to deal only with the more reputable ones.

Rolnick and Weber's Challenge

In a widely cited article in the *Journal of Political Economy*, Arthur Rolnick and Warren Weber (1986) challenge not just the naïve, "universal" version of Gresham's law, but the more careful version defended here. The naïve version of the law, they correctly note, "is simply contradicted by history." But the qualified version, they say, is no better "because it relies on the existence of a fixed rate of exchange that is different from the market price." There is, according to Rolnick and Weber,

> no evidence that such a fixed rate of exchange ever existed, and that is not surprising since it is hard to believe it ever could exist. If such a rate ever were managed-through a mint policy or a legal tender law, for example – it would imply potentially unbounded profits for currency traders at the expense of a very ephemeral mint or a very naive public. (ibid, p. 186)

Having on these grounds declared Gresham's Law a "fallacy," Rolnick and Weber propose in its place two alternative propositions. The first of these holds that "bad" and "good" money may in fact circulate together, with one type circulating at par, and the other at premium (if "good") or a discount (if "bad"). The other, which becomes relevant in the presence of substantial market-based costs of nonpar exchange, is that par money – that is, whichever money serves as the medium of account – will drive nonpar money out of circulation. The second possibility is the one previously referred to as "Rolnick and Weber's Law."

In a reply to Rolnick and Weber, Robert Greenfield and Hugh Rockoff (1995) focus their attention on the first proposition, which they claim is contradicted by nineteenth-century US experience. But even if correct, that finding falls short of vindicating Gresham's Law, for it leaves open the possibility that US experience supports, not the Gresham's Law prediction that "bad money drives good money from circulation" but "Rolnick and Weber's Law" that par money drives nonpar money from circulation.

Indeed, as Greenfield and Rockoff themselves recognize, one of the cases they consider – California's continued adherence to the gold standard during the Greenback era – appears consistent with Rolnick and Weber's Law rather than Gresham's: although greenbacks were legal tender, Californians refused to accept them in lieu of gold. Since greenbacks depreciated considerably relative to gold, in California (and also in Oregon), in contrast to the rest of the United States, it was "good" gold that drove away "bad" greenbacks. Commenting on this outcome, Greenfield and Rockoff (1995) wonder whether it lies "outside the scope of

Gresham's law because Californians considered the greenback foreign exchange, not money." But if Californians were able to take this view, it was because Washington either would not or could not force them to obey its laws. In other words, it meant that at least one of the two sufficient conditions for having Gresham's Law take effect was lacking.

The fundamental problem with Rolnick and Weber's criticism of Gresham's Law is not that their alternatives, or at least the one I have dubbed "Rolnick and Weber's Law," are never valid, but that they simply cannot account for the frequent (though far from universal) tendency, in premodern times especially, for "bad" rather than "good" money to became nations' de facto medium of account (Selgin 1996). Rolnick and Weber succeed in downplaying that frequent outcome, and with it the empirical relevance of Gresham's Law, in part by simply not considering those early episodes of debasement that were the law's original inspiration. Instead, they refer only to instances of bimetallism or the introduction of paper substitutes for coin; yet more often than not, even in those episodes, as Greenfield and Rockoff's appraisal makes clear, "bad" money tended to prevail.

And what about Rolnick and Weber's claim that Gresham's Law "relies on the existence of a fixed rate of exchange" that would have bankrupted any mint that attempted to maintain it? It is true that Gresham's Law is often said to rely on the presence of a "fixed exchange rate." Friedman and Schwartz (1963, p. 27n16), for example, observe that it "applies only where there is a fixed rate of exchange" between two distinct monies; F. A. Hayek (1978 p. 43) likewise claims that Gresham's law "applies only if a fixed rate of exchange between different forms of money is enforced." The literature is full of similar statements.

But statements to the effect that Gresham's Law depends on "fixed exchange rates" need not *and should not* be understood as implying that it depends on the presence of an *operational* fixed rate, as opposed to a declared legal equivalence, as when coins containing different amounts of metal, or made of two different metals, or coins on one hand and irredeemable paper on the other are officially deemed to represent identical sums. Certainly no one familiar with bimetallic coining arrangements ever supposed that the mints involved offered to exchange gold for silver and vice versa at their implied mint ratios! Rather than depending on mint officials' risking bankruptcy by actually offering to exchange "good" money for its face value equivalent in "bad" money, Gresham's Law refers to the quite unintended consequence of governments' futile attempts to force their citizens to treat moneys they consider unequal equally.

Cross-References

► Money and Prices in Colonial America
► Rise and Demise of the Global Silver Standard

References

Balch TW (1908) The law of Oresme, Copernicus and Gresham. Proc Am Philos Soc 47(188):18–29
Bezanson A (1951) Prices and inflation during the American Revolution: Pennsylvania, 1770–1790. University of Pennsylvania Press, Philadelphia
Breckenridge SP (1903) Legal tender: a study in English and American monetary history. University of Chicago Press, Chicago
Bridrey E (1906) La Théorie de la Monnaie au XIV Siecle: Nicolas Oresme. V. Giard et E. Brière, Paris
Calomiris CW (1988) Institutional failure, monetary scarcity, and the depreciation of the continental. J Econ Hist 48(1):47–68
Chown J (1994) A history of money from AD 800. Routledge, London
Dutu R (2004) Moneychangers, private information and Gresham's law in late medieval Europe. Rev Hist Economica – J Iber Lat Am Econ Hist 22(3):555–571
Fisher I (1894) The mechanics of bimetallism. Econ J 4(15):527–537
Flandreau M (2004) The glitter of gold: France, bimetallism, and the emergence of the international gold standard, 1848–1873. Oxford University Press, New York
Friedman M (1992) Bimetallism Revisited. In: Money mischief: episodes in monetary history. Harcourt, Brace, Jovanovich, New York, pp 126–156
Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1863–1960. Princeton University Press and NBER, Princeton
Giffen R (1891) The Gresham law. Econ J 1(2):304–306
Greenfield RL, Rockoff H (1995) Gresham's Law in nineteenth-century America. J Money Credit Bank 27(4, Part 1):1086–1098
Harris SE (1930) The Assignats. Harvard University Press, Cambridge
Hayek FA (1978) Denationalisation of money: the argument refined. Institute of Economic Affairs, London
Head BV (1888) A catalogue of the greek coins in the british museum: Attica-Megaris-Aegina. Longmans & Co, London
Hicks J (1969) A theory of economic history. The Clarendon Press, Oxford
Jevons WS (1875) Money and the mechanism of exchange. D. Appleton and Company, New York
Johnson C (1956) The de moneta of nicholas oresme and english mint documents. Thomas Nelson and Sons Ltd, London
Kindleberger CP (1989) Third lecture: Gresham's Law. In: Economic laws and economic history: the Raffael Mattilioni lectures. Cambridge University Press, Cambridge, UK, pp 43–66
Li L-F (2009) After the great debasement, 1544–51: did Gresham's Law apply? LSE Department of Economic History working papers no. 126/09, October
Macleod HD (1858) Elements of political economy. Longmans, Green & Co., London
Macleod HD (1896) The history of economics. Bliss, Sands, and Co, London
Macleod HD (1899) Gresham's Law and bimetallism. Sound Currency 6(5):57–72
Meissner CM (2015) The limits of bimetallism. NBER working paper 20852, January
Miskimin H (1989) The enforcement of Gresham's Law. In: idem, Cash, credit and crisis in Europe, 1300–1600. Variorum Reprints, London, pp 147–61
Mundell R (1998) Uses and abuses of Gresham's Law in the history of money. Zagreb Journal of Economics 2(2):3–38
Munro JHA (2012) Coinage debasements in Burgundian flanders, 1384–1482: monetary or fiscal policies? In: Nicholas D, Bachrach BS, Murray JM (eds) Comparative perspectives on history and historians: essays in memory of Bryce Lyon (1920–2007). Medieval Institute Publications, Western Michigan University, Kalamazoo, pp 314–360
Redish A (2000) Bimetallism: an economic and historical analysis. Cambridge University Press, Cambridge
Rolnick AJ, Weber WE (1986) Gresham's Law or Gresham's fallacy? J Polit Econ 94(1):185–199

Rolnick AJ, Velde FR, Weber WE (1996) The debasement puzzle: an essay on medieval monetary history. J Econ Hist 56(4):789–808
Sargent TJ, Smith B (1997) Coinage, debasements, and Gresham's Laws. Economic Theory 10(2):197–226
Sargent T, Velde F (2002) The big problem of small change. Princeton University Press, Princeton
Selgin G (1996) Salvaging Gresham's Law: the good, the bad, and the illegal. J Money, Credit, Bank 28(4):637–649
Selgin G (2008) Good money: birmingham button makers, the royal mint, and the beginnings of modern coinage. University of Michigan Press, Ann Arbor
Spencer H (1851) Social statics. John Chapman, London
State of Pennsylvania (1852) Minutes of the supreme executive council, vol 11. Theodore Fenn & Company, Harrisburg
Summers B (1976) Private coinage in America. The Freeman 26(7):436–440
Sumner WG (1892) The financier and the finances of the American revolution, vol 1. Dodd, Mead & Company, New York
Velde FR, Weber WE, Wright R (1999) A model of commodity money, with applications to Gresham's Law and the debasement puzzle. Rev Econ Dyn 2(1):291–323
Webster P (1791) Political essays on the nature and operation of money, public finances, and other subjects. Joseph Crukshank, Philadelphia
White AD (1933) Fiat money inflation in France: how it came, what it brought, and how it ended. D. Appleton-Century Company, New York
White LH (1984) Competitive payments systems and the unit of account. Am Econ Rev 74(4):699–712

Part III

Trade, Money Markets, and International Currencies

Money, Trade, and Payments in Preindustrial Europe

9

Meir Kohn

Contents

Abstract

The means of payment and remittance initially available in preindustrial Europe—coin and bullion—were quite inadequate to the needs of commerce. So commerce developed its own system of payment and remittance based on the transfer of IOUs. We will examine how this system of payment and remittance evolved—first its initial development during the Commercial Revolution, and then its continuing evolution in subsequent centuries. We will conclude by considering how the evolution of this system contributed to the development of commerce and to economic progress in general.

M. Kohn (✉)
Department of Economics, Dartmouth College, Hanover, NH, USA
e-mail: Meir.G.Kohn@Dartmouth.EDU

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_15

Keywords
Payments · Remittance · Pre-industrial Europe · Banks · Bills of exchange

Merchants were in constant need of making and receiving payment – often in large amounts. They did so when buying and selling goods and services, when borrowing and lending, when paying tolls and taxes, and when discharging debts and distributing profits. Also, because merchants operated in multiple locations distant from one another, they frequently needed to remit funds from one place to another. The costs of payment and remittance constituted a significant part of transactions costs.

The means of payment and remittance initially available in preindustrial Europe – coin and bullion – were quite inadequate to the needs of commerce. So commerce developed its own system of payment and remittance based on the transfer of IOUs. This system developed in the places where the need was greatest – in the organized markets of major commercial cities.

We will examine how this system of payment and remittance evolved – first its initial development during the Commercial Revolution, and then its continuing evolution in subsequent centuries. We will conclude by considering how the evolution of this system contributed to the development of commerce and to economic progress in general.

Means of Payment During the Commercial Revolution

Payment is made by transferring something of value – a means of payment – from one person to another. The means of payment can be an object of intrinsic value – usually a quantity of precious metal, commonly minted into coin. Or it can take the form of an IOU – a promise to pay at a later date.

Coin and Its Problems

Coin, in preindustrial Europe, came in a bewildering profusion of types. Each small jurisdiction minted its own coin, and coin from different jurisdictions varied widely in weight and fineness (Spufford 1988). Because of the crude technology of minting, even new coins differed in weight, and coin was easy to clip and counterfeit. The quality of a given issue deteriorated over time as a result of wear and because of the culling of heavier coins. To keep up with this deterioration, successive issues tended to be of progressively lower weight and fineness. And not only was the coinage of poor quality, but there was never enough of it. There was a chronic shortage of coin throughout the period, punctuated now and then by an acute shortage.

Another problem for merchants was that the relatively low value of silver coins made them inconvenient for large payments: the necessary counting, sorting, weighing, and testing of thousands or even millions of coins could take weeks.

Consequently, for large payments, merchants generally used silver bars or gold coins – when the latter became available in the late thirteenth century. Even so, payment in coin or bullion remained difficult and expensive.

Payment by IOU

As a result of the difficulties of using coin in payment, people generally used IOUs instead whenever they could. However, accepting an IOU requires trust: there has to be some confidence that the IOU will be honored. Consequently, the use of IOUs in payment developed initially only among people who knew one another well.

Most exchange involved sales credit: that was how one merchant sold to another whom he knew and trusted. But it was also how one villager sold to another and how shopkeepers, craftsmen, or innkeepers sold to their regular customers. Selling on credit in this way meant, essentially, accepting an IOU in payment: in exchange for the actual delivery of goods or services, the seller accepted from the buyer a personal promise to pay at a later time.

Managing Payment by IOU

Exchange based on sales credit requires periodic settlement. Someone who has extended credit may wish to liquidate the debt in order to be able to make purchases himself – although, if he too can buy on credit this may not be essential. A more important reason for periodic settlement is the need to limit exposure to default by the debtor. Periodic settlement constitutes a "test of solvency."

Of course, if sales credit must ultimately be settled in currency, the need for currency is not eliminated but merely deferred. However, the need for actual payment in currency can be reduced by setting off one debt against another – by *netting*. With netting, only a part of the original debt remains to be settled in coin, and the more successful the netting the smaller the part. And, in practice, what remained was often settled in kind rather than in cash.

When sales credit was reciprocal, as it often was in local trade, there was considerable opportunity for bilateral netting: a baker who sold bread on credit to a shoemaker would eventually need a new pair of shoes. The extended periods for which debts were often carried made this sort of netting easier. In larger communities, however, and with more complicated patterns of trade, there were fewer opportunities for bilateral netting. Some form of multilateral netting was therefore desirable.

The simplest form of multilateral netting is the assignment of third-party debt. For example, a merchant who owes 100 florins to a supplier and is in turn owed 80 florins by a customer can assign the customer's debt to the supplier in partial settlement. Repeated assignment of debts in this way creates opportunities for bilateral netting – for example, if the merchant's supplier happens to be in debt to the merchant's customer.

The assignment of third-party debt was common among merchants. In Mediterranean Europe, merchants generally assigned debts via the books of the parties

concerned. Merchants in the North did not keep books: instead, they recorded debts by means of promissory notes ("letters obligatory"). So it was these they assigned in settlement. The assignment of third-party debt was quite common, too, within local communities. Complex chains of assignment were common, even in the country.

Limitations on Payment by IOU

A system of payment based on IOUs works only within a community of people who know one another. Each knows the other's credit, and the need to preserve that credit gives each an incentive to honor his debts. Debts can be carried for extended periods, because no one is going to flee to escape his obligations. And if trust between the parties breaks down, there is a common formal order to call upon (of village, association, or city). Multilateral netting through the assignment of third-party debt is possible within a community, because credit is common knowledge: a creditor would know whether or not to accept the debt of a particular third party in settlement.

The principal problem for a payments system based on IOUs is what to do when a participant fails to settle as promised. This poses a danger to the system as a whole, because each participant relies on timely settlement by others to be able to settle his own debts. One failure to settle can therefore lead to another in a domino effect that eventually brings the whole system to a halt.

The problem can largely be eliminated, of course, through flexibility – by creditors being willing to give debtors extra time to settle when necessary and appropriate. Such flexibility is possible, once again, only within a community. Within a community, a creditor is likely to know whether a particular debtor is temporarily short of funds or insolvent – flexibility being appropriate in the former case but not in the latter.

When it comes to payments between strangers, however, none of this works. The credit of strangers is unknown, the incentive of a continuing relationship is usually absent, and social constraints on behavior are weak. The obvious alternative in these circumstances is payment in cash. Indeed, despite the problems of the coinage, cash was the chosen means of payment for the myriad, relatively small, transactions in town markets. Cash was not, however, a practical solution for payments between merchants. Instead, in the major commercial centers, there developed a system of payment based on the assignment of the IOUs of banks.

Deposit Banks

The banks in question had their origin in moneychanging. Moneychangers facilitated payment in cash by exchanging one type of coin for another and by checking the quality of coin tendered. At some point, rather than checking coin each time it was used in payment, moneychangers began to check it once and then place it in storage for their customers; their customers could then make payment, not by transferring the coin itself, but by assigning the claim to the stored coin. In this way, moneychanger-bankers created a sort of book-entry money by "immobilizing" the actual coin and transferring the claim to it on their books (there is an obvious

parallel with the very modern practice of immobilizing stocks in a depository). Rather than physically handing over coin, a payer could transfer a credit on the books of his bank to the benefit of the payee. This practice was known as "payment in bank."

This form of banking, known as deposit banking, seems to have originated in Genoa in the twelfth century. By the fourteenth century, payment in bank predominated in all the major trading centers, both commercial cities and fairs. Every merchant residing in such a center, whether native or foreign, would hold deposits at one or more deposit banks; visiting merchants would open temporary accounts. In cities, even large ones, that were not primarily commercial centers, deposit banking either did not develop at all (for example, in Paris and London), or it did not develop to the same extent (for example, in Florence).

Payment in Bank as Assignment of Third-Party Debt

While payment in bank originated as a way of improving the efficiency of payment in coin, it was at the same time an example of settlement by the assignment of third-party debt. As we have seen, for the assignment of such a debt to be acceptable, the third party in question must be known to the potential assignee and considered a good credit. In major commercial centers where trade with strangers predominated, deposit banks played the role of such a third party – universally known and trusted.

Deposit banks were not the only ones to play the role of trusted third party. Many merchants held balances with innkeepers and brokers and they settled with one another by transferring balances on the books of these intermediaries (Nicholas 1992; Nicholas 1979). Also, merchants often held working deposits with merchant banks (about which we will learn presently) and used these too for payment purposes (de Roover 1944).

Trust in deposit banks was reinforced by public regulation. Banks were licensed by the city authorities, who required from potential bankers a solemn oath – then a very serious matter – together with third-party sureties. The authorities also restricted a bank's permissible assets and activities. Cities sometimes even went so far as to provide guarantees themselves: for example, in Bruges in 1309, when a series of bank failures infuriated foreign merchants and threatened the city's trade, the authorities instituted a form of deposit insurance (de Roover 1974 Ch. 15).

Payment by assignment of bank deposits had additional advantages. Because the debt of the *same* third party was assigned in many payments, netting was easier (Usher 1934). As a result, bank deposits could be used to mediate a large volume of transactions with very little need for final settlement in cash. Moreover, because bank deposits were so useful as a means of payment, depositors tended to hold on to them rather than converting them immediately into cash.

Bank Lending and Overdraft

The willingness of people to hold on to bank deposits made it possible for banks to become lenders. They did so by creating a deposit for a borrower without any corresponding coin actually having been deposited. Since such a "fictitious deposit" was indistinguishable from the real thing, the borrower could nonetheless use it

to make payment. As a consequence of this practice, the amount of deposits outstanding came to exceed the amount of coin held against them – fractional reserve banking.

The ability of banks to lend enabled them to offer a solution to the problem noted earlier of traders being unable to settle on time – the problem of liquidity risk. The solution that worked within a community – flexibility – was not feasible in transactions between strangers. Instead, the deposit bank provided a "liquidity facility." If a depositor temporarily lacked a sufficient balance in his account to settle, the banker would allow him to "overdraw" his account. That is, the banker would make him a loan that permitted him to settle. Such lending was, of course, at the discretion of the banker. Since transactions were conducted orally, overdraft was impossible without the banker's consent (Usher 1934 p18).

The banker was in a position to provide such credit, because he knew his customers and because his customers had an incentive to repay in order to preserve the banking relationship. Such overdraft lending was a natural extension of the bank's function as a payments intermediary.

Banking at the Fairs

The purest example of deposit banking as assignment of debt was to be found at the great medieval fairs, particularly at the Fairs of Champagne. Trading at Champagne was divided into periods. The first period were devoted mainly to the sale of cloth, with the sellers predominantly from Flanders, and the buyers from Italy. A subsequent period was devoted to the sale of oriental spices and drugs: now it was the Italians who were the sellers and the Flemings, the buyers. (This is a considerable simplification of the actual procedure.)

The fair banks provided the payments system that made this separation into trading periods possible. In the first period, the Italians paid for their purchases from Flemish merchants by transferring to them credits on the books of the fair banks: payment was final in that the banks guaranteed settlement. In the later period, the Flemish merchants used the credits they had accumulated in this way to pay for their own purchases of spices and drugs from Italian merchants.

The credits, in this case, were not really deposits. Although evidence is scarce, it seems the Italians did not deposit coin with the bankers in advance. Rather, the bankers created the credits on their books by allowing Italian merchants to overdraw their accounts during the first period of trading. The bankers at Champagne were in a position to extend credit in this way, because they too were from northern Italy, predominantly from Asti and Piacenza (Abulafia 1997). These overdrafts were then largely extinguished by payments the Italians received during the later period of trading (de Roover 1954 p204). Consistent with this interpretation, "deposits" at the fair banks were not payable on demand, but only during the settlement period that concluded the fair.

Interfair Deposits

In the settlement period, merchants whose balances were negative had to cover them. Although they could, in principle, have settled in cash, it was more common to roll

over the unpaid balance to the following fair. This was usually done by borrowing from other merchants who had positive balances with the bankers on completing their trading. Merchants with positive balances could receive payment in cash if they wished, but many preferred instead to lend them. To facilitate the borrowing and lending of bank balances, there soon developed an active market in interfair loans or "interfair deposits."

Interfair deposits were a convenience both for borrowers and for lenders, since it enabled them to balance their purchases and sales over time without having to ship bullion back and forth between their home cities and the fairs. For example, a merchant who bought more than he sold at one fair could balance this by selling more than he bought at another: he did not have to settle his "deficit" in cash. The market for interfair loans enabled netting to be extended *across* fairs (Nicholas 1992 p171). Grafe (2001) describes the practice at the Castilian fairs and Farmer (1991), at the English fairs.

The market for interfair deposits also played the role of a liquidity facility. If a debtor was unable to settle a debt to a stranger as promised, he could borrow temporarily in this market – from someone who knew and trusted him – to pay his creditor on time.

Means of Remittance During the Commercial Revolution

Initially, merchants traveling to distant trading centers had either to balance their purchases and sales there or to transport coin or bullion to bridge the gap. For example, if a merchant planned to purchase goods in a distant market, he could either bring with him goods of equal value and sell them there to provide the necessary funds or he could bring the necessary funds in the form of bullion or coin. However, because each place had its own coinage, bullion or coin from other places had to be melted down and reminted. And, because coin and bullion were especially vulnerable to theft and robbery en route, their transportation required extra security. Remittance by IOU offered a less-expensive alternative.

Remittance by IOU

A merchant could deliver funds to someone in his home city in exchange for a promise of funds in the distant market; he could then use the funds he received there to cover any excess of his purchases over his sales. Similarly, a merchant who sold more than he purchased in a distant market could repatriate the excess funds by delivering them to someone there in exchange for a promise of repayment at home.

Cambium Contracts and Letters Obligatory

Remittance by IOU seems to have originated in the late twelfth century in the inland cities of northern Italy as part of their system of trading with the Fairs of Champagne (Ashtor 1983). The earliest instrument of remittance was the *cambium*

contract – a promissory note, drawn up formally by a notary (Blomquist 1990). In such a contract, one merchant acknowledged receipt from another of payment in local currency and promised to repay him in a different place at a future date in the currency of that place. To settle the transaction on the specified date, the two parties concerned, or their agents, had to be present at the place of repayment.

A similar instrument, the letter obligatory, appeared a little later in northern Europe. The letter obligatory was a promise by one merchant to pay another. Payment could be in the same place or in another and was often in the same currency. The letter obligatory was not a formal document drawn up by a notary like the *cambium* contract, but rather an informal document written in the hand of the debtor.

Like sales credit, the *cambium* contract and the letter obligatory were direct obligations of one party to another: no intermediary was involved. Consequently, again like sales credit, they required trust, and their use was therefore limited to merchants who knew one another and who had recourse, if necessary, to a common system of formal order. Expansion of remittance by IOU to transactions between strangers had to await the emergence of a mediated instrument.

Bills of Exchange

By the late thirteenth century, there had emerged a number of large trading companies, each having branches in many of the principal commercial centers. Such companies found it easy to offer the service of remittance as a sideline. They accepted payment at one of their branches in local coin and sent instructions to a branch at a different location to repay in the coin current there. The instrument they used for this transaction was called a *lettera di cambio* or bill of exchange. (The Knights Templar, a military order, had monastic houses around Europe and in the Holy Land; they used these "branches" in much the same way to provide remittance services (Barber 1994).)

Remittance by bill of exchange involved four parties. The "deliverer" delivered funds to the "taker" and received from him the bill of exchange. The deliverer sent the bill to his own agent at the designated place of payment – the "payee." The payee presented the bill to the agent of the taker there – the "payer." In the meantime, the taker would send notification (an *avisa*) to the payer, instructing him to honor the bill when it was presented.

The bill of exchange, like the letter obligatory, was an informal document, not a notarial deed. This lowered the cost and increased the speed of the transaction but also made it unenforceable initially in civil courts. However, merchant courts did recognize such informal documents. Moreover, since the taker-payer was typically a large trading company which provided remittance services as part of its business, it had an interest in preserving its reputation for honesty and reliability. For additional security, it was usual to begin the text of the bill with "*Al nome di Dio*," making default blasphemous.

Remittance by bill of exchange depended, therefore, on the existence of large international trading companies with multiple foreign branches willing to provide this service – willing to act as "merchant banks." Such companies were predominantly Italian – initially Lucchese, then mostly Florentine. Non-Italians did

play a minor role: Catalan merchant banks, for example, were active in Bruges and Palermo.

The Exchange Network

Cities served by merchant banks were known as "banking places." In northern Italy, the center of medieval commerce, most cities were banking places as were many other cities in the Mediterranean. There were fewer banking places in northern Europe – the fairs of Champagne, and when Champagne declined, Paris, Bruges, and London. Banking places constituted a network, and only within this network was it possible to remit funds by bill of exchange (de Roover 1954; Mueller 1997; Spufford 1988).

Centers of the Exchange Network

Banking places were not all, however, of equal importance. At any one time, the exchange network had a center and perhaps one or two subcenters. The first great center, from the late twelfth century, was Champagne, with Genoa as a subcenter; Lucchese merchant banks dominated both. By 1320, Venice had replaced Champagne as the center of the exchange network, which was by then dominated by the Florentines. Bruges was the subcenter for northern Europe; exchange there was largely in the hands of fewer than a dozen Florentine houses and a few Lucchese.

As a rule, it was the great commercial centers that became centers of the exchange network. It is notable that Florence, for example, never became a center of the exchange network, even though Florentines dominated the business of exchange for centuries. However, Florence – unlike Genoa or Venice – was more a center of manufacturing than a center of commerce. The great commercial centers were by their nature the source and destination of a large volume of remittance. The large volume created a liquid and competitive market. And the liquid and competitive market lowered costs and attracted more business.

Consequently, while it was possible to remit directly from one ordinary banking place to another, most remittance was indirect, routed through one of the centers (Boyer-Xambeu et al. 1994 p74). For example, a fourteenth-century Italian merchant remitting funds from London to northern Italy would usually find it cheaper to do so by remitting to an agent in Bruges and having the agent remit from there to the final destination.

Organized Markets for Exchange

In most banking places, and certainly in the centers of the network, there would be an organized market for exchange. In Bruges, for example, the exchange market was held in the square that faced the inn of Van der Beurse, called the Beursplein or Place de la Bourse. Trading there took place each day during fixed hours, opened and closed by the ringing of a bell (Blockmans 1992).

An organized market provided a mechanism for setting prices – which were in this case rates of exchange. For any particular transaction, the rate was negotiated

between the parties, but negotiation was constrained by the general state of the market. Potential takers and deliverers generally relied on specialized brokers to find the best rate available, and the activity of these brokers kept rates reasonably uniform.

Organized exchange markets relied on deposit banks for settlement. Indeed, settlement of bills of exchange soon provided deposit banks with most of their business (Usher 1934 p7). Settlement in bank minimized the need for settlement in cash, thereby lowering the cost of transactions and speeding them up. Indeed, deposit banks were so indispensable to the smooth functioning of the exchange market that when the banks were closed in Venice, due to snow or a local feast day, the exchange market shut down as well (Mueller 1997 p 570).

Overdraft lending by deposit banks and markets for interfair loans provided liquidity facilities for the exchange market. However, the exchange market was itself a source of liquidity. A merchant temporarily lacking the funds to honor a bill presented for payment could – if his credit was good – become the taker on another bill, thereby acquiring the funds he needed to meet his obligation on time.

The Further Development of the Payments System

In the centuries that followed the Commercial Revolution, the system of payments and remittance continued to develop. Its development was largely driven by the need to address the system's central problem – the chronic instability of deposit banks. Throughout the period, the frequent failure of banks disrupted markets, paralyzing trade and exchange.

Two solutions to this problem eventually emerged – a system of payments without deposit banks and a new kind of deposit bank that was less likely to fail. To understand the nature of these solutions, we first need to understand why it was that deposit banks were so prone to failure.

The Instability of Deposit Banks

As we have seen, the lending of deposit banks began with the provision of short-term overdraft to depositors. But banks soon expanded their lending well beyond this, and as a result their deposit liabilities come to exceed by a large margin the amount of coin they held against them. This situation exposed banks to problems of liquidity and solvency: they needed to be able to convert deposits into coin on demand, and they had to ensure that losses on their loans did not compromise their ability to honor their commitments to their depositors.

For several reasons, ensuring the liquidity and solvency of banks was much more difficult then than it is today. First, banks were extremely small. No bank in fourteenth century, Bruges for example, had more than 100 depositors or more than $1 million, in today's money, in deposits (de Roover 1948). Venice's banks were probably the largest: in 1500, its ten banks had between them some 4000 deposits

totaling about one million ducats – roughly $60 million in today's money (Lane 1937). Because bank failures were frequent, depositors were nervous about the safety of their deposits, and the slightest rumor could set off a run. Since a dozen large depositors might hold two thirds of the bank's deposits, simultaneous withdrawals by only a few of them could cause the bank to fail.

A second reason why banks were so unstable was the lack of safe and liquid assets in which they could invest: no such assets then existed. Much of a typical bank's assets took the form of equity investments in trading ventures. These were risky, illiquid, and of relatively long maturity (a trading voyage could take 2 years). The other common bank asset, the overdraft loan, was hardly much better: the maturity of such a loan was indefinite, and, just like an equity investment, it could not be turned into cash in an emergency. Moreover, whatever the form of a bank's assets, they were generally poorly diversified: a small number of large investments or loans accounted for most of the total.

A third reason for the instability of banks was that they operated in a highly unstable environment. War, or even the threat of war, could interrupt trade for extended periods, and the needs of war finance could empty a territory of coin. Commodity prices were highly volatile, and a collapse of prices could be disastrous for banks that had invested directly in commodities or that had extended loans or guarantees to those who traded in them. The monetary environment, too, was highly unstable. A general shortage of coin would increase withdrawals, depleting bank reserves. A strengthening of the coinage (deflation) would raise the real value of the bank's liabilities, decrease the value of its reserves, and, at best, leave other assets unchanged. The result could be catastrophic. (Since debasement – inflation – had just the reverse effects, banks actually benefited from it.)

All of these problems came to a head in the second half of the fifteenth century (de Roover 1954; Van der Wee 1977). The "bullion famine," which peaked at that time, was devastating for banks. The scarcity of coin strained their liquidity and the general deflation and slump threatened their solvency. Bank failures increased the demand for coin, which further exacerbated the shortage. Problems were contagious, as the market for bills of exchange transmitted tight liquidity from place to place (Mueller 1997 Ch. 4). Bank failures occurred in waves that continued into the sixteenth century – most notably in Venice, Florence, and Bruges.

Payments Systems Without Banks

One solution for the instability of deposit banks was simply to do without them. This became possible as a result of developments in the exchange network. As we have seen, Venice replaced Champagne as the primary center of the network in the early fourteenth century. Venice, in its turn, was succeeded by Geneva and then by Lyons in the fifteenth century; and Antwerp emerged as a secondary center to Lyons in the sixteenth. By then, merchant bankers from Genoa were the dominant group.

Indeed, the Genoese did enough business to be able to establish an alternative center of their own, first at Besançon, and then later at Piacenza – where it continued

to be known as the "Bisanzone." The Bisanzone fairs were unique in that they were created solely for the purpose of clearing bills of exchange and served no other function. (The Bisanzone fairs were the keystone of the "Genoese system" of finance and remittance that connected Seville, Genoa, and Antwerp.) With the decline of Lyons and Antwerp late in the sixteenth century, the Bisanzone fairs came to dominate the exchange network and continued to do so well into the seventeenth century.

Settlement at Lyons and Bisanzone

The exchange markets of Lyons and Bisanzone – unlike those of Champagne and Venice – did not rely on deposit banks for settlement. Unlike Champagne, they were not public markets but "inside markets," in which merchant bankers transacted only with one another to settle bills drawn on them. Because the participants were relatively few and of known reputation, they did not need to assign the debt of a trusted third party – of a deposit bank. They could just as easily assign the debt of other merchant banks.

Settlement proceeded roughly as follows. Participating merchant bankers compared their "market books," in which each had listed his bills payable and receivable. Pairs of bankers netted bilaterally payments due to each other. Any remaining imbalance was covered either by assigning debts due from third parties, which might be netted later in the process, or by drawing new bills. These bills might be payable at the next fair or at another banking place. Instead of relying on deposit bank overdraft for liquidity, participants accommodated each other through "overdraft" on their own market books. Instead of relying on interfair deposits to balance credits and debits over time, participants drew new bills of exchange.

The development of this highly efficient means of clearing and settlement greatly increased the efficiency of the exchange network: "De Rubys thus reports that a million *livres tournois* were paid in the course of one morning without a single *sol* being disbursed." (Boyer-Xambeu et al. 1994 p93). By the end of the sixteenth century, a group of some 60 merchant bankers was meeting regularly 4 times a year at the Bisanzone fairs, "to settle the debits and credits of half of Europe" (Van der Wee 1977). Rather than transacting with one another directly, a large number of strangers – the customers of the merchant banks – engaged in financial transactions with one another indirectly through a small group of merchant bank intermediaries.

Settlement at Antwerp

This system of settlement worked for the exchange markets of Lyons and Bisanzone because the number of participants was small. The exchange market at Antwerp, however, had a large number of participants, coming from many different countries. In such a public market, one would have expected settlement to continue to rely on payment in bank. However, Antwerp had no banks. In 1489, following a series of banking crises, the rulers of the Low Countries had banned banks altogether (de Roover 1974).

This left the exchange market at Antwerp with no choice but to rely on the assignment of *private* third-party debt – primarily letters obligatory. Netting and

settlement took place on each of the four "quarter days" that had originally marked the Brabant fairs: letters obligatory and bills on Antwerp were usually written to mature on those days. Not surprisingly, this reliance on private debt did not work well. However, once the problems were addressed in a series of legal innovations, performance improved significantly.

To understand the nature of these innovations, let us consider a hypothetical example. Suppose Jan has an IOU from Charles; Jan assigns this IOU to Robert in payment. The first problem is that if Charles defaults, only Jan, the original creditor, has the right to sue him: Robert has no direct recourse. This problem was solved in 1507 when the Antwerp civil courts recognized the transferability of letters obligatory. What transferability means is that if Jan assigns Charles's IOU to Robert, it becomes Charles's IOU to Robert rather than to Jan.

Solving this problem, however, created another. If Charles does not pay up, Robert can now sue him, but he no longer has any recourse against Jan: Jan's debt to Robert is considered discharged by the transfer to him of Charles's IOU. Clearly, under these conditions, no one would be willing to accept in settlement the paper of someone of unknown or inferior credit.

This second problem was solved when the courts adopted the principle of negotiability. Under this principle, if Charles fails to pay up, Robert has recourse against Jan, who assigned him Charles's IOU in payment. Jan's debt to Robert is considered discharged only when the assigned IOU is actually paid. Moreover, if Charles's IOU has been assigned more than once, then Robert has recourse against each assigner in turn – from the first to the last. As a result, under negotiability, each additional assignment strengthens the credit of the instrument. The negotiability of letters obligatory was first established by edict for Antwerp in 1536 and for the whole of the Low Countries in 1541.

To keep track of the chain of assignment, it became customary for each assigner to endorse the instrument – to sign it on its back – another legal innovation. Endorsement legally bound the assigner/endorser (in our case, Jan) to indemnify the assignee (Robert) if the issuer (Charles) and the previous assigner/endorsers failed to pay up. The first examples of endorsement date from the 1570s and by the turn of the century the practice was widespread.

Initially, only letters obligatory circulated in Antwerp. However, with the increasing presence there of South Germans and Italians, whose preferred instrument was the bill of exchange, it became necessary to bring bills of exchange into the settlement system. Consequently, by the 1530s, it had become the practice to assign bills of exchange in the same manner as letters obligatory (Van der Wee 1993 Ch. 10). Over the course of the sixteenth century, northern merchants gradually switched over from using the letter obligatory to using the bill of exchange (Van der Wee 1977).

As commercial paper (letters obligatory and bills of exchange) evolved into a negotiable instrument, it became a kind of convertible money, similar in many respects to the deposit of a deposit bank. Merchants in Antwerp used commercial paper as a means of payment: instead of transferring the debt of a deposit bank on the books of the bank, they transferred from hand to hand the debt of nonbanks.

Commercial paper commonly changed hands ten or twenty times before maturity, and a hundred times was not unusual. Circulating paper often found its way back to the original issuer, so extinguishing the debt, and eliminating the need for settlement (Van der Wee 1963). Obviously, the paper with the best "names" circulated the most freely, and the paper of the Fugger merchant bank, the *Fuggerbriefe*, passed from hand to hand almost like modern banknotes (Van der Wee 1977).

The usefulness of commercial paper as a means of settlement was further enhanced by the practice of discounting. Although assignment of commercial paper was acceptable in settlement, when money was tight, creditors were willing to pay a premium for settlement in cash. Those who had cash could exploit this situation by offering to buy commercial paper for cash at a discount. (The practice of discounting was not itself new: there had been earlier examples of the discounting of other types of receivable.)

Those most active in discounting were the "cashiers" or money dealers, unofficial pseudobanks that had appeared after deposit banks had been prohibited. Cashiers escaped the prohibition by insisting that they were not banks – Heaven forbid! – but simple cash-managers.

Discounting commercial paper was authorized officially in 1540, and by the 1550s, it was common practice (Van der Wee 1963). By providing liquidity, discounting improved the quality of the market and increased the acceptability of commercial paper in settlement. It also provided a way for letters obligatory to be used as an instrument of pure credit: a merchant wishing to raise cash could simply write a letter obligatory and discount it with a cashier.

Payments Systems Based on Public Banks

While fairs, like Bizanzone and Antwerp (initially one of the fairs of Brabant), learned to manage without banks, several commercial cities found a different solution to the payments problem in a new kind of bank – the public bank.

The public bank was more stable than a private deposit bank, because its function was limited to payments: any form of lending was prohibited. Giving up lending to protect the bank's payments function in this way made perfect sense in terms of the general good. The payments function of deposit banks was crucial, while their contribution to financing was quite minor. Of course, giving up lending made the bank much less profitable, but this was not much of a problem for a public institution.

The Banco di Rialto

Venice was not the first city to create a public bank: the first was Barcelona, which established the *Taula de Canvi* in 1401. But Venice was the most influential in establishing the practice. Bank failures had been a continuing problem in Venice, repeatedly disrupting commerce and tying up depositors' funds in lengthy and litigious liquidations. In a speech in 1584, a Venetian senator noted that of the 103 private banks founded during the life of the Republic, 96 had failed

(Mueller 1997 p 122). A severe banking crisis in 1584 proved to be the last straw: the Venetian Senate decided to establish a public bank, and the *Banco di Rialto* opened for business in 1587.

The *Banco di Rialto* accepted deposits and allowed deposit transfers and cash withdrawals, but it paid no interest on its deposits, and it allowed no overdraft or other lending. This was clearly not a profit-making enterprise, and the bank's expenses had to be paid out of public funds. The management of the bank was licensed to a private individual – which was the normal practice for public services – but the bank was guaranteed by the Venetian Senate. Although Venice did not prohibit private banks, none were established for several years.

The *Banco di Rialto* was a great success. Its deposits were so useful as a means of payment that there were actually few cash withdrawals, and payment *in banco* was soon at a premium over payment in coin (Van der Wee 1977). The prohibition of overdraft seems to have been ignored, which no doubt improved the efficiency of the payments system. Furthermore, since there was only a single bank, there was no need for interbank clearing, and this too increased efficiency.

The success of the *Banco di Rialto* inspired imitation and many Italian cities set up similar public banks – for example, Milan in 1593 and Rome in 1605. In Genoa, an existing public financial institution, the *Casa di San Giorgio*, began to accept deposits in 1586 (Parker 1977).

The Bank of Amsterdam

The idea of the public bank spread to northern Europe with the establishment of the Bank of Amsterdam (*Amsterdamsche Wisselbank*) in 1609 (Van der Wee 1977; de Roover 1948). However, the motivation in this case was somewhat different.

The city authorities had been suspicious of the new-fangled practices that were spreading to Amsterdam from Antwerp – settlement by assignment and discounting by cashier-bankers. The authorities believed that commercial paper ought to be settled in hard cash. However, when they tried to shut the cashiers down, they met with strong resistance from the city's merchants, who considered the cashiers' services indispensable. So the city set up a "public cashier" as a substitute.

The Bank of Amsterdam offered transferable deposits convertible into good coin. (The Bank traded extensively in precious metals to provide the mints with the bullion needed to produce this coin, and it later came to play a central role in the European and world market for bullion (van Dillen 1934).) There was no overdraft or other lending, so that deposits were fully backed by coin, and all services were free. To get merchants to use the Bank, commercial paper over 600 guilders was required to be payable in deposits of the Bank.

The Bank was spectacularly successful, and similar public banks were soon established in other northern cities. By the end of the seventeenth century, there were some 25 public banks across Europe (Parker 1977).

Despite the success of public banks, private cashier-banks did not disappear. In many places, including Amsterdam, they competed successfully with the public banks by offering better service – particularly by being more willing to extend

overdraft. Amsterdam gave up its attempts to prohibit private cashiers in 1621 and chose instead to regulate them (van der Borght 1896).

Although public banks were more stable than private banks, they were not entirely immune to the temptations of lending – particularly to the city government. In fact, some public banks were created specifically to provide the city government with inexpensive credit. The lending of public banks, therefore, like that of private banks, often ended badly. Indeed, the first public bank, Barcelona's *Taula de Canvi*, failed in the 1460s as a result of excessive lending to the city.

Antwerp and Venice, therefore, found two very different solutions to the problem of banking instability – negotiable bills of exchange and discounting on the one hand and the public bank on the other hand. The two practices were later combined in Amsterdam and in London as the foundation for a more stable system of private banking (Van der Wee 1963). By 1600, goldsmiths and scriveners (scribes) in England were starting to take on the banking functions performed by cashiers in Antwerp and Amsterdam, and the Bank of England was founded in 1694.

The negotiable bill of exchange was a relatively safe and liquid asset. The discounting of bills therefore provided a form of lending well suited to small banks – precisely what earlier had been lacking. Moreover, the public bank lent additional stability to the system by providing liquidity to private banks in times of crisis.

Modern commercial banking that combined these two elements was to play an essential role in financing the expansion of English trade in the eighteenth century and in financing the subsequent Industrial Revolution (Pressnell 1956).

The Economic Contribution of Payment and Remittance by IOU

The development of a system of payment and remittance through the transfer of IOUs contributed to economic progress directly. It also contributed to it indirectly by facilitating financial development.

The Direct Contribution to Economic Progress

The development of a system of payments and remittance that used IOUs rather than bullion and coin reduced transactions costs. Obviously, it reduced the cost of payments and remittance, but it also reduced transactions costs in other ways.

The Reduction in the Costs of Making Payment and Remittance

Payment by IOU greatly reduced the need to count, weigh, examine, and sort coin. And remittance by IOU similarly reduced the need for the costly shipment of coin or bullion from place to place.

Of course, payment and remittance by IOU was not without cost itself. Payment in bank required both payer and payee to be present, since the order to pay generally had to be given to the banker orally. And the banker had to register the credit and the

debit in his books by hand: a dozen such transactions kept him busy all day. Remittance by IOU similarly involved time and paperwork. However, unlike the cost of using currency or bullion, the cost of payment and remittance by IOU were largely independent of the sum involved. So the use of IOUs was much less expensive for large transactions.

Coin still had a role in small transactions, especially between strangers, and it remained important as an ultimate medium of settlement for the system of payment by IOU. As a result it retained its role as the standard of value and unit of account. And there remained a significant demand for coin and bullion as stores of value.

The Saving in Resources and the Stimulus to Trade in Bullion

Means of payment and remittance are "money." So the cost of providing Europe with money was greatly reduced by the use of IOUs. If all payments and remittance had had to be executed in coin or bullion, it would have required a far larger amount of both. This would have increased the demand for bullion in Europe and so raised its value there, leading to more investment and more effort in prospecting and mining. The use of IOUs for payment and remittance therefore reduced not only the cost of individual transactions but also the social cost of the system.

As a result of the substitution of IOUs for bullion, bullion – especially silver – remained cheaper in Europe than it was elsewhere. It was therefore profitable for Europe to export large quantities of silver in exchange for goods from the Islamic world and Asia. So, instead of having bullion tied up as money, Europe traded it for foreign goods.

This effect was strengthened in the sixteenth century, of course, when Europe began to import even cheaper American silver and re-export it to Asia. The profitability of this arbitrage stimulated interzone trade, with all the growth-enhancing benefits of market expansion that this implied.

The Facilitation of Multilateral Trade

Remittance by IOU facilitated multilateral trade. In its absence, trade between two places would have had to be balanced bilaterally either in goods or through the shipment of bullion. One place would have been able to buy from another more than it sold to it only by shipping bullion to cover the difference. Remittance by IOU made it possible to offset an excess of imports from one place with an excess of exports to another: that is, it made it possible to net trade flows. As a result, trade needed to be balanced, not bilaterally between each pair of places, but only multilaterally between each place and all other places.

For example, Lucca was an important manufacturer of silk cloth, which it exported to Champagne. To produce the cloth, Lucca imported the raw silk it needed through Genoa. Lucchese merchants buying raw silk in Genoa could cover their purchases there by writing IOUs payable in Champagne. When these merchants sold the raw silk in Lucca, they could remit the proceeds to Champagne to cover their IOUs. They did so by purchasing IOUs payable in Champagne from merchants who exported silk cloth to the Fairs and therefore had funds available in Champagne from the sale of the cloth (Blomquist 1990). In the absence of remittance by IOU, Lucca

would have had to bring home the proceeds of its exports to Champagne in bullion and to pay for its imports from Genoa by shipping bullion to that city. Remittance by IOU eliminated the need for both shipments of bullion.

The Impact on the Organization of Commerce

This reduction in transactions costs affected patterns of trade and the extent of the market, but it also affected the organization – and so the productivity – of commerce.

In the absence of remittance by IOU, each individual merchant would have had to export goods to a distant market equal in value to those he imported from there or to cover the difference by shipping bullion – a costly option. Remittance by IOU removed the requirement that the trade of each merchant be balanced in this way by making it possible to net across merchants. One merchant could import goods without exporting anything while another could export goods without importing anything – and nobody had to ship any bullion.

This possibility made it much easier for merchants to specialize – both in particular sorts of goods and in particular trading destinations. For example, some Lucchese merchants became experts in buying raw silk and knew the market in Genoa. Others became experts in selling silk cloth and knew the market at Champagne. By facilitating this sort of division of labor and specialization, the use of bills of exchange increased the productivity of commerce.

The Contribution to the Development of Finance

In several different ways, the development of the system of payments and remittance contributed significantly to the development of finance. The IOUs that served as means of payment and remittance were still IOUs; as such, they themselves played an important role in the financing of commerce. Also, the liquidity facilities that arose to serve the system of payments and remittance expanded to become more general sources of financing. And a workable system of payments and remittance was necessary for the efficient functioning of the system of financial markets in Europe.

Means of Remittance and Payment as Short-Term Loans

Means of remittance were, by their nature, simultaneously instruments of financing. Funds were delivered in one place to be repaid later somewhere else: the delay made the transaction a loan. Such loans became an important source of working capital for merchants. For example, returning to our earlier example, Lucchese merchants who purchased raw silk in Genoa financed their working capital by writing IOUs payable at Champagne.

As we have seen, the use of bills of exchange in remittance generally involved the participation of a merchant bank. When merchants' banks took funds in return for a bill of exchange, they were borrowing; this borrowing enabled them to lend – delivering the funds to someone else in return for their bill of exchange or making

some other type of loan. In this way, merchant banks became important financial intermediaries.

Deposit banks were also financial intermediaries: they borrowed in their own name – by accepting deposits – and re-lent to others. As we have seen, overdraft lending grew naturally out of their payments function, and banks also invested in commercial ventures. Banks also made noncommercial loans to princes, nobles, municipalities, craftsmen, petty traders, and even peasants. Such noncommercial loans were generally secured. The most common security was a pawn – in the case of princes and nobles usually of jewelry or plate; the king of England once borrowed from bankers in Bruges by pawning his crown jewels. Total lending by deposit banks was, however, on a much smaller scale than lending by merchant banks. And, as we have seen, it was not entirely a good thing, since it tended to undermine bank stability.

The Development of Liquidity Facilities into General Sources of Financing

The broadening of bank lending is an example of a liquidity facility, overdraft, expanding to became a more general source of financing. This happened too, more consequentially, with another liquidity facility – the market for interfair deposits at the Fairs of Champagne.

This market started out as an adjunct to the settlement process, but before long most loans were unrelated to settlement. During the thirteenth century, the borrowing and lending that surrounded settlement developed into a general money market – a market for short-term loans. Not only traders at the Fairs, but others seeking short-term credit or having funds they wished to lend short term, came to the Fairs to borrow or lend.

The maturity of an interfair deposit was typically to the following fair – say 3 months. However, the maturity was sometimes "for two fairs." And, in any case, it was easy to roll the loan over – usually at the new market rate. Later fairs, including Lyons and the Brabant fairs at Antwerp, imitated Champagne in their use of the interfair deposit. In their case too, markets for interfair deposits developed into general markets for short-term loans.

The market for bills of exchange began as a market for instruments of remittance, and, as we have seen, it also served as a liquidity facility for the settlement of those instruments. But it, too, evolved into a general money market. Merchants could raise funds by selling bills; merchants and nonmerchants alike could invest their funds by purchasing bills.

The use of the bill of exchange as a pure instrument of credit was soon facilitated by a practice known as "dry exchange" (de Roover 1974, Ch. 4; Mueller 1997, Ch. 8). This combined two bills of exchange, back to back, to create a loan in the same place. For example, a merchant in Venice would borrow there by writing a bill payable in London. When the bill was presented to his representative in London for payment, the representative would borrow the necessary funds to settle by writing a bill to be paid by the merchant himself in Venice. If, as was the custom, each bill was payable after 3 months, the round trip to London created a 6-month loan *in Venice*

(it was purely the great distance from Venice to London that made London an important center of Venetian exchange (Mueller 1997, Ch. 8)). If a longer loan was required, a merchant could extend the term further by rolling the amount over for one or more additional "returns."

Trading Centers Become Financial Centers

Before long, it became common even for debts contracted elsewhere – from London to Palermo – to be made payable at a particular fair. This was convenient for several reasons. Fairs possessed efficient arrangements for settlement, and it was easy to remit funds to them or from them. Also, if the debtor was temporarily short of funds, he could refinance his loan by borrowing in the market for interfair deposits or in the exchange market. Similarly, if the creditor did not have immediate use for the funds repaid to him, he could easily lend them again short term in one of these same markets.

In these ways, the facilities that developed in trading centers for payment and for remittance by IOU lowered the cost of financial transactions as well the cost of trade transactions. So, growing out of their function as trading centers, the various fairs also became important financial centers. And they often continued to function as financial centers even after their importance as trading centers declined.

Integration of Financial Markets

As financial markets developed, bills of exchange were used increasingly to remit funds for financial as well as for commercial reasons. In particular, merchant bankers used them to transfer funds to where they could earn the highest return. For example, the Bruges correspondent of Francesco Datini of Prato wrote to him on April 26, 1399: "It appears that there is an abundance of specie in Genoa; so do not send our money to Genoa, or only if you can get a very good price for it; put it rather in Venice or Florence, or here in Bruges or in Paris or Montpellier." (Braudel 1984).

A merchant banker could go a step further. Rather than merely directing funds to where rates were highest, he could borrow funds where rates were low (by being a taker on bills) and lend them where rates were high (by being a deliverer) (de Roover 1948, Ch. 4). This practice helped to integrate financial markets across Europe and thereby to equalize interest rates.

It also generated an enormous volume of transactions in bills of exchange. Indeed, as early as the fourteenth century, most bills of exchange were related to finance rather than trade. For example, between 1336 and 1340, the Covoni company of Florence registered 443 exchange transactions: of these, only 70 were trade-related, while the remaining 373 were financial (of the financial bills, 335 were "speculative," to exploit interest-rate differentials, 159 were from Florence to Venice and 176 were from Venice to Florence). The remaining 38 involved "dry exchange" (Mueller 1997, Ch 8). Because of the increasing predominance of financial bills, the volume of exchange soon exceeded the volume of trade by a wide margin.

In all of these ways, the development by commerce of a system of payments and remittance using IOUs rather than bullion and coin facilitated the financial development of preindustrial Europe and, more generally, contributed to its overall economic progress.

References

Abulafia D (1997) The impact of Italian banking in the late middle ages and renaissance, 1300–1500. In: Teichova A, van Hentenryk GK, Ziegler D (eds) Banking, trade, and industry: Europe, America, and Asia from the thirteenth to the twentieth centuries. Cambridge University Press, Cambridge, pp 17–34

Ashtor E (1983) Levant trade in the later middle ages. Princeton University Press, Princeton

Barber M (1994) The new knighthood: a history of the order of the temple. Cambridge University Press, Cambridge/New York

Blockmans W (1992) Bruges, a European trading center. In: Vermeersch V (ed) Bruges and Europe. Fonds Mercator, Antwerp

Blomquist TW (1990) Some observations on early foreign exchange banking based upon new evidence from thirteenth century Lucca. J Eur Econ Hist 19(2):353–375

Boyer-Xambeu M-T et al (1994) Private Money & Public Currencies: the 16th century challenge. M.E. Sharpe, Armonk

Braudel F (1984) The perspective of the world. Harper & Row, New York

de Roover R (1944) Early accounting problems of foreign exchange. Account Rev 19:381–407

de Roover R (1948) Money, banking and credit in mediaeval Bruges. The Mediaeval Academy of America, Cambridge

de Roover R (1954) New interpretations of the history of banking. J World Hist 2:38–76

de Roover R (1974) Business, banking, and economic thought. University of Chicago Press, Chicago

Farmer D (1991) Marketing the produce of the countryside, 1200–1500. In: Miller E (ed) The Agrarian history of England and Wales, vol III. Cambridge University Press, London, pp 1348–1500

Grafe R (2001) Northern Spain between the Iberian and Atlantic worlds: Trade and regional specialization, 1550–1650. In: Economic History. London School of Economics and Political Science, London

Lane FC (1937) Venetian bankers, 1496-1533: a study in the early stages of deposit banking. J Polit Econ 45:187–206

Mueller RC (1997) The venetian money market: banks, panics, and the public debt 1200–1500. Johns Hopkins University Press, Baltimore

Nicholas D (1979) The English trade at Bruges in the last years of Edward III. J Mediev Hist 5:23–61

Nicholas D (1992) Medieval Flanders. Longman, London

Parker G (1977) The emergence of modern finance in Europe 1500–1730. In: Cipolla CM (ed) The Fontana Economic History of Europe. Volume 2: The Sixteenth and Seventeenth Centuries. Harvester Press, New York, pp 527–594

Pressnell LS (1956) Country banking in the industrial revolution. Clarendon Press, Oxford

Spufford P (1988) Money and its uses in medieval Europe. Cambridge University Press, Cambridge

Usher AP (1934) The origins of banking: the primitive banks of deposit 1200–1600. Econ Hist Rev 4:399

van der Borght R (1896) The history of banking in the Netherlands. In: A history of banking in all the Leading nations, vol IV. The Journal of Commerce and Commercial Bulletin, New York, pp 191–371

Van der Wee H (1963) The growth of the Antwerp market and the European economy. Martinus Nijhoff, The Hague
Van der Wee H (1977) Monetary, credit and banking systems. The Cambridge economic history of Europe. In: Rich EE, Wilson CH (eds) The Economic Organization of Early Modern Europe, vol V. Cambridge University Press, Cambridge, pp 290–393
Van der Wee H (1993) The low countries in the early modern world. Variorum, Aldershot
van Dillen JG (1934) The Bank of Amsterdam. In: van Dillen JG (ed) History of the Principal Public Banks. Martinus Nijhof, The Hague, pp 79–123

Money Markets and Exchange Rates in Preindustrial Europe

10

Pilar Nogues-Marco

Contents

Abstract

This chapter focuses on money markets and exchange rates in preindustrial Europe. The foreign exchange market was mostly based on bills of exchange, the instrument used to transfer money, and provides credit between distant centers in preindustrial Europe. In this chapter, first I explain bill of exchange operations, money market integration, usury regulations, and circumventions to hide the market interest rate, as well as the evolution of bills of exchange in history, focusing mainly on the most relevant features generalized during the first half of the seventeenth century: endorsement and the joint liability rule, which facilitated the full expansion of the foreign exchange market beyond personal networks. Then, I describe the European geography of money in the mid-eighteenth century, characterized by a very high degree of multilateralism with the triangle of Amsterdam, London, and Paris as the backbone of the European settlement system. Finally, I measure the cost of capital and relate it to liquidity. I show evidence of interest rates in the eighteenth century for Amsterdam, London, Paris, and Cadiz. While Amsterdam, London, and Paris presented low and similar

P. Nogues-Marco (✉)
University of Geneva and CEPR, Geneva, Switzerland
e-mail: Pilar.NoguesMarco@unige.ch

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_16

interest rates, Cadiz had higher interest rates, mostly being double the cost of capital. These results seem to show a high inverse correlation between liquidity and interest rates, suggesting that the share in international trade of European centers might have been a powerful driver of international monetary leadership. While more empirical evidence and further research are needed, this approach opens the scope of the analysis beyond the national institutional explanation.

Keywords
Money market · Bills of exchange · Monetary geography · Usury regulations · Cost of capital

Introduction

This chapter focuses on money markets and exchange rates in preindustrial Europe. The foreign exchange market was mostly based on bills of exchange, the instrument used to transfer money, and provide credit between distant centers in preindustrial Europe. In section "Bills of Exchange: Transfer and Credit Instrument" I define bill of exchange operations in order to highlight the two functions of transfer and credit involved in bill of exchange transactions. The function of transfer strengthened trade relations because bills of exchange facilitated cashless payments, as I explain in the second section. The function of credit developed due to the hidden interest rate embedded in the exchange rate at maturity that circumvented usury regulations on credit, as I describe in the third section.

European money markets became connected by virtue of utilizing bills of exchange, which had been created for the purpose of transferring money between cities without having to ship specie. Section "European Money Market Integration" measures European money market integration in the eighteenth century. Because shipping precious metals was costly, the exchange rates normally fluctuated within the "specie point," that is, the ratio of the bullion market prices in two cities plus/minus the cost of shipping bullion. An efficient and integrated money market would suffer only few and nonpersistent "breaks" of the "specie points." In Medieval and Early Modern Europe, the well-functioning of the specie-point mechanism was conditioned by the foreign exchange policy, which distorted market integration. However, by the early eighteenth century, core countries had liberalized bullion movements, which permitted market integration. The measurement of the specie-point mechanism shows that London and Amsterdam money markets were highly integrated in the mid-eighteenth century. This result is consistent with our knowledge of London-Amsterdam securities market integration in the eighteenth century. Nevertheless, more research is needed to improve our knowledge on money market integration for a broader sample of European centers.

Section "Usury Regulation and Circumventions in Commercial Finance" explains how the credit function of bills of exchange developed due to the hidden interest rate embedded in the exchange rate at maturity since it circumvented usury

regulations on credit. Usury regulation influenced the development of capital markets in the Middle Ages and the Early Modern period by defining a ceiling on interest rates. To conceal usurious interest rates in local loans, the borrower sometimes acknowledged a debt larger than the sum actually received. Other times official documents just recorded the amount of capital to be repaid, so interest rates are unknown. As a consequence of usury regulations, local interest rates cannot be found, or, if they can be found, they are only a biased measure of the opportunity cost of capital. Bills of exchange, however, constitute the benchmark to calculate interest rates because this instrument circumvented usury regulations. Church doctrine accepted that the price charged on bills of exchange was motivated by the risks and efforts associated with overcoming the obstacles of the foreign settlement. In practice, the price of bills of exchange was determined by the geographical distance as well as by time, so the interest rate was hidden by the exchange rate at maturity.

Section "Bills of Exchange: Endorsement and the Joint-Liability Rule" focused on the most relevant features of bills of exchange generalized during the first half of the seventeenth century: endorsement and the joint liability rule. Endorsement made bills of exchange a common negotiable instrument because the joint liability provided high financial protection against nonpayment as all parties involved in a bill transaction (the payer, all the endorsers, and the drawer) had a joint liability for the payment. As a consequence, endorsement facilitated an expansion of commercial-financial activities beyond personal networks by supporting long-distance transactions which involved quasi-impersonal relations between parties. The emergence of a liquid European market for bills of exchange organized along lines defined by trading relations provided the infrastructure for financial development.

In section "The Geography of Money Before the Industrial Revolution," I describe the European geography of money in the mid-eighteenth century, characterized by a very high degree of multilateralism with the triangle of Amsterdam, London, and Paris as the backbone of the European settlement system. This result underlines the critical importance of multilateral commercial finance as opposed to bilateral trade relations in shaping foreign exchange transactions. More research is needed to map the geography of monetary relations before the eighteenth century in order to understand the dynamics of change in commercial finance geographical relations during the Early Modern period.

Finally, in section "Measuring the Cost of Capital in the Eighteenth Century Europe" I measure the cost of capital and relate it to liquidity. I show evidence of interest rates in the eighteenth century for Amsterdam, London, Paris, and Cadiz. While Amsterdam, London, and Paris presented low and similar interest rates, Cadiz had higher interest rates, mostly being double the cost of capital. These results seem to show a high inverse correlation between liquidity and interest rates, suggesting that the share in international trade of European centers might have been a powerful driver of international monetary leadership. While more empirical evidence and further research are needed, this approach opens the scope of the analysis beyond the national institutional explanation.

Bills of Exchange: Transfer and Credit Instrument

A bill of exchange was a letter by which one merchant ordered his correspondent in some other city to make a payment on his behalf to another merchant in that distant city. It was the European merchants' instrument of cashless payment and a means of providing credit between the Middle Ages and the twentieth century. Thus, the European money and financial market was historically built on the bill of exchange.

Figure 1 explains bill of exchange operations. Four persons participated in the transaction: two at the city where the bill was drawn and two at the city where it was payable. First, there was the drawee (deliverer, giver, remitter, or negotiator), the person who delivered the money to the drawer in exchange for the bill; second, the drawer (taker) who received the money by selling the bill to the drawee; third, the party who paid the money at the destination point by virtue of the bill drawn on him, commonly termed the payer (accepter); and fourth, the person to whom the bill was made payable at the destination point, called the payee (possessor or holder). Typically the payer and the drawer would keep accounts with each other and could offset the payment of the bill with claims from other transactions so that only small net amounts would have to be occasionally settled (Beawes 1773; De Roover 1953; Einzig 1962; Mueller 1997; Neal 1990; Nogues-Marco 2011; Jobst and Nogues-Marco 2013).

Suppose that an English agent needs to settle a debt in Amsterdam and he does not have his own correspondent there. This English agent (drawee) will buy a bill of exchange from a merchant in London who has a Dutch correspondent (drawer), paying out in local currency, sterling pounds (£). The bill is drawn in the foreign currency, *shellinge bank* (β) to be paid out in Amsterdam. The drawee will remit by post the bill to the payee in Amsterdam to settle his debt. The payee will show the bill to the payer for acceptance, and the payer will pay the bill to the payee at maturity. If the bill was not accepted or if the accepted bill was not paid, the payee would protest the bill and the drawer would be obliged for the payment of the bill plus the charges of protest, postage, commissions, and brokerage.

The exchange rate is defined by the ratio between the quantity of sterling pounds paid for the purchase of the bill in London and the quantity of *shellinge bank* paid in Amsterdam for the payment of the bill. The operation of exchange entails indeed two operations: the transfer of money from London to Amsterdam and the lending involved between the purchase of the bills in London and its payment in Amsterdam. Thus, the exchange rate at maturity is comprised of two components: a shadow spot exchange rate for the transfer and a shadow interest rate for the loan. This interest rate is determined by the interest rate at the destination point in accordance with the origin center. In our example (Fig. 1), it would be the interest rate in Amsterdam. To understand why, just think about the following arbitrage operation: A time bill in London on Amsterdam is equivalent to a spot transfer of funds from London to Amsterdam at the current exchange rate and a subsequent deposit of the *shellinge* thus obtained at the Amsterdam interest rate (Flandreau et al. 2009a).

In the next sections, we study the two functions of transfer and credit of bills of exchange. The function of transfer strengthened trade relations because bills of

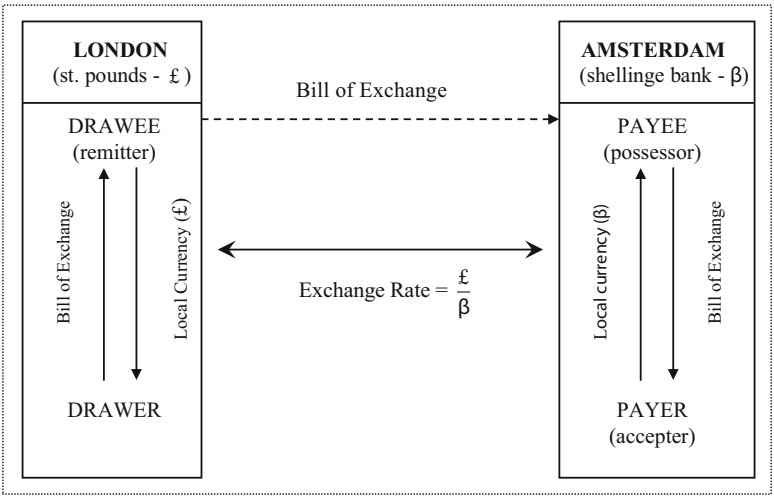


Fig. 1 The bill of exchange

exchange facilitated cashless payments. The function of credit developed due to the hidden interest rate embedded in the exchange rate at maturity that circumvented usury regulations on credit.

European Money Market Integration

European money markets became connected by virtue of utilizing bills of exchange, which had been created for the purpose of transferring money between cities without having to ship specie. Bills offered several advantages in relation to precious metals: First, shipping precious metals had a high risk of shipwreck, piracy, and/or brigandage. Second, bills avoided transportation costs of gold or silver, and finally, bills eliminated the cost of minting coins at the destination point, that is, cost associated with specie shipment except for those few coins that enjoyed international acceptance (Munro 1979: 172–173).

Because shipping precious metals was costly, the exchange rates normally fluctuated within the "specie point," that is, the ratio of the bullion market prices (gold or silver) in two cities plus/minus the cost of shipping bullion between the two cities (Morgenstern 1959; Officer 1996; Flandreau 2004; Nogues-Marco 2013). As a consequence, agents would prefer to transfer bills than specie, except when the exchange rates "broke" the "specie point" so it was more profitable to ship specie instead of bills. Suppose that an English agent needs to settle a debt in Amsterdam. Gold and silver, along with bills of exchange, may be used to settle international payments. Normally, the best way to settle the debt was to buy a bill of exchange in London on Amsterdam, provided enough such bills were available. But if bills were scarce, their price would rise. If the bill price increased above the level at which it

became preferable to send metal than bills as payment, the English debtor would buy gold or silver on the London market and ship it to Amsterdam.

The exchange rate must lie within the "specie points" to avoid metal shipments and, therefore, eliminate the risks and costs of transferring specie to cash payments. A well-functioning and integrated money market would suffer only few and non-persistent "breaks" of the "specie points."

In Medieval and Early Modern Europe, the well-functioning of the specie-point mechanism was conditioned by the foreign exchange policy, which included several measures that distorted market integration, such as bans or licenses on the export of bullion, fixed exchange rates for foreign coins, bans on unauthorized exchange transactions, taxes on foreign exchange transactions, debasements, official foreign exchange operations, etc. Governments were interested in influencing foreign exchange in order to achieve goals such as retaining or attracting bullion, preventing or reversing a rise in domestic prices resulting from a depreciation of the exchange rate, safeguarding the interests of national industries by stimulating exports or handicapping imports, keeping down or reducing the cost of government remittances abroad, or causing difficulties to hostile governments (Einzig 1962: 155–168).

The implementation of foreign exchange policies that distorted market integration decreased during the eighteenth century because the spirit of laissez-faire was gaining ground. Indeed, many of the foreign exchange policies used during the sixteenth and seventeenth centuries became obsolete in the early eighteenth century (Einzig 1962: 168). European countries liberalized bullion movements at the end of the seventeenth century, which permitted money market integration: Dutch law had allowed ingots and foreign coin exports since 1646. Dutch mint masters attempted to ban bullion exports, but mercantile circles in Amsterdam opposed them and prevailed in keeping bullion exports open. England's bullion trade was liberalized in 1663. At that time, bars and foreign coins could be exported, although this was not possible for English coins up to 1819 (Attman 1983: 27; Gillard 2004: 132; Viner 1955: 4; Munro 1979: 212). To the end of the eighteenth century, only Spain and Portugal, the empires that produced the vast majority of precious metals, maintained bullion export prohibitions (Pallavicino 1855: 8; Larruga 1787–1800: (3) 49–57; Nogues-Marco 2010).

Once bullion movements were free and money markets integrated, shipping bullion was only occasionally profitable, and bills of exchange were normally used for payments. Money market integration supported commercial exchanges between European centers because an efficient functioning of the "specie-point mechanism" facilitated the use of bills of exchange, so eliminated the transaction costs associated with the shipping of specie.

Measuring money market integration in the Early Modern period is complicated because systematic data on bullion prices are difficult to find in primary sources before the nineteenth century. Nevertheless, from the end of the seventeenth century, London registered prices on bullion in the financial press *The Course of the Exchange*. Additionally, Amsterdam bullion prices are available for the eighteenth century in the commercial bulletin *Kours van Koopmanschappen tot Amsterdam*. Using these

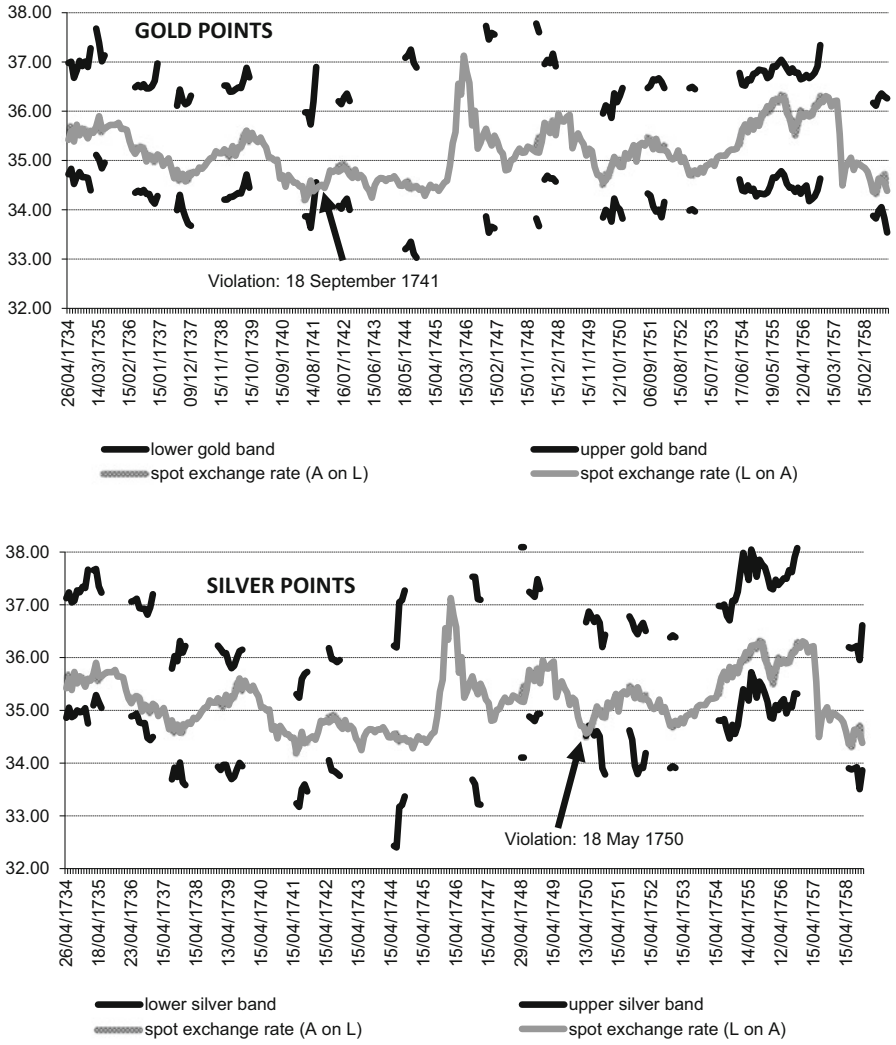


Fig. 2 Specie-point mechanism (1734–1758), London and Amsterdam, shellinge bank/sterling pound (monthly observations). (Source: Nogues-Marco 2013: 452–453)

data, Nogues-Marco (2013) measured money market integration between London and Amsterdam in the mid-eighteenth century. Figure 2 shows results.

Figure 2 demonstrates the highly integrated money market between London and Amsterdam in the mid-eighteenth century. These results are consistent with our knowledge of London and/or Amsterdam securities market integration in the eighteenth century (Neal 1990, 2000; Nogues-Marco and Vam Malle 2007; Koudijs 2011). While integration for the cases of London and/or Amsterdam is well known,

more research is needed to improve our knowledge on money market integration for other European centers.

Usury Regulation and Circumventions in Commercial Finance

Usury regulation influenced the development of capital markets. In the Middle Ages, usury was defined as a ban on lending money at interest, that is, anything beyond the principal of a loan was deemed usury (De Roover 1967: 258). According to Koyama (2010), usury regulation endured centuries because it created a barrier to entry that enabled secular rulers, the Church and a small number of merchant-bankers to earn monopolistic rents. This argument suggests that the prohibition restricted lending to a small group of merchants who were able to write contracts sufficiently complex to evade the Church's laws. The rents generated by the usury prohibition were appropriated by medieval rulers, elite merchant-bankers, and the Church.

To conceal usurious interest rates in local loans, the borrower acknowledged a debt larger than the sum actually received. Usury suits provide evidence of this practice. Two examples can be obtained from the *Archivio di Stato* (Pisa 1230): In the first suit, the borrower stated that the lender had forced him to appear in front of a notary and swear that he had borrowed £26 when in fact he had received but £20. In the second suit, the borrower claimed that he was forced to stipulate that he had received £7 in order to borrow only £4 (Blomquist 1971: 469). Many other cases illustrate that contemporaries used the same device to camouflage interests, that is, merchants recorded a price for the amount of capital to be repaid at a level that would incorporate a hidden interest rate. For example, local merchants of Prato used a parallel account to keep track of large loans with a full description of all the conditions: In 1385, the tailor Domenico di Jacopo recorded in his accounting books that he had received a loan of 70 florins, while the lender had given him a written receipt for a loan in the amount of 90 florins. Similarly, in 1333 in *Santa Maria dell'Impruneta*, a moneylender confessed in his will that he was owed 90 ½ florins and 37 *staia* of grain but had lent only 68 florins and no grain. In 1402, a wool manufacturer made a loan of 20 florins but had it notarized as 22 florins. In 1450, Francesco di Matteo Castellani signed an agreement in a banker's book to repay a loan of 110 florins when – according to his own accounts – he had in fact borrowed only 100 florins (Marshall 1999: 98–99).

Medieval usury regulation did not distinguish between money lent to poor people or to businessmen who intended to invest it in a profitable venture. Calvin (1509–1564) was the first who made such a distinction between business loans on which it was all right to take interest and distress loans which should be made free of charge (De Roover 1967: 258; Kerridge 2002: 30–34). For business loans, the ban was replaced in some countries by a legal maximum rate that prevented interest rates from raising this legal ceiling, which would constitute "usury." The legal ceiling declined in the long run, suggesting an improvement of efficiency in European capital markets. For example, in the case of England, Henri VIII defined the legal ceiling at 10% in 1545, but Edouard VI reintroduced the ban on lending at interest in

1552. Queen Elisabeth accepted again the ceiling at 10% in 1571 due to merchants' interests, despite the opposition of the scholastic doctors and the Church. The ceiling was reduced to 8% in 1624, 6% in 1661 and 5% in 1714 (Nelson 1949: 83; De Roover 1953: 125).

The practice of concealing interest rates remained until the end of the Old Regime as the consequence of the usury regulation. In France, while certain debt instruments and short-term loans among merchants permitted interest rates if they did not exceed the usury ceiling, the ban on interest rates for long-term loans endured until the French Revolution (Hoffman et al. 2000: 14). Luckett (1992: 29) explains that the annual yield of perpetual annuities was limited by law to 5% and, therefore, notarized acts of the sale of rentes thus invariably list the purchase price as exactly 20 times the coupon. According to him, historians have frequently accepted this evidence at face value, asserting that the long-term rate of interest was unusually stable under the Old Regime. But he considers that the formal purchase price of rentes found in notarized contracts was nothing but a legal fiction, an amount that the buyer pretended to pay and that the seller pretended to receive when they signed the act. Mercier (1782–1788: (2) 36) considered that "when [the notary] then writes 'currency counted and delivered', this is most often a fiction." Similarly, Hoffman et al. (2000: 15–16) explains the ways by which the Early Modern notaries hid the interest rates on loans. For instance, obligations (promissory notes) could not openly specify the payment of interest, but evidence shows that it was de facto paid. Suppose, for example, that Monsieur Martin lent 1000 livres to Baron du Pont for a year at 10% interest. The contract might hide the interest rates if it did not register the loan actually received but specified only the amount of repayment. In our example, the contract might stipulate that du Pont had to repay 1100 livres in a year's time. In the late eighteenth century, notaries developed payment schedules that joined interest payments and capital repayment in obligations. Suppose, for example, that Martin lent du Pont 1000 livres to be repaid in 5 equal annual installments at an interest rate of 10%. The notary might draw up a contract stating that du Pont had borrowed 1320 livres, which he would repay at the rate of 264 livres a year.

Therefore, the mechanism to hide interest rates remained similar from the Middle Ages to the end of the Old Regime. These mechanisms were mainly based on recording in the official documents only the amount of capital to be repaid. As a consequence of usury regulation, local interest rates cannot be found, or, if they can be found, they are only a biased measure of the opportunity cost of capital.

Bills of exchange, however, constitute the benchmark to calculate interest rates because this instrument circumvented usury regulations. Unlike other financial instruments, which had a local circulation and were thus subject to the regulation of interest rates, bills of exchange escaped usury ceilings because the price charged on bills of exchange was motivated, according to contemporary bankers, by the risks and efforts associated with overcoming the obstacles of foreign settlement. As we have seen, in practice, the price of bills was determined by the geographical distance as well as by time, so the interest rate was hidden by the exchange rate at maturity (Beawes 1773; De Roover 1967; Flandreau et al. 2009a; Malynes 1601).

Indeed, we can distinguish between "real exchanges," that is, the plain vanilla bills of exchange drawn to cash real mercantile transactions between distant cities, and "fictitious exchanges," which were credit instrument based on derivatives of bills of exchange used in Medieval and Early Modern Europe to lend locally to yield a return that did not consider usury constraints. One way of doing so was called "dry exchange" (*cambium siccum*) that worked as follows: merchant *A* went to local banker *B* in London to borrow money; banker *B* wanted to charge an usurious interest rate, so he gave the money to merchant *A* in exchange for a bill drawn by merchant *A* to an imaginary person settled in Amsterdam at *x* number of months, according to the maturity of the loan. When it reached maturity, the bill protested for nonpayment came from Amsterdam, so merchant *A* must repay to banker *B* in London for the money he had borrowed. An alternative way to lend locally at usurious interest rates was "exchange and reexchange" (*cambium et recambium*): merchant *A* went to local banker *B* in London to borrow money, and the banker B gave the money to the merchant *A* in exchange for a bill drawn by the merchant *A* payable in Amsterdam (exchange operation). When the bill was presented for payment at maturity in Amsterdam, the payer *C* (merchant's *A* agent there) settled the bill by sending a second bill in Amsterdam on London for payment to the banker *B*, the original lender in London (re-exchange operation). Using these two bills back to back, the first bill (exchange operation) created a local loan in London to be repaid in London when the second bill (reexchange operation) returned for payment. The operation was covered because participants had agreed in advance on the price of the return bill (fictitious exchange), so the operation was bound by arbitrage to yield the same market return as a local loan (Beawes 1773; De Roover 1944; Denzel 2010; Flandreau et al. 2009a; Hayes 1719–1777; Jobst and Nogues-Marco 2013; Neal 1990).

Bills of exchange became a safe and cheap financial instrument from late Medieval Europe, rather than simply a means of cashless payments because exchange transactions provided wealthy lenders with a way of making profit while circumventing the religious interest ban. Plain vanilla bills of exchange were used to arbitrate for profit by buying and selling bills in different European cities at different exchange rates. Additionally, fictitious exchange operations (such as dry exchange or exchange and reexchange) facilitated the use of bills of exchange as a legal financial instrument to hide market interest rates. Borrowers were willing to pay a premium, that is, the interest rate hidden in the exchange rates, to have access to capital. According to available empirical evidence, the volume of financial bills surpassed the volume of bills used for commercial purposes already in the fourteenth to fifteenth centuries. For instance, the ledger of the Covoni Family Company (Florence, 1336–1340) registered 443 exchange operations: 70 purely commercial remittances from Venice to Florence, covering liabilities incurred; 335 speculative remittances (159 from Florence to Venice, 179 from Venice to Florence); and 38 dry exchange contracts (Mueller 1997: 317–318).

The expansion of the financial use of bills of exchange in late Medieval Europe gave businessmen the incentive to establish organizations capable of spreading the use of bills of exchange from relatively small and personal networks to a broader

European semi-impersonal system. According to Rubin (2010), Italian bankers established interregional branches in all the major financial centers of Europe to take advantage of exchange rate differences and capital scarcity (thus implicitly lending at interest) while at the same time diversified portfolios to shield against risk. Calculations of the return of exchange and reexchange operations between Bruges, London, and Venice (c. 1437–1481) indicate rates ranging from 15.6% to 20.9% and rising as the length of the loan rose from 1 to 6 months (Neal 2015: 37).

The extension of the credit network achieved by the branching system allowed for semi-impersonal credit relations to arise because from the point of view of the headquarters in Italy, most financial activities were conducted with impersonal relations in Europe (Rubin 2010). In late fourteenth and fifteenth centuries, a dense network of bank branches and corresponding merchant-bankers linked the main European trade and finance centers. For instance, Datini (1335–1410) established a bank in Prato with branches in Florence, Pisa, Avignon, Barcelona, Valencia, and Palma de Mallorca. Although Datini's bank had no branch in Bruges, he was represented there by several correspondents (De Roover 1948a: 30, 55–56). Similarly, the Medici bank established in Florence (1397–1494) opened branches in many European centers, such as Ancona, Avignon, Basel, Bruges, Geneva, London, Pisa, Rome, and Venice (De Roover 1948b: 13–23, 1963: 67–69). Quasi-impersonal credit relations consolidated during the Early Modern period with the adoption of endorsement and the joint liability rule, as we are going to see in next section.

Bills of Exchange: Endorsement and the Joint Liability Rule

European trade and finance supported the emergence of a liquid market for bills of exchange, which was organized along lines defined by trading relations and provided the infrastructure of financial development. Endorsement facilitated the expansion of bills because it made them a common negotiable instrument from the early seventeenth century (Usher 1914; De Roover 1953). Endorsement means that the holder (endorser) could transfer the bill to another person (endorsee) by signing the back of the document. Endorsement implied a joint liability, that is, all parties involved in a bill transaction (the payer, all the endorsers, and the drawer) were legally responsible in solidum for the payment (joint liability rule).

The adoption of endorsement is difficult to trace. Usher (1914: 576) considered that the practice began in Italy in the second half of the sixteenth century and spread northward along the commercial routes. Lapeyre (1956: 296) and De Roover (1953: 101) claimed that endorsement already existed in sixteenth century, according to the empirical evidence obtained in *Archivo provincial de Valladolid* (Castilian bills of exchange endorsed in 1575) and *Archivio di stato di Firenze* (Italian bills of exchange already endorsed in 1519), although endorsement seems to have been only occasional in the sixteenth century. Evidence of bills of exchange endorsed in northwestern Europe dates from the late sixteenth century. The first bill of exchange discovered for Antwerp is from 1571, although endorsement was not generally

practiced in Antwerp until the first decades of the seventeenth century (Van der Wee 1963: (II) 349).

The legal origin of the joint liability rule that supported the expansion of endorsement is also difficult to trace. According to Van der Wee (1963: (II) 337–348, 1977: 325–329) and Puttevils (2015), bills of exchange adopted the Antwerp legal principle of assignment when they were introduced in the Antwerp market in the sixteenth century. From the Middle Ages, merchants in Antwerp had used the instrument "letter obligatory" (promissory note) to postpone payments: A merchant who underwrote one committed himself to paying a given sum within a given period. The medieval letter obligatory used quite frequently the clause "payable to bearer," although it did not facilitate their circulation from hand to hand because of the lack of legal guarantees in transferability. To be able to take a defaulting debtor to court, the bearer of a letter obligatory, even if it carried a bearer's clause, had first either to obtain an explicit authority from the original creditor – an authority which incidentally could be revoked at any time – or to have an official transfer by means of a formal cession. But in 1507, a judicial verdict in Antwerp granted the bearer of letter obligatory the same rights as the original creditor with regard to the prosecution of an insolvent debtor. This principle was confirmed by Imperial Edict for the whole of the Netherlands in 1537. Additionally, financial protection for the bearer improved: The courts had tended to regard all transfers of letters obligatory as cessions that relieved the transferring creditor of all responsibility. But this practice changed by 1532 when assignment of letter obligatory had become familiar practice. In 1541, an Imperial Edict recognized the principle of assignment for transferred letter obligatory, that is, the legal right of the bearer to assign the collection of the debt from the original signatory of the letter to the assigning debtor (i.e., the ceding creditor). This principle added a new important guarantee for the circulation from hand to hand of letter obligatory, as the assigning debtor was also bound until the payment had really been made. Once this principle had been recognized for the first transfer, it had logically to apply to the ensuing transfers as well, with the proviso however that only the last assigning debtor was held. When the use of bills of exchange expanded in Antwerp by influence of the more important firms of Southern German and Italian merchant-bankers, the clause "payable to bearer" was naturally introduced in bills of exchange, so the ordinance of 1541 recognized the principle of assignment covered both letter obligatory and bills of exchange. However, the increasing degree of anonymity created problems in terms of identifying who the latest assigned debtor was – because in letter obligatory only the original creditor and debtor were named in the document itself. As a consequence, endorsement emerged to identify the last debtor by the signature within the bill of exchange, and the liability was no longer limited to the latter but extended to all previous endorsers, that is, the principle of the joint liability rule (see also Hunt and Murray 1999: 212–213; Munro 1991).

The joint liability attached to endorsement played a major role in ameliorating fundamental information problems in long-distance trade and finance, so facilitated the expansion of the foreign exchange market beyond personal networks. Santarosa (2015) has empirically examined the role of the joint liability rule in

reducing asymmetric information in long-distance trade. First, negotiable bills of exchange contained a threat of adverse selection caused by the incentive of sellers to pass on the bills of riskier debtors. The joint liability rule diminished adverse selection because it made the drawer, all the endorsers, as well as the payer legally responsible for payment. Second, bills of exchange also included a standard moral hazard problem in agency relations between correspondents. On the one hand, the drawer had to deal with clearing risk, that is, the payer refusing to pay his bill. On the other hand, the payer had to deal with the risk of the drawer's account being overdrawn. The joint liability rule induced endorsers to participate in the enforcement of agency relationships because it encouraged endorsers to monitor the payers. An increase in the number of endorsers improved monitoring and thereby reduced the chance of default. As a consequence, the joint liability rule permitted bills of exchange to expand commercial-financial activities by supporting long-distance transactions which involved quasi-impersonal relations between parties.

The joint liability rule turned the bill of exchange into the most important instrument of transfer and short-term credit in the Early Modern period. The geography of the market for bills of exchange indicates the density of commercial finance transactions in Europe. The next section explains the geography of money before the industrial revolution.

The Geography of Money Before the Industrial Revolution

In a joint research, Flandreau et al. (2009b) mapped the geography of money in the mid-eighteenth century. We constructed a systematic map of European intercity monetary relations to study the contours of foreign exchange market linkages in Europe before the industrial revolution, whose main results are summarized in this section.

Why do we consider intercity monetary relations? The European monetary linkages in the Early Modern period were not international linkages but intercity linkages. The modern national monetary systems emerged in Europe in the nineteenth and first half of the twentieth centuries, as part of the process of nation-state building, through the establishment of national issuing banks and the development of branch-banking transfers (Helleiner 2003). Such changes consolidated national payment systems based on paper currencies circulating within national territories, which reduced the cost of moving money across domestic locations practically to zero. But before the process of money nationalization, money moved between cities – within the same country or between countries – according to the "specie-point mechanism," as we have seen in previous section. Therefore, the relevant unit of analysis for documenting monetary relations was not the country but the individual city, seen as a "node" in a "network" of inter-cities monetary linkages.

Why do we focus on Europe? The market for bills of exchange was restricted to Europe until the late eighteenth century. We documented the bills-of-exchange connections between nearly 80 cities in the mid-eighteenth century. A high

proportion of those cities (close to one half) were ports. There were typically several markets per country, except for the case of England that stands out as the one large political entity with only one exchange center in London. Locations were evenly scattered all over Europe with outreaches on the fringes of the Orient. There were no American, Asian, or African cities, while only two cities in the Ottoman Empire, Constantinople and Smyrna. This European bias is the result of a structural characteristic of the foreign exchange network: European sources did not direct to non-European centers, while non-European sources directed back to Europe. The global financial system of that period had a distinct European focus.

Why do we analyze the mid-eighteenth century? European intercity monetary linkages are measured by the existence (or not) of exchange rate quotations, that is, city *X* quotes city *Y*. Our primary sources are the traditional handwritten exchange rates reported in bankers' correspondence together with semi-printed and printed foreign exchange bulletins whose use expanded in the eighteenth century (McCusker and Gravesteijn 1991). The large sample avoids biased results. This research has provided us with the opportunity to identify Europe's monetary geography on the eve of the Industrial Revolution. However, more research is needed to map the geography of monetary relations before the eighteenth century in order to understand the dynamics of change in commercial finance geographical relations during the Early Modern period.

A crucial characteristic of foreign exchange bulletins – handwritten, semi-printed, or published bulletins – is that certain centers are quoted but not others. Flandreau et al. (2009b) argue that the existence or not of a price quotation gives a reasonably good indication for the liquidity of the underlying market because it reveals the existence of a sufficiently large demand and supply to warrant the posting of exchange rates.

We are likely to observe the development of liquid, well-organized foreign exchange markets, where not only commercial and financial intercourse is intense, but also other bankers have already established connections, as that which occurs in agglomeration economies. The existence or not of quotations gives a strong indication of liquid markets in international financial claims and the existence of financial linkages between financial centers. Therefore, collecting lists of foreign centers quoted allows sketching the geography of commercial finance.

Figure 3 shows the ranking of cities according to the number of quotes they received differentiating according to the listed versus quoted criterion. We identified a population of 78 centers. We found exchange rate bulletins for 64 of them. According to the available sample, Amsterdam was quoted almost everywhere (54 out of the 64 possible markets, or 84%), London was quoted in 46 markets (72%), and Paris was quoted in 39 markets (61%), implying that multilateral settlement using Amsterdam, London, or Paris as a clearing center was definitely feasible by the mid-eighteenth century. Another feature is the dominance of Northwestern European financial centers (Amsterdam, London, and Paris; and also Hamburg) along with the continued relevance of cities in the Southern/Mediterranean arena. Leghorn and Genoa (and also Venice) were the most relevant centers of the Mediterranean zone.

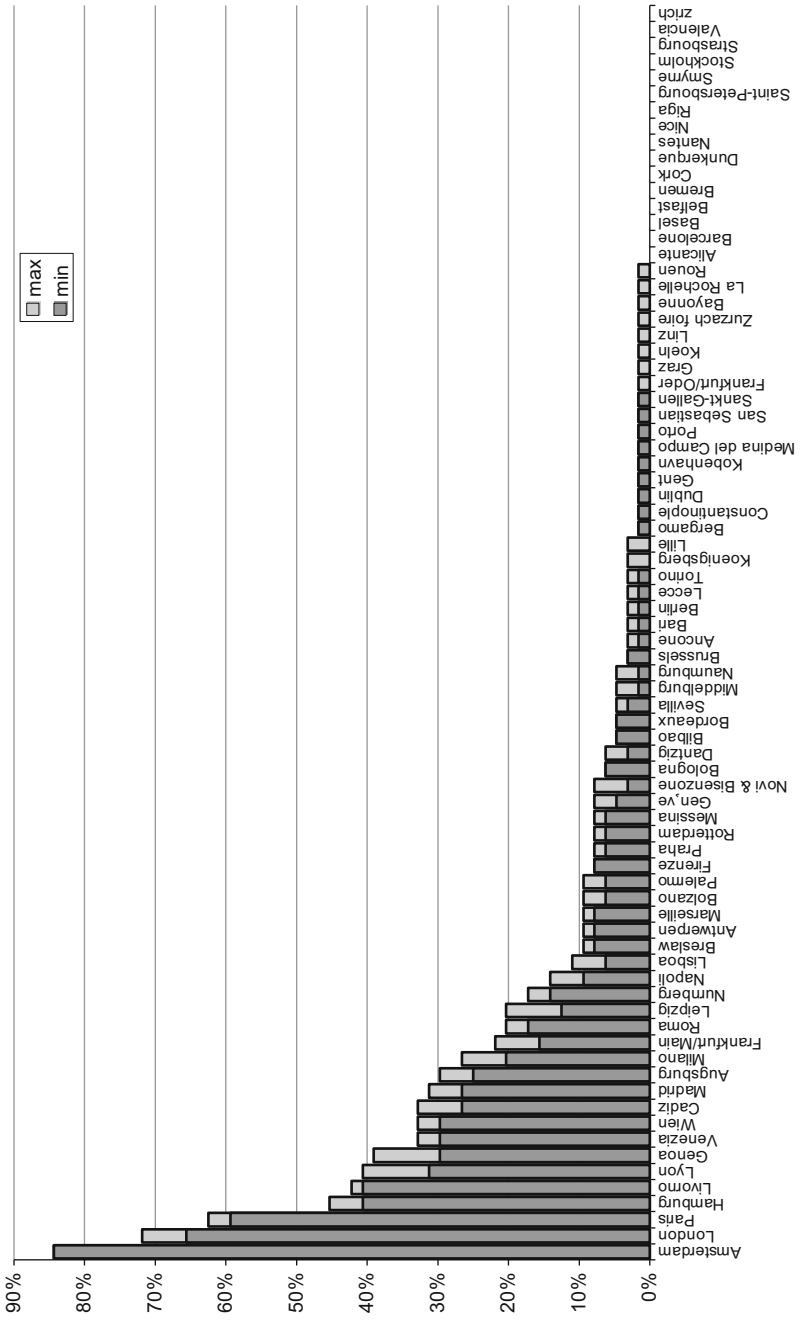


Fig. 3 Monetary popularity in the mid-eighteenth century. (Source: Flandreau et al. 2009b: 160). Min and max correspond to the intersection/reunion of four networks created as the narrow/broad listed/quoted databases generated from several primary sources: printed commercial and/or financial bulletins, semi-printed financial bulletins, and hand-written financial bulletins

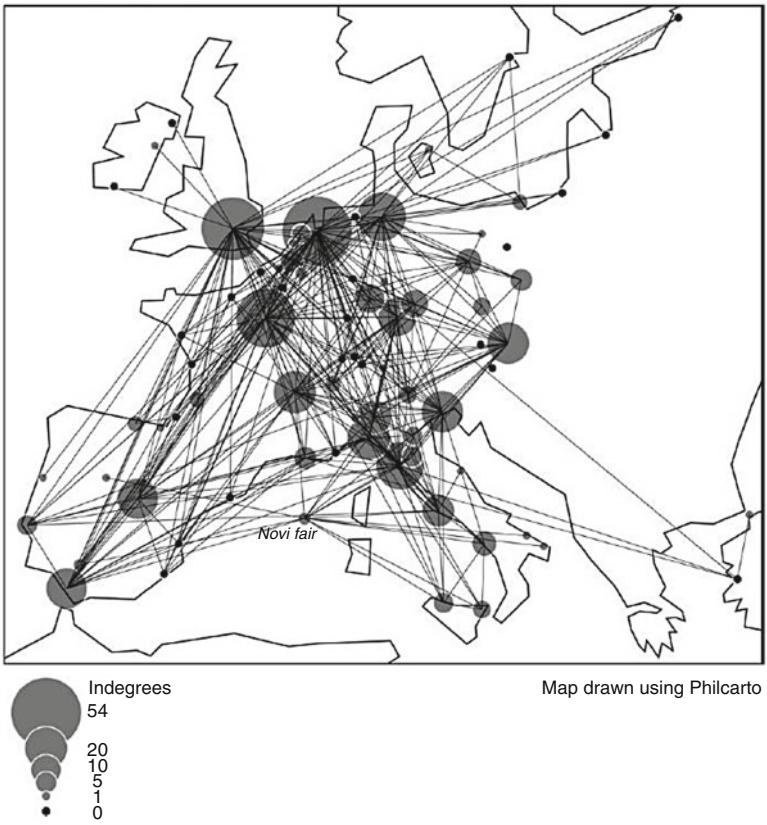


Fig. 4 Monetary agglomeration in the mid-eighteenth century. (Source: Flandreau et al. 2009b: 161)

Graphically, the information can be depicted in a map, where a line between two cities indicates the existence of a liquid market for mutual claims. These linkages are directed, i.e., they might go from center *A* to center *B* (*A* quotes *B*), from center *B* to center *A* (*B* quotes *A*), or both ways (*A* quotes *B* and *B* quotes *A*). The sum of all links gives the network. Figure 4 shows monetary agglomeration: It plots the number of links that a particular market received by the mid-eighteenth century. About 20% of links between cities were direct, 75% had to pass through an intermediary center, while only 7–8% had two intermediaries. This reinforces the notion of an encompassing multilateral settlement system with Amsterdam, London, Paris, and also Hamburg and Genoa as the main connecting hubs. Therefore, the European system was a dense web with an area of intense financial linkages that stretched over Amsterdam-London-Paris-Hamburg and shrunk as it headed toward Italy. This area overlaps with Brunet's current Blue Banana area of modern economic prosperity (Brunet 2002).

While the network appears dense, only a small share of between 11% and 13% of the total number of possible links is active. In fact, the eighteenth-century system

exhibited a high degree of concentration, comparable to what Flandreau and Jobst (2005) obtained for the later, country-based, network of the late nineteenth century. This result underlines the critical importance of *multilateral* commercial finance as opposed to *bilateral* trade relations in shaping foreign exchange transactions.

To sum up, the geography that emerges from the network analysis shows evidence of a two-part system. One was the older Mediterranean system, now revolving around Genoa and Leghorn. The other was the newer "Northern Atlantic" system with Amsterdam, London, and Paris on top. Amsterdam was the leading clearing center for multilateral settlements and the entrepôt for commercial credit throughout Europe (Neal 2000: 121–122), but Dutch currency did not possess a natural monopoly that had crowded out the other international currencies. On contrary, the contact between North Atlantic and Mediterranean systems was guaranteed by the pivotal roles of Amsterdam, London, and Paris. It was also achieved through the agency of older European continental financial centers that had been the cradle of the Commercial Revolution.

Measuring the Cost of Capital in the Eighteenth Century Europe

We have seen in previous sections that the bill of exchange was the instrument that sustained European commercial finance. International trade supported the making of a liquid international market for trade bills of exchange, which as a result tended to be organized along lines defined by trading relations. The global trading network turned out to provide the infrastructure of financial development. The eighteenth-century European money markets were integrated and Europe's monetary geography was seamless owing to the existence of bills of exchange, which were widely traded. In turn, the market for commercial bills was the benchmark money market of the eighteenth century, which had grown outside the reach of usury legislation, through regulatory arrangements in which merchants were able to prevail.

The commercial bills market was the benchmark money market of the eighteenth century, so the commercial interest rate was the benchmark interest rate. But systematic direct evidence on commercial interest rates in the eighteenth century is difficult to find. Whichever financial center we are looking at, there are no recorded series of "money market" rates for the eighteenth century. This is consistent with the fact that we are dealing with an over-the-counter market. In order for "one" price to be recorded and quoted, a formal centralized market must be organized. This requirement was not met by the credit markets of the time, since interest rates resulted from bilateral drawing arrangements that were in turn put to work as a lever for operating on the foreign exchange market. Formalization and centralization prevailed in the foreign exchange market, not in the money market. As a result, a precise notion of the "general interest rate," meaning probably the typical conditions that the best merchant-bankers in a center would extend to their correspondent in another center, must have existed as a kind of "mental average" in the mind of contemporary practitioners but was nowhere physically quoted.

Interest rate collectors have tended to be eclectic in their choice of sources, using mainly bankers' archives – although not exclusively – as illustrated by Homer and Sylla (2005). An indirect way in line with direct evidence to know the commercial interest rate for the eighteenth century is to estimate the shadow interest rate comprised in the bills of exchange. Suppose that we know the price for a foreign bill bought in a given market A and drawn on another market B where it matures at a certain future date (a_{AB}). Suppose next that we also know the price for a "similar" bill, bought in market A and payable in market B and involving the same risks and returns, but hypothetically maturing today – spot bill (x_{AB}). It is obvious that there is a relation between the price of the first and the second bill that involves the interest rate for the maturity period for a commercial loan in center B according to center A between today and the maturity period (r_B^A) (Officer 1996):

$$x_{AB} = a_{AB} \cdot \left(1 + r_B^A\right)(\text{currency } A/\text{currency } B) \tag{1}$$

Scholars have used variants of this formula to calculate the shadow interest rate. Perkins (1978) provided the interest rate in London from New York, 1835–1900. Schubert (1989) estimated the interest rate in Amsterdam and Paris from London, 1731–1795. Boyer-Xambeu et al. (2001) calculated the interest rate in Paris from London and in London from Paris, 1833–1873. Flandreau et al. (2009a) gave the interest rate in Amsterdam from London, in London from Amsterdam, and in Paris from London, 1720–1789. Nogues-Marco (2011) measured the interest rate in Cadiz from London, 1729–1788, and, finally, Edvinsson (2011) made the interest rate available in Stockholm, 1660–1685.

According to Flandreau et al. (2009a), it is possible to calculate the shadow interest rate for the main centers in the eighteenth century (Amsterdam, London and Paris) because these centers quoted bills of exchange at two maturities, sight and 2 months. For example, let us calculate the shadow interest rate in Amsterdam (denoted as A) according to London (denoted as L), supposing that we know the sight and 2 months exchange rate in London on Amsterdam. The long maturity exchange rate (a_{LA} [n_l days]) and the short maturity exchange rate (a_{LA} [n_s days]) can be rewritten according to Eq. (1) as:

$$a_{LA}[n_l] = x_{LA} / \left(1 + r_A^L \cdot \frac{n_l}{365}\right)(\text{sterling pound}/\text{schellingbank}) \tag{2}$$

$$a_{LA}[n_s] = x_{LA} / \left(1 + r_A^L \cdot \frac{n_s}{365}\right)(\text{sterling pound}/\text{schellingbank}). \tag{3}$$

Substituting for x_{LA} in the Eqs. 2 and 3 gives the arbitrage condition that derives shadow interest rates:

$$r_A^L = \frac{(a_{LA}[n_s] - a_{LA}[n_l]) \cdot 365}{(a_{LA}[n_l] \cdot n_l - a_{LA}[n_s] \cdot n_s)} \tag{4}$$

However, this methodology is limited to the centers whose exchange rates quoted at two maturities, which are mainly the core centers, such as Amsterdam, London, and Paris for the eighteenth century. We need to explore other methodologies for those noncore centers whose exchange rates quoted only at the longer maturity. Nogues-Marco (2011) proposed a method that derives an implicit spot exchange rate assuming that market participants had an idea of the interest rate that was used in correspondent markets. The intuition is the following: suppose that the interest rate in a given market B is known for bankers located in center A. Then it is possible to derive a "shadow" spot exchange rate in center A on center B, provided that A quotes B at certain maturity (Eq. 1). Assuming that the spot exchange rate in A on B is, by arbitrage, essentially identical to the spot exchange rate in B on A, it is possible to construct the interest rate in A if we know the exchange rate at maturity in center B on center A (Eq. 1).

In other words, suppose that people in market A can form an estimate of interest rate in center B from A, ($\hat{r}_B^A$). This estimate is indexed by A reflecting that it is a market A's opinion of the price of credit in market B. Of course, there can be as many such estimates as there are markets where people must form an opinion of credit conditions in market B. If agents are reasonable enough, we can assume that on average, they will guess things adequately so that:

$$\hat{r}_B^A = E\left(r_B^A | I_A\right) \tag{5}$$

This is the first assumption we make. The $E\left(r_B^A | I_A\right)$ operator indicates the expectation conditional upon the information available in market A, denoted as I_A. Plugging this value into Eq. (1) yields an estimate of the "shadow" spot exchange rate in market A on market B or

$$\tilde{x}_{AB} = E(x_{AB} | I_A) = x_{AB} \cdot \left(1 + \hat{r}_B^A\right) \tag{6}$$

Now suppose further that the "shadow" exchange rate of A on B is identically the shadow exchange rate of B on A so that we can write:

$$\tilde{x}_{AB} = \tilde{x}_{BA} \tag{7}$$

This is the second assumption we make. At one level it is a simple arbitrage condition. But in practice, since there are delays in information delivery and transaction costs, "cross" spot exchange rates, when they exist, are not necessarily the same and assuming that shadow cross spot exchange rates are identical is not innocuous. A priori, we may surmise that the validity of this assumption is

influenced by the degree of development of money markets, the efficiency of arbitrage and information technology, and the quality of expectations of what is happening in other markets. Nogues-Marco (2013: 467) notes that cross spot exchange rates between Amsterdam and London in mid-eighteenth century display only tiny differences.

Conditional on this assumption being a valid one, we can plug our estimates of the shadow price of a bill on A traded in B and combine it with the known price of a time bill on A traded in B to generate a measure of the interest rate in center A. It is the following:

$$\hat{r}_A^B = (\tilde{x}_{BA}/a_{BA}) - 1 \tag{8}$$

Figure 5 shows the long-run behavior of the three commercial interest rate series computed for Amsterdam (from London), London (from Amsterdam), and Paris (from London) according to Eq. 4 (Flandreau et al. 2009a) and the interest rate series in Cadiz (from London) according to Eq. 8 (Nogues-Marco 2011). Differentials between Amsterdam, London, and Paris remain small throughout, especially for the Amsterdam-London pair. Paris interest rates were slightly higher – say, between 4% and 5% when London and Amsterdam were between 3% and 4.5% – but the salient fact is that differences across core countries are not large and actually disappear toward the end of the century.

A different result appears for the case of Cadiz that shows high interest rates in comparison with core countries. While average interest rate for the whole period, 1720–1789, is 3.92% for Amsterdam, 3.8% for London, and 4.59% for Paris, it is 8.79% for the case of Cadiz. These results are consistent with the limited direct evidence that we have for the four centers. In the case of Amsterdam and Paris, we can compare our estimated interest rates with some empirical evidence on commercial rates, as well as VOC overdrafts at the Bank of Amsterdam for the case of Amsterdam. Similarly, we have some evidence of Bank of England's discount rate for the case of England (Flandreau et al. 2009a: 187–188). For the case of Cadiz, the discount rate in Cadiz was 8% in 1786, according the evidence available in the archive of the Bank of San Carlos (Tedde 1988: 131). In all cases, the direct evidence is consistent with our estimations.

These results challenge the view given by North and Weingast (1989) about the relevance of credible institutions as a requisite for financial development. It is well worth to note the crucial importance of Paris as an international center. This is at odds with the traditional emphasis on the inadequacy of the constitutional underpinnings of France's political regime. Understanding better the underlying mechanisms and the reason why Paris-based commercial paper could prosper despite the financial difficulties of the French crown would go a long way toward getting a clearer view of the degree to which agglomeration economies can substitute for "sound" institutional infrastructure.

Figure 4 in the previous section has shown monetary agglomeration in mid-eighteenth century. Amsterdam, London, and Paris were the most liquid centers

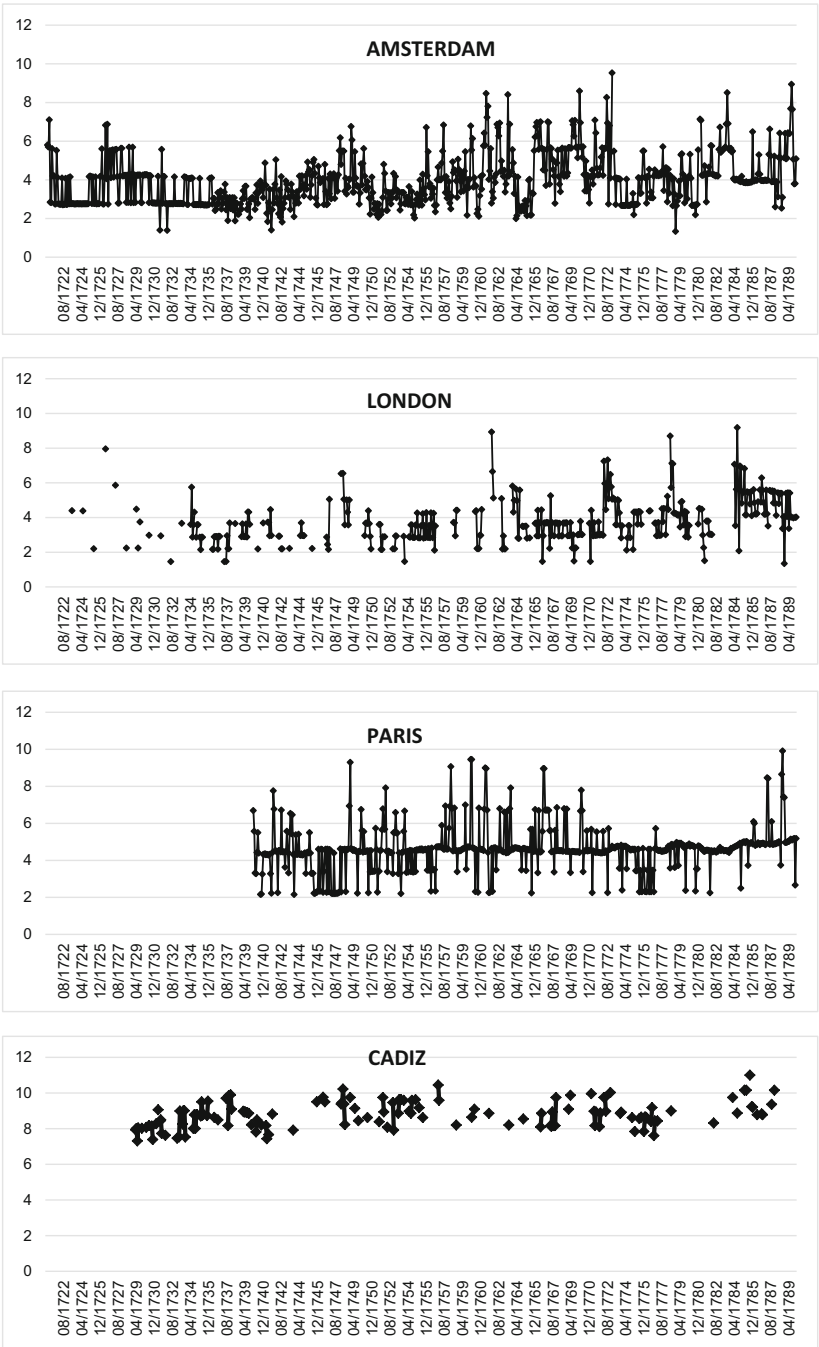


Fig. 5 Commercial annual interest rates (%), 1720–1789. (Source: Amsterdam, London and Paris in Flandreau et al. 2009a: 187–188. Cadiz in Nogues-Marco (2011: 65–92). Outliers have been removed)

at that time. According to Flandreau and Jobst (2009: 653) and Eichengreen et al. (2018: 101), the larger the number of foreign quotations a currency received, the lower its interest rates. That is, interest rates of leading currencies were lower because lots of agents were using them, so that their currencies were more likely to be quoted abroad and this further strengthened their leadership.

According to Kindleberger (1967), size is the main driver of currency leadership. Flandreau and Jobst (2009) tested the empirical determinants of international currency status in 1900 and found that size measured by the share in international trade is a powerful driver of international monetary leadership. While the drivers of international monetary leaderships in the Early Modern period have not been studied yet, size would also explain currency leadership in the eighteenth century. If we approach the share in international trade by the merchant fleet, England, France, and Holland had the highest shares of the European merchant fleet circa 1790 – being 26.15%, 21.62%, and 11.79%, respectively – while Spain had one of the lowest, being only the 4.43% (Romano 1962: 578). Size would explain leadership and low interest rates in core centers (Amsterdam, London and Paris), and high interest rates in secondary centers (Cadiz). While more empirical evidence and further research is needed, this approach opens the scope of the analysis beyond the national institutional explanation.

References

Attman A (1983) Dutch enterprise in the world bullion trade: 1550–1800. Kungl. Vetenskaps – och Vitterhets-Samhället, Göteborg

Beawes W (1773) Lex maercatoria rediviva: or, the merchant's directory, 6th edn. Printed for James Williams, London

Blomquist TW (1971) The Castracani family of thirteenth-century Lucca. Speculum 46(3):459–476

Boyer-Xambeu M-T, Deleplace G, Gillard L (2001) L'intégration des marchés monétaires au XIXe siècle. Les places financières de Paris et de Londres (1833–1873). Hist Mesure XVI(1/2):113–155

Brunet R (2002) Lignes de force de l'espace européen. Mappemonde 66(2):14–19

De Roover RA (1944) What is dry exchange? A contribution to the study of English mercantilism. J Polit Econ 52(3):250–266

De Roover RA (1948a) Money, credit and banking in medieval Bruges. The Mediaeval Academy of America, Cambridge, MA

De Roover RA (1948b) The Medici Bank. Its organization, management, operations, and decline. New York University Press/Oxford University Press, New York/London

De Roover RA (1953) L'évolution de la lettre de change. XIVème XVIIIème siècles. Armand Colin, Paris

De Roover RA (1963) The rise and decline of the Medici Bank, 1397–1494. Harvard University Press, Cambridge, MA

De Roover RA (1967) The scholastics, usury, and foreign exchange. Bus Hist Rev 41(3):257–271

Denzel MA (2010) Handbook of world exchange rates, 1590–1914. Ashgate, Farnham

Edvinsson R (2011) Shadow interest rates in Stockholm and the integration of the early financial markets, 1660–1685: was Heckscher right? Financ Hist Rev 18(3):309–329

Eichengreen B, Mehl A, Chitu L (2018) How global currencies work. Past, present and future. Princeton University Press, Princeton/Oxford

Einzig P (1962) The history of foreign exchange. Macmillan, London

Flandreau M (2004) The glitter of gold. France, bimetallism, and the emergence of the international gold standard, 1848–1873. Oxford University Press, Oxford
Flandreau M, Jobst C (2005) The ties that divide: a network analysis of the international monetary system 1890–1910. J Econ Hist 65(4):977–1007
Flandreau M, Jobst C (2009) The empirics of international currencies: network externalities, history and persistence. Econ J 119(537):643–664
Flandreau M, Galimard C, Jobst C, Nogues-Marco P (2009a) The bell jar: commercial interest rates between two revolutions, 1688–1789. In: Atack J, Neal L (eds) The origins and development of financial markets and institutions. From the seventeenth century to the present. Cambridge University Press, Cambridge, pp 161–208
Flandreau M, Galimard C, Jobst C, Nogues-Marco P (2009b) Monetary geography before the industrial revolution. Camb J Reg Econ Soc 2(2):149–171
Gillard L (2004) La Banque d'Amsterdam et le Florin Européen au temps de la République néerlandaise (1610–1820). Éditions de l'École des Hautes Études en Sciences Sociales, Paris
Hayes R (1719–1777) The Negociator magazine of monies and exchanges. London: printed for R. Hayes (1st, 1719); printed for J. Noon (2nd, 1724); printed for W. Meadows (3rd, 1730); printed for J. Noon (4th, 1739); printed for J. Noon (6th, 1740); printed for J. Noon (8th, 1754); printed for G. Keith (9th, 1764); printed for G. Keith (11th, 1777)
Helleiner E (2003) The making of national money. Territorial currencies in historical perspective. Cornell University Press, Ithaca/London
Hoffman PT, Postel-Vinay G, Rosenthal JL (2000) Priceless markets. The political economy of credit in Paris, 1660–1870. The University of Chicago Press, Chicago/London
Homer S, Sylla R (2005) A history of interest rates. Rutgers University Press, Hoboken
Hunt ES, Murray JM (1999) A history of business in medieval Europe, 1200–1550. Cambridge University Press, Cambridge
Jobst C, Nogues-Marco P (2013) Commercial finance in Europe, 1700–1815. In: Caprio G (ed) Handbooks in financial globalization: handbook of key global financial markets, institutions and infrastructure. Elsevier, London, pp 95–105
Kerridge E (2002) Usury, interest and the reformation. Ashgate, Aldershot
Kindleberger C (1967) The politics of international money and world language. Essays in international finance, vol 61. Princeton University Press, Princeton
Koudijs PAE (2011) Trading and financial market efficiency in eighteenth-century Holland. PhD Dissertation, Universitat Pompeu Fabra
Koyama M (2010) Evading the "taint of usury": the usury prohibition as a barrier to entry. Explor Econ Hist 47:420–442
Lapeyre H (1956) La banque, les changes et le crédit au XVIe siècle. Rev Hist Mod Contemp 3(4):284–297
Larruga E (1787–1800) Memorias políticas y económicas sobre los frutos, comercio, fábricas y minas, con inclusión de los reales decretos, órdenes, aranceles y cédulas expedidas para su gobierno y fomento. Imprenta de B. Cano, Madrid. 45 volumes
Luckett TM (1992) Credit and commercial society in France, 1740–1789. PhD dissertation, Princeton University
Malynes G (1601) A treatise on the canker of England's commonwealth. Printed by Richard Field for William Iohnes printer, London
Marshall K (1999) The local merchants of Prato: small entrepreneurs in the Late Medieval economy. The Johns Hopkins University Press, Baltimore/London
McCusker JJ, Gravejstijn C (1991) The beginnings of commercial and financial journalism. NEHA, Amsterdam
Mercier LS (1782–1788) Tableau de Paris. Amsterdam
Morgenstern O (1959) The international solidarity of money markets. In: International financial transactions and business cycles. Princeton University Press, Princeton, pp 166–279
Mueller R (1997) The venetian money market: banks, panics, and the public debt, 1200–1500. Johns Hopkins University Press, Baltimore

Munro JH (1979) Bullionism and the bill of exchange in England, 1272–1663: a study in monetary management and popular prejudice. In: The dawn of modern banking (conference 1977, Los Angeles). Yale University Press, New Haven/London, pp 169–239
Munro JH (1991) The international law merchant and the evolution of negotiable credit in late-medieval England and the low countries. In: Banchi pubblici, banchi privati e monti di pietà nell'Europa preindustiale: atti del convegno 1–6 ott. 1990, vol 1. Società ligure di storia patria, Genoa, pp 49–80
Neal L (1990) The rise of financial capitalism. International capital markets in the age of reason. Cambridge University Press, Cambridge
Neal L (2000) How it all began: the monetary and financial architecture of Europe during the first global capital markets, 1648–1815. Financ Hist Rev 7(2):117–140
Neal L (2015) A concise history of international finance. From Babylon to Bernanke. Cambridge University Press, Cambridge
Nelson B (1949) The idea of usury. From tribal brotherhood to universal otherhood. Princeton University Press, Princeton
Nogues-Marco P (2010) Bullionism, specie-point mechanism and bullion flows in the early 18th-century Europe. PhD Dissertation, Sciences-Po Paris
Nogues-Marco P (2011) Tipos de cambio y tipos de interés en Cádiz en el siglo XVIII (1729–1788). Banco de España – Estudios de Historia Económica 56, Madrid
Nogues-Marco P (2013) Competing bimetallic ratios: Amsterdam, London and Bullion Arbitrage in mid-18th century. J Econ Hist 73(2):445–476
Nogues-Marco P, Vam Malle-Sabouret C (2007) East India bonds, 1718–1763: early exotic derivatives and London market efficiency. Eur Rev Econ Hist 11(3):367–394
North DC, Weingast B (1989) Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century Britain. J Econ Hist 49(4):803–832
Officer LH (1996) Between the dollar-sterling gold points: exchange rates, parity, and market behaviour. Cambridge University Press, Cambridge
Pallavicino de Grimaldi MC (1855) A Legislação monetaria em Portugal. Typographia do Progresso, Lisboa
Perkins EJ (1978) Foreign interest rates in American financial markets: a revised series of dollar-sterling exchange rates. J Econ Hist 38(2):392–417
Puttevils J (2015) Tweaking financial instruments: bills obligatory in sixteenth-century Antwerp. Financ Hist Rev 22(3):337–361
Romano R (1962) Per una valutazione della flotta mercantile europea alla fine del secolo XVIII. In: Studi in Onore di Amintore Fanfani vol. V (Evi moderno e contemporaneo). Giuffrè, Milan, pp 575–591
Rubin J (2010) Bills of exchange, interest bans, and impersonal exchange in Islam and Christianity. Explor Econ Hist 47:213–227
Santarosa VA (2015) Financing long-distance trade: the joint liability rule and bills of exchange in eighteenth-century France. J Econ Hist 75(3):690–719
Schubert ES (1989) Arbitrage in the foreign exchange markets of London and Amsterdam during the 18th century. Explor Econ Hist 26(1):1–20
Tedde P (1988) El Banco de San Carlos. Alianza Editorial, Madrid
Usher AP (1914) The origin of the bill of exchange. J Polit Econ 22(6):566–576
Van der Wee H (1963) The growth of the Antwerp market and the European economy, fourteenth-sixteenth centuries. Martinus Nijhoff, The Hague
Van der Wee H (1977) Monetary, credit and banking system. In: Rich EE, Wilson CH (eds) The Cambridge economic history of Europe. The economic organization of early modern Europe. Cambridge University Press, Cambridge, pp 290–392
Viner J (1955) Studies in the theory of international trade. George Allen & Unwin Ltd., London

International Money Markets: Eurocurrencies

11

Stefano Battilossi

Contents

S. Battilossi (✉)
Department of Social Sciences, Universidad Carlos III de Madrid, Madrid, Spain
e-mail: battilos@clio.uc3m.es

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_57

Abstract

Eurocurrencies are international markets for short-term wholesale bank deposits and loans. They emerged in Western Europe in the late 1950s and rapidly reached a global scale. A Eurocurrency is a form of bank money: an unsecured short-term bank debt denominated in a currency (for instance, US dollars) but issued by banks operating offshore, in a geographical location or a legal space situated outside of the jurisdiction of the national authorities presiding over that currency (for instance, the Federal Reserve). In Eurocurrency markets, banks intermediate mainly between foreign residents. They borrow funds by "accepting" foreign currency deposits and lend foreign currency-denominated funds by "placing" deposits with other banks, by granting short-term loans or investing in other liquid assets. Historically, Eurodollars accounted for the largest share of Eurocurrencies, although other international currencies (Deutsche Marks, Japanese yens, and especially Euros since 1999) played an important role. Eurocurrency markets were a manifestation of financial integration and interdependence in a globalizing economy and performed critical functions in the distribution and creation of international liquidity. At the same time, their fast growth was a recurrent source of concerns for central bankers and policymakers due to their implications for macroeconomic policies and financial stability. This chapter analyzes different aspects of the historical development of Eurocurrency markets and their role in the international monetary and financial system. The first part discusses theoretical interpretations, presents estimates of markets' size, describes their structure, and explains the determinants of their growth. The second part analyzes the spread between Eurodollar rates and other US money market rates, the role of arbitrage, the evolution of risk factors, and the causes of historical episodes of stress and contagion in the interbank market. The last part discusses political economy issues, such as the role of governments and market forces in the emergence of Eurodollars in the 1950s and the failed attempts to impose multilateral controls on Eurocurrency markets in the 1970s.

Keywords

International banking · Wholesale banking · Money markets · International liquidity · Offshore finance · Liability management · Interest arbitrage

Introduction

Eurocurrencies are a form of bank money: an unsecured short-term bank debt denominated in a currency (for instance, US dollars) but issued by banks operating offshore – that is, in a geographical location or a legal space situated outside of the jurisdiction of the national authorities presiding over that currency (for instance, the Federal Reserve) – in order to fund short-term loans. Eurodollars were the first and

more prominent example of Eurocurrency. The term was coined in the late 1950s, when banks in London and other European financial centers started bidding for dollar liquid balances in the hands of foreign wealth owners (other commercial banks, central banks, official institutions, commercial companies) and used them to fund short-term loans to other foreign banks and nonfinancial companies. At the same time, smaller offshore markets for time deposits denominated in other international currencies emerged; the most relevant examples were the EuroDeutscheMark, EuroSwissFranc and Euroyen markets in London, and the Eurosterling market in Paris. In the 1970s, as banks operating in extra-European financial centers expanded their Eurocurrency (mostly Eurodollar) activities, the market acquired a global dimension, turning into "one of the fastest-growing as well as the most vital and important capitalist institutions" of the twentieth century (Stigum and Crescenzi 2007, p. 209). By the mid-1980s, London accounted for 25% of global Eurocurrency assets, followed by Tokyo (10%), Paris (7%), and offshore financial centers in the Caribbean (Bahamas, Cayman Islands) and the Far East (Singapore and Hong Kong), with a share between 4% and 6% each (Lewis and Davis 1987, pp. 230–231). Since then, the prefix "Euro" survived more as a remnant of the origins of the market, than as a characterization of its geographical scope. Given the historical importance of Eurodollars, their impact on global monetary and credit conditions, and the extensive literature on their origins, development, and implications, this chapter will often make special reference to them.

As a financial product, a Eurocurrency is simply a time deposits (or a certificate of deposit, that is, a negotiable receipt of a deposit) yielding a fixed rate and with maturities ranging from overnight to 6 months. Time deposits are a form of near (or quasi) money: short-term stores of value that cannot be used directly as a medium of exchange to settle debts but can be very easily converted into cash. In spite of their simplicity, Eurocurrencies represented a financial innovation with enormous consequences. Their "essential feature" (Niehans 1984) was the separation between the location of the issuing bank and the currency in which transactions were denominated. This resulted in the unbundling of currency risk from political risk and more generally in the ability to circumvent regulations imposed by national authorities (Lewis and Davis 1987, pp. 217–219 and 269–270). For instance, the most important advantage of booking dollar deposits offshore was to avoid interest rate ceilings, reserve requirements, and deposit insurance fees imposed by US authorities on deposits held with domestic banks. By significantly reducing the costs of bank intermediation, this allowed depositors to yield higher interest rates and borrowers to have access to cheaper short-term loans.

Eurodollars and the other Eurocurrencies were an innovative form of wholesale banking (i.e., transactions with large customers – including other banks – involving large sums), which led to the development of an international money market with a specific microstructure and autonomous sets of interest rates. Since the 1960s, therefore, Eurocurrency markets played a critical role as a channel for the redistribution of international liquidity. Banks' borrowing and lending in the Eurocurrency wholesale market was also conducive to a major structural change in banking business: the marketization of liabilities, pioneered by US commercial banks both domestically and internationally in the 1960s, and subsequently adopted by banks

in both industrialized and developing economies. Liability management – the active management of short-term debt instruments with different rates, maturities, and currency of denomination to match the size and characteristics of asset portfolios – created unprecedented scope for leverage, thus enhancing the fast expansion of banks' balance sheets. However, it also made banks much more vulnerable to currency, liquidity, interest rate, and counterparty risk and facilitated the international transmission of financial shocks (Kane 1979; Lewis and Davis 1987, pp. 81–128; Battilossi 2010).

By drawing on an extensive economic and historical literature, this chapter analyzes different aspects of the historical development of Eurocurrencies and their role in the international monetary and financial system between the late 1950s and the Great Financial Crisis of 2007–2009. The first section "Theory" discusses alternative interpretations based on different approaches to monetary economics. The second section "Scale" presents estimates of the size of Eurodollar and Eurocurrency markets in the long run. The third section "Structure" describes the microstructure of the market and its interbank segment. The fourth section "Growth" discusses the different phases of expansion of the market and their determinants. The fifth section "Arbitrage" analyzes the relationships between rates in the Eurodollar market and other US money markets. The sixth section "Risk" explores how risk in the Eurocurrency markets evolved over time and analyzes historical episodes of severe stress in the interbank market. The seventh section "Political Economy" focuses on the attitude of British authorities in the emergence of Eurodollars and the international debates of the 1970s on the multilateral regulation of Eurocurrencies. A brief summary and final considerations are offered in the last section.

Theory

Multiplier Versus Portfolio

The business of "accepting" (borrowing) and "placing" (lending) Eurodollars emerged in London in the second half of the 1950s as a consequence of a set of favorable circumstances. Early accounts (Einzig 1960, 1964; Holmes and Klopstock 1960; Altman 1961, 1963) focused mainly on supply-side factors, such as the availability of a critical mass of dollar deposits booked with European banks. Some of them were related to Cold War political tensions. In case of national emergencies, the Trading with the Enemy Act of 1917 and its subsequent amendments allowed US presidents to "freeze" or seize assets held in the USA by governments and residents of a foreign country. Therefore, by holding dollar balances with banks in Europe, the Soviet Union and other communist regimes limited their exposure to country risk – that is, the risk of seeing their dollar reserves affected by sanctions such as those imposed by the USA on Communist China and North Korea after the outbreak of the Korean War in 1950 (Coates 2018). Multinational companies, central banks in nonindustrial countries, and international institutions,

such as the Bank for International Settlements (BIS), were also considered as important sources of dollars for European banks. In fact recent research confirmed that the BIS placed dollar deposits with banks in the city in connection with its swap operations with the Bank of England in support of the pound (Yago 2013, pp. 160–163). Specific demand-side shocks were another factor emphasized in the early literature, namely, the restrictions imposed by British authorities during the currency crisis of 1957 on the use of sterling-denominated trade credit for non-sterling area trade, which induced British banks to offer dollar-denominated facilities funded by dollar term deposits. More recent research cast doubts on whether these elements provide a sufficient explanation for the emergence of the market and points to other factors. They include the opening of opportunities for international interest arbitrage between money markets in New York and London by the mid-1950s, thanks to the emergence of interest rate differentials; the gradual return of currencies to external current account convertibility, culminated in 1958; the presence in London of the largest foreign exchange market in Europe and the reopening of forward markets for hedging exchange risk; the reactivation of traditional bank correspondent connections between New York, London, and continental financial centers; a cartelized banking system and extensive domestic and external regulations that encouraged British banks to pursue innovative lines of business; and the acquiescent attitude of British monetary and supervisory authorities, which allowed dollar intermediation to flourish unregulated (Schenk 1998, 2002; Burn 1999, 2006, pp. 99–134; Battilossi 2000, 2002a, b).

The fact that banks freely competed for deposits in foreign currencies was at odds with a financial environment in the USA, UK, and Continental Europe still strongly influenced by the legacy of interwar and war finance, with extensive exchange and capital controls, pervasive regulations of interest rates and binding constraints on the size and composition of banks' portfolios. Initially, analysts and practitioners struggled to grasp how the market worked and what implications it had for the international monetary system and the conduct of national economic policies. The magnitude of this "surprise effect" was such that, more than 20 years after the first transactions had taken place in London, Eurocurrencies "continue[d] to appear to be an enigma even to those who operate in them continually" (Dufey and Giddy 1978, p. 2). In fact, the exponential growth of the market generated lively debates among economists and raised serious concerns among policymakers about its possible consequences for exchange rate stability and the international propagation of inflation. An early source of controversy was its relationship with the US balance-of-payments deficit that emerged at the end of the 1950s and the fast growth of foreign dollar liabilities, both private and official, which in 1960 led Robert Triffin to formulate his "dilemma" (Eichengreen 2006, pp. 116–117; Eichengreen 2011, pp. 50–51; Bordo and McCauley 2017). Some pinned the balance-of-payments hypothesis on the fact that dollar reserves of foreign central banks were an important source of Eurodollar deposits, either directly or indirectly (Klopstock 1970). Others argued that foreign dollar holdings were not a sufficient condition for the growth of the market and that Eurodollars were essentially a monetary phenomenon (Friedman 1971).

Although most subsequent economic analyses underwrote Friedman's view, economists disagreed on its theoretical underpinnings. The main debate opposed supporters of a "multiplier" approach to those of a "portfolio" approach. This divide reflected a more fundamental controversy between the "fractional reserve theory" and the "financial intermediation theory" of banking that emerged in monetary economics in the 1960s and 1970s (Werner 2016). The key intuition of the multiplier approach, as formulated by Friedman (1971) and formalized by Fratianni and Savona (1972), was that the Eurodollar system operated similarly to a domestic banking system of fractional reserves. As a consequence, the expansion of the market was understood as the consequence of a credit or deposit multiplier mechanism, through which a portion of the liquid funds lent by banks to nonbank borrowers was redeposited with other banks in the system, thus generating an endogenous process of credit creation. Critics of the multiplier approach emphasized its inconsistency with some of the observed characteristics of the system – especially the large and increasing role of interbank transactions and its nature of an open system connecting different national financial systems – and the failure of empirical studies to provide robust estimates of the base and magnitude of the multiplier (Machlup 1970; Masera 1972; Crockett 1976; Niehans and Hewson 1976; Dufey and Giddy 1978, pp. 135–154; Mayer 1979; Johnston 1981; De Cecco 1987).

The alternative "portfolio approach" was based on the financial intermediation theory proposed by Gurley, Shaw, and Tobin, among others, in which banks and nonbank financial intermediaries compete for loanable funds with securities markets by issuing liabilities and purchasing claims from borrowers. In this view, the portfolio preferences of wealth owners and the characteristics (return, riskiness, liquidity) of the liabilities issued by banks determine their ability to expand their balance sheets. The portfolio approach adequately captured two essential characteristics of Eurodollars: the "near money" nature of time deposits (not means of payments, but closed substitutes for money held in anticipation of payments) and the fact that they were imperfect substitutes of domestic deposits and money market assets. The conclusion was that its growth was driven by the ability of banks to compete with other intermediaries and markets in attracting a larger share of the fast-growing global market for dollar-denominated short-term credit – "a growing slice of an expanding pie" (Dufey and Giddy 1978, pp. 107–130; Niehans 1982, pp. 17–19). As banks outside the USA competed with banks located in the USA for the intermediation of dollar-denominated liquid funds owned by wealth owners worldwide (Goodfriend 1981), Eurodollars were a substitute for domestic deposits and worked as a "parallel market" deeply integrated with national money markets in the USA and elsewhere. It offered alternative opportunities for the placement of short-term funds both to US investors for transactions in domestic currency (thus competing with the New York money market) and to non-US investors for transactions in foreign currencies (thus competing with other national money markets, for instance, the London market) (Johnston 1983, p. 76).

International and Eurocurrency Banking

Since the 1960s, when statistics about Eurodollars and other Eurocurrencies began to be compiled by central banks, Eurocurrency banking has been conventionally identified on the base of two elements: the currency of denomination (domestic vs. foreign) of banks' assets and liabilities and the residence of banks' counterparties (national vs. foreign). More precisely, it was defined as both cross-currency and cross-border intermediation (short-term bank assets and liabilities denominated in foreign currencies vis-à-vis foreign residents). This convention, however, reflected more the existing regulatory and supervisory procedures of the time, which differentiated between business in domestic and external currency, or with resident and nonresident customers, than a conceptual difference between Eurocurrency and other types of international banking. As Bryant (1987, p. 24) made clear, "from an analytical perspective, there is nothing logically compelling about the conventional definition [. . .] The establishment of asset and liability relations with foreigners and the denomination of obligations in external currencies . . . is a pervasive feature of life in interdependent national economies. There is no good reason for isolating one aspect of international banking and analysing it independently of the rest of the nexus of financial relations linking nations together."

In fact, banks continued to engage in "traditional" foreign banking activities (Lewis and Davis 1987, pp. 220–221). These included short-term lending to nonresident customers in domestic currency (cross-border) or to resident customers in foreign currency (cross-currency) in order to finance international trade or the raising of capital funds. It also entailed placing interest-bearing deposits denominated in foreign currency with foreign banks – e.g., a British bank keeping a dollar balance with a US correspondent bank (cross-border and cross-currency: in the traditional jargon of correspondent banking, a nostri claim) – to facilitate their customers' access to foreign exchange and international trade. In many cases, Eurocurrency banking originated from and was closely linked to traditional international banking, mainly channeled through foreign exchange markets (Mayer 1985). Some banks could possibly enter Eurodollar banking in order to use more efficiently the dollar liquidity owned in the form of traditional balances held with US correspondents (Lewis and Davis 1987, p. 284). Traditional international banking expanded rapidly in the postwar period and continued to provide the bulk of means for international payments, including short-term credits for trade financing and hedging forward against exchange risk (McKinnon 1977, pp. 4–5). In the early 1980s, the size of claims on foreign residents in domestic currency and claims on domestic residents in foreign currencies was, respectively, 57% and 42% of the total claims on foreign residents in foreign currencies (the conventional measure of Eurocurrency banking) in OECD countries. Taken together, assets with some international characteristics (cross-border, cross-currency, and both) represented one quarter of the gross balance sheets of banks in industrialized countries (including domestic assets); this share varied from 72% in the UK to 8% and 11% in Germany and Japan, respectively (Bryant 1987, pp. 26–27).

Scale

Measures

The main source of quantitative information on Eurodollars and other Eurocurrencies is the Bank for International Settlements, which in 1964 began to publish data on international banking activities (i.e., cross-border and cross-currency short-term assets and liabilities) reported to central banks by resident banks. Countries reporting to BIS were initially limited to G-10 members, but the geographical coverage was gradually extended (see Table 1). In 2015 BIS locational banking statistics covered banking offices of domestic and foreign institutions located in 44 countries. Data include the currency composition of resident banks' balance sheets, as well as a geographical breakdown of their counterparties (banks, nonbanks, official institutions) in more than 200 countries (BIS 2015).

BIS data can be used to proxy the scale of the different types of international banking illustrated in Fig. 1: cross-border transactions (external positions in domestic currency: e.g., US banks' dollar claims vis-à-vis non-US residents), cross-currency transactions (local positions in foreign currencies: e.g., UK banks' dollar claims vis-à-vis British residents), and Eurodollar/Eurocurrency transactions (external positions in foreign currencies: e.g., UK banks' dollar claims vis-à-vis non-British residents).

Figure 2 shows three different estimates. The first series, external total assets, is the real outstanding stock (expressed in 2010 US dollars) of short-term assets of banks located in all reporting countries vis-à-vis nonresident counterparties and denominated in foreign and domestic currencies. This can be considered as an

Table 1 Chronology of BIS reporting countries

1964	Belgium-Luxemburg, France, Italy, Netherlands, Sweden, Switzerland, UK, West Germany
1968	Canada, Japan
1970	USA (including selected US banks' branches in offshore centers in the Caribbean and Far East)
1977	Austria, Denmark, Ireland
1983	Finland, Spain, Norway, Bahamas, Bahrain, Cayman Islands, Hong Kong, Netherlands Antilles, Singapore
1997	Australia, Portugal
2000	Turkey
2001	Greece, Guernsey, India, Jersey, Isle of Man
2002	Bermuda, Brazil, Panama, Chile, Taipei
2003	Mexico
2005	South Korea
2006	Macao
2008	Malaysia
2009	Cyprus

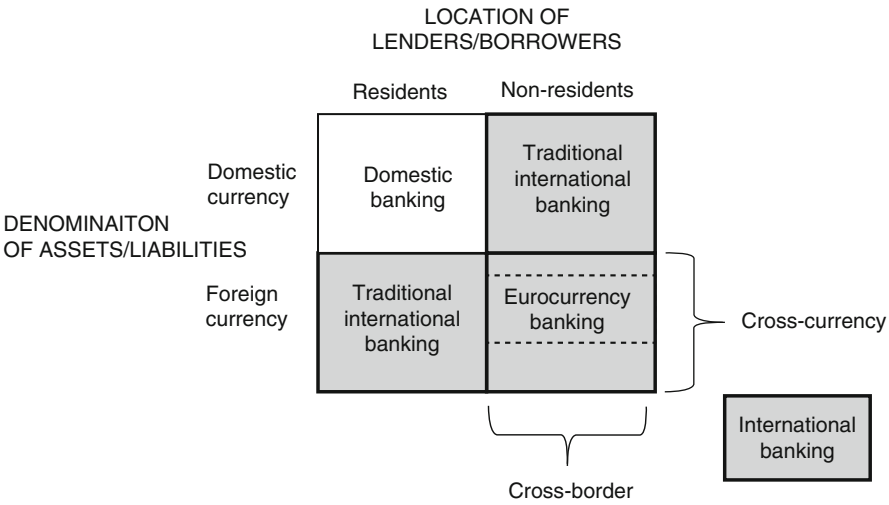


Fig. 1 Eurocurrency and international banking

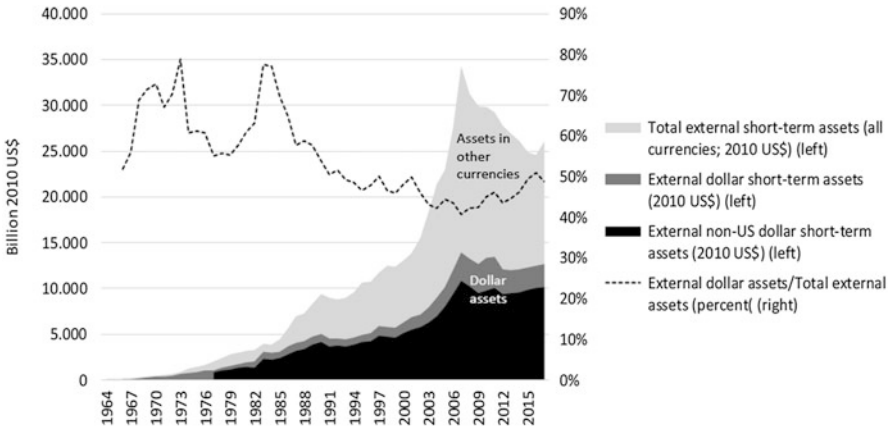


Fig. 2 External short-term assets in dollars and all currencies, 1964–2018 (real 2010 US$). Note. Total external short-term assets are positions of banks in all BIS reporting countries denominated in all currencies (foreign and domestic) vis-à-vis nonresidents. External dollar short-term assets are position of banks in all BIS reporting countries (the USA included) denominated in US dollars vis-à-vis foreign residents. External non-US dollar assets exclude from the latter the dollar positions vis-à-vis foreign residents of banks located in the USA. (Source: 1964–1976, BIS Annual Report (printed edition); 1977–2018, BIS Statistical Bulletin (online dataset at www.bis.org/statistics/index.htm)). Current US$ series are deflated by using the US GDP deflator. (Source: World Bank online data)

upper bound for the total size of the global Eurocurrency market, as it includes also external assets denominated in local currencies (part of traditional cross-border banking in the taxonomy of Fig. 1). The second series, external dollar

assets, is the real outstanding stock of dollar-denominated short-term assets of banks located in all reporting countries (the USA included) vis-à-vis nonresidents. Again, this does not fit exactly with the conventional definition of Eurodollars, as it includes dollar assets of banks operating in the USA vis-à-vis foreign residents (in principle, again, traditional cross-border banking). For this reason, the third series, external non-US dollar assets, is based only on dollar external assets of banks outside the USA (which of course includes foreign branches of US-owned banks).

Although the latter measure fits in better with the conventional definition of Eurodollars, there are reasons to prefer the second series (external dollar assets) as a more realistic measure of the size of the Eurodollar market. First, before 1983 (when offshore centers began to report directly to the BIS), foreign claims of US banks' branches in offshore centers in the Caribbean area, Central America, and the Middle and Far East were reported to the BIS by US authorities. Offshore facilities had been authorized in 1969 by the Federal Reserve in order to give US banks the opportunity to conduct transactions with foreign residents or in foreign currencies without the burden of domestic regulation, supervision, and taxation. In some cases, such as Bahamas and the Cayman Islands, offshore branches were simply segregated accounting units ("shell branches") where *entrepôt* business (dollar deposits of and dollar claims on foreign branches and subsidiaries of US residents, often contracted at head offices in the USA) was booked to circumvent domestic restrictions. In other cases, foreign banks had been allowed to establish and operate in special deregulated enclaves, such as the Asian currency units in Singapore and the offshore banking units in Bahrain (Cassard 1994). Until the mid-1980s, banking transactions conducted from offshore centers had been growing very rapidly, at an annual compound rate ranging between 30% and 50%. Overall, by 1982 offshore foreign currency assets were estimated to represent approximately one third of world claims on foreigners (Bryant 1987, pp. 134–140). Therefore, dropping data reported from US authorities would exclude the offshore assets of US banks' branches (which were big players in offshore locations), resulting in an underestimation of the size of the Eurodollar market for the period until 1983.

Second, in 1981 the Federal Reserve allowed US banks (and US branches of foreign banks) to operate International Banking Facilities (IBFs), a sort of onshore-offshore center (Palan 1998, p. 33) that is, a deregulated legal enclave in US territory, under which banks could borrow from and lend to foreign residents without being subject to domestic banking regulation. As a consequence, a large volume of Eurodollar transactions with nonresidents which until then banks had booked in the balance sheets of their overseas or offshore branches was shifted back to the USA (Key and Terrell 1989). As a matter of fact, IBFs based in New York accounted for approximately 8% of global Eurocurrency claims in the mid-1980s, a share larger than any other individual offshore center (Lewis and Davis 1987, pp. 230–231). Therefore, excluding data reported from US authorities (which include IBFs) would underestimate the size of the Eurodollar market for the period after 1981.

Growth Phases

The long-run pattern of growth of the Eurodollar market can be divided into four main stages: a long phase of almost uninterrupted expansion (with a temporary setback between 1982 and 1985), which brought the size of the market to 4 trillion dollars (expressed in real terms as 2010 USD) in 1990; a mild and short-lived reversal in the early 1990s; a new phase of exponential growth between the mid-1990s and 2007, when its size almost tripled in real terms, approaching 11 trillion dollars; and a sharp and long-lasting contraction during and after the Great Financial Crisis of 2007–2009, when the market failed to recover its pre-crisis level. In turn, other Eurocurrencies (which represented a relatively small share until the early 1980s, when the share of Eurodollars fluctuated between 60% and 80% of the total) took off in the 1980s (as countries such as Germany and Japan liberalized the international use of their currencies) and continued to expand at a faster pace than Eurodollars until the crisis of 2007. This led to a secular contraction of Eurodollars' share in total Eurocurrencies, which fell below 50% between 2000 and 2007 for the first time in history. This of course reflects the impact of the creation of the Euro and the large cross-border activities of banks within the Eurozone.

Figure 3 shows the annual and trend growth of external total assets (Eurocurrencies) and external dollar assets (Eurodollars). The annual rate at which the two markets expanded experienced a secular decline during the last quarter of the twentieth century, falling from 25–30% in the late 1960s to 5% in the late 1990s. The new century reversed this trend, bringing the two markets back to a growth pace comparable to that of the 1970s. The slowing down of Eurocurrencies and

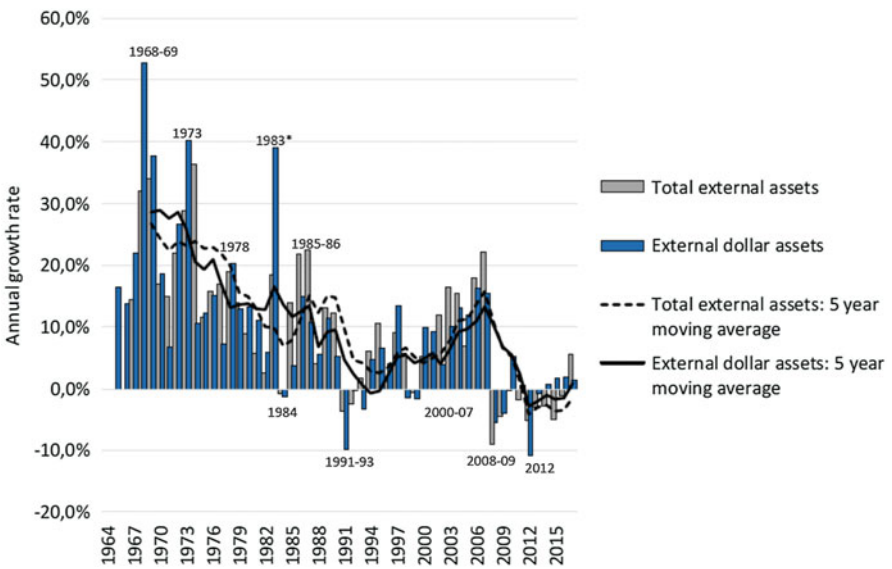


Fig. 3 Growth of external short-term assets, 1964–2018. (See Fig. 2)

Eurodollars in the 1980s is explained by securitization – that is, the emergence of financial innovations that allowed banks and commercial companies to borrow and unbundle different risk factors by issuing new types of short-term liabilities in international money markets. Examples of securitization included Euronotes, Eurocommercial paper, currency and interest rate swaps and options, and forward rate agreements (FRAs). In this new context, the role of banks gradually shifted from traditional direct intermediation to services related to the origination, underwriting, and placing of marketable securities, with a contraction of their traditional Eurodollar balance sheets and an expansion of off-balance sheet business (BIS 1986; Bryant 1987, pp. 51–57). However, as will be seen in the next section, the new phase of expansion of Eurodollars after the turn of the century was mainly driven by European banks' dollar borrowing from US money market funds to invest in asset-backed securities originated by the shadow banking system in the USA (He and McCauley 2012).

Eurocurrencies, Globalization, and Technological Innovation

In the Bretton Woods system, the US dollar was the main reserve and intervention currency. This generated a worldwide demand to hold dollar balances with US banks as a means of settling international current and capital account transactions and to borrow dollars as a medium for deferred payments. The US dollar also emerged quickly as the vehicle currency in foreign exchange transactions (Lewis and Davis 1987, p. 270). As a matter of fact, the exponential growth of the Eurodollar market in the 1960s coincided chronologically with the period in which Eichengreen et al. (2016) detected structural shifts in the determinants of the composition of international reserves. This suggests that Eurodollars may have actively contributed to the dominant role of the dollar as a source of incremental international liquidity in the second half of the twentieth century (He and McCauley 2010; Eichengreen 2012). Therefore, the important (though elusive) question arises to what extent the expansion of Eurodollars and Eurocurrencies in general simply followed the expansion of international trade and the multinationalization of companies or was instead also an autonomous process driven by other factors. The latter include technological innovations in the transmission and processing of information, the execution of transactions and payments, and the transfer of funds, which increased banks' sensitivity to cross-border and cross-currency arbitrage and investment opportunities. Cross-country asymmetries in regulation, supervision, and taxation of financial intermediaries were a third crucial factor that gave private sector agents strong incentives to relocate geographically part of their activities in order to benefit from less constraining environments – a form of "regulatory arbitrage" (Bryant 1987, pp. 62–73) extensively discussed in the next section.

Figure 4 shows the historical evolution of the ratio of total external and dollar assets (including interbank positions) to global trade. Although in principle only claims vis-à-vis nonbank end users should be directly associated with the financing of trade, part of interbank positions also might be driven by trade transactions.

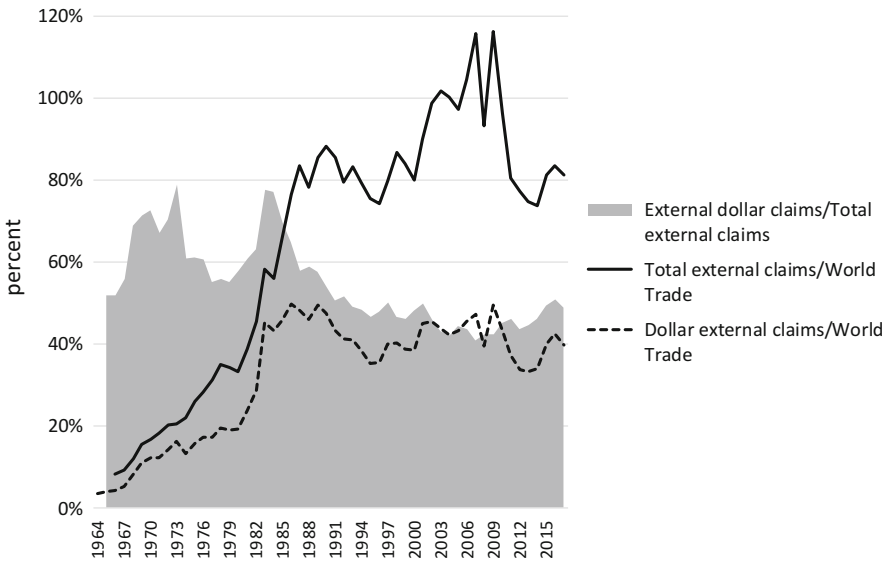


Fig. 4 International banking and world trade. Note. External dollar claims are short-term dollar-denominated assets of banks from BIS reporting countries (including the USA) vis-à-vis foreign residents. Total external claims are short-term assets denominated in all currencies (including domestic) of banks from BIS reporting countries vis-à-vis foreign residents. (Source: 1964–1976, BIS Annual Reports (printed edition); 1977–2018, BIS Statistical Bulletin (online dataset)). World trade is total merchandise exports and imports in current US$. (Source: World Bank online dataset)

In fact, by purchasing a large volume of dollars in the spot market and investing them in Eurodollar interbank deposits with different maturities to match the schedule of future payments, traders would significantly reduce transaction costs (Swoboda 1968; Makin 1972; Johnston 1983, pp. 76–81). Data show that until the mid-1980s and then again in the period that preceded the Great Crisis, the growth of total external short-term claims outpaced the growth of international trade, increasing from 20% to 80% of world merchandise trade between 1970 and 1985 and approaching 120% in 2007. This suggests that, at least in these two periods, cross-border flows of short-term banking funds responded also to factors not directly related to the globalization of trade and production. Since the late 1980s, however, its share remained stable, fluctuating between 40% and 50% of global trade. Interestingly the dollar component of external claims followed a different pattern, as its ratio to international trade reached a peak around 40% in the very early 1980s and then hovered around this level for the rest of period. This suggests that non-trade factors – for instance, portfolio diversification – were especially relevant for cross-border flows in other international currencies, such as the Japanese yen in the 1980s and the Euro after 2000.

Finally, at least in the case of Eurodollars, the growth of Eurocurrencies was not driven by the expansion of domestic bank money. As shown in Fig. 5, between the late 1960s and the early 1980s, external dollar-denominated short-term assets

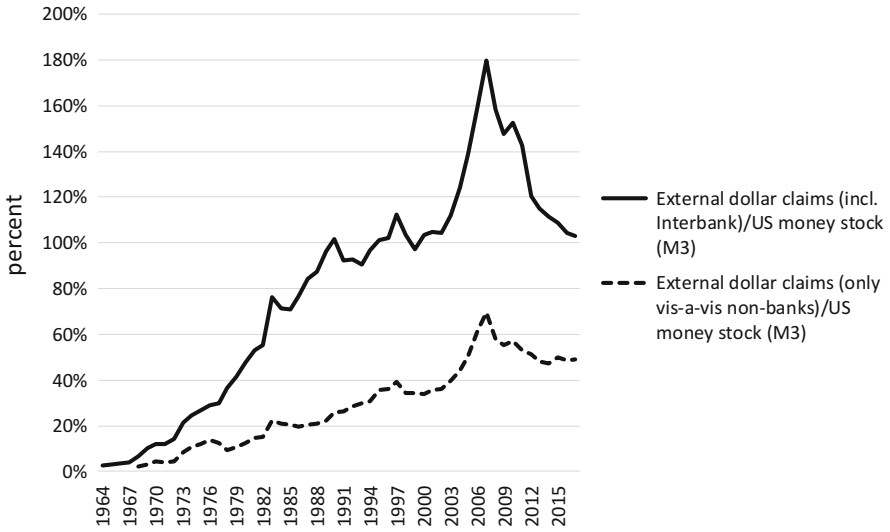


Fig. 5 Eurodollars and US money stock. Note. External dollar claims are short-term dollar-denominated assets of banks from BIS reporting countries (including the USA) vis-à-vis foreign residents. (Source: 1964–1976, BIS Annual Reports (printed edition); 1977–2018, BIS Statistical Bulletin (online dataset)). US Money stock is M3. (Source: OECD data retrieved from FRED (Federal Reserve Bank of St. Louis))

escalated from 7% to 48% (including interbank positions) and from 2% to 12% (considering only positions vis-à-vis nonbanks) of the US money stock, measured on the base of the monetary aggregate M3. These figures went up to 100% and 35% by the end of the twentieth century and to 180% and 70% in the run-up to the crisis of 2007–2009. This is just one aspect of the "excess elasticity" (insufficient constraints on the availability of external finance and the creation of credit) of a deregulated international monetary and financial system, which systematically favored the emergence of financial imbalances in response to expansionary monetary policies since the 1990s (Borio and Disyatat 2011).

Structure

Brokers, Dealers, and Tiers

Eurocurrencies are a modern example of international money markets. Their closest historical antecedent was the pre-1914 international money market in London, when the city acted as "the center of world liquidity." In that period, Sterling-denominated bankers' acceptances (bills of exchange endorsed by merchant banks on behalf of foreign borrowers) were not only the key instrument to finance world trade and short-term capital movements – including US international trade (Eichengreen 2011, pp. 14–19) – but also money market assets globally

demanded by banks as short-term investment to manage their liquidity and by official institutions (central banks and Treasuries) as a substitute for gold reserves (Flandreau and Jobst 2005; Bignon et al. 2012; Flandreau and Ugolini 2013; Accominotti and Ugolini 2019).

A money market works as a clearing mechanism for the surpluses and deficits of funds in an economy, which allows borrowers and lenders to manage liquidity efficiently. Liquidity management (the ability to minimize the holding of idle cash balances in case of liquidity surplus and access readily available cash to meet payments and other short-term financial obligations in case of liquidity deficit) is an essential function of treasuries and finance departments of both financial and nonfinancial organizations. The development of a secondary money market in which short-term debt securities are traded greatly enhances the efficiency of liquidity management. A key role in money markets is played by brokers and dealers. Money brokers manage communication networks that allow a smooth flow of information between borrowers and lenders; they act as agents and earn a fee for their service. Money dealers are banks or specialized intermediaries that act as principals and perform the function of market makers by trading on their own account. They bid for surplus liquid funds of wealth owners by issuing short-term liabilities in the form of indirect securities, such as time deposits and CDs. On the asset side, dealers invest in a portfolio of short-term debt securities issued by economic units with temporary shortages of liquid funds. They make the market by quoting continuously bid and ask prices at which they are prepared to buy or sell and earn a profit from the spread between borrowing costs and the return on their portfolio.

A large proportion of Eurodollar transactions was linked to foreign exchange transactions, as liquid funds were converted from other currencies into dollars and the exchange risk was covered with forward contracts. In fact, in the early years, most banks in London dealt with Eurodollars and other Eurocurrencies from their foreign exchange dealing rooms (Einzig 1971, p. 138). In a similar fashion, foreign currency brokers (not money market brokers) acted as the main intermediaries. US banks' branches were an exception, as they handled the business from their money market departments since the dollar for them was the domestic currency. The information pooled by brokers was vital in the initial stages of the market, as they allowed banks to save in information and transaction costs, to approach the market on their own terms, and to preserve anonymity in the early stage of a dealing. As the market became consolidated, banks built specialized units of Eurodollar and Eurocurrency dealers in order to monitor the market directly, and the intermediary role of brokers was partially replaced by direct dealings between banks. In fact, there were very few banks specialized in Eurodollar and Eurocurrency banking. "Eurobanks" were branches, subsidiaries, or specialized departments of domestic banks, and their Eurocurrency book represented a specific compartment of their balance sheets (Lewis and Davis 1987, p. 271).

The market operated on the base of a multitier structure, in which tiers reflected differences in creditworthiness and risk assigned to banks. Since the mid-1960s, the top tier was occupied by foreign branches of prime US banks from money centers

such as New York and Chicago, who acted as main dealers and market makers. They could always borrow at marginally lower rates, ran large books of Eurodollar deposits, and quoted bid and ask rates on a permanent basis. Spreads between tiers (the differential rate at which banks in different tiers could borrow, with higher rates reflecting the risk premium demanded by depositors) were not constant and tended to become more pronounced during period of liquidity strain. For instance, in 1974 (a year characterized by a number of important banking crises), the market operated on the base of nine tiers, with US banks at the top and Italian and Japanese banks at the bottom (Sarver 1988, pp. 28–30). Tiering also responded to banks' exposure to specific risk; for instance, during the Asian crisis of 1997–1998, Japanese banks were heavily penalized and could borrow only at a rate significantly higher than European banks (Stigum and Crescenzi 2007, pp. 849–850).

Transaction Technology

Eurocurrency transactions were based on standardized contracts for large fixed amounts, as typical of wholesale banking. In the 1970s deposits of 1 million USD were the norm, but larger transactions (up to 10 million USD) were not unusual, as the market appealed especially to very large depositors in its initial stage. Maturities ranged from overnight to 1 year, but 3 months was the standard, and the rate on 3-month deposits became the benchmark rate quoted in the market. After 1966 US banks' branches in London introduced in the market Eurodollar certificates of deposits, a negotiable instrument with the same characteristics as the CDs introduced a few years earlier by US commercial banks in the domestic money market. Negotiability allowed holders of Eurodollar CDs to convert easily their investment into cash in an efficient secondary market. This was especially useful in case of unexpected payments or anticipation of adverse movements of the dollar exchange rate – circumstances which could be accommodated only by negotiating option or penalty clauses in a fixed term deposit contract. Eurodollar CDs were mainly issued for smaller amounts than time deposits, mainly in a range between 25,000 and 100,000 USD, with maturities from 3 months to 2 years, and with rates at a slight discount compared to equivalent time deposits (Einzig 1971, pp. 155–167; Shaw 1978, pp. 94–113).

Transactions were dealt with by telephone or telex, and the market was essentially a global network of telephones, telexes, and monitor screens connecting banks and brokers in different financial centers (Einzig 1971, pp. 136–139; Lewis and Davis 1987, p. 271). Technological development rapidly reduced information barriers and transaction costs, thus enhancing multilateral trading and the development of an almost perfectly efficient market (Agmon and Barnea 1977; Frenkel and Levich 1975, 1977). The efficiency of Eurodollar interbank trading was greatly enhanced in the 1970s by the establishment of computerized international network systems of private clearing such as CHIPS (Clearinghouse Interbank Payments System, managed by the New York Clearing House) and message transmission such as SWIFT (Society for Worldwide Interbank Financial Telecommunications), and

new advanced information services offered to interbank traders by Reuters and Telerate (Sarver 1988, pp. 207–221).

Bid-ask spreads on Eurodollar transactions were narrower than in domestic intermediation. For dollar-holding investors and borrowers, US banks and money markets were the natural outlet and source of funds. As a consequence, banks in the Eurodollar market had to offer competitive terms, that is, higher yields on deposits and CDs, and lower rates on loans. This resulted in margins between 0.125% and 0.0625% or less on annual basis for short-term interbank transactions. In spite of very narrow margins, Eurodollar dealings could generate significant profits, thanks to the absence of costs from reserve requirements and other regulatory constraints. They also had very low transaction and information costs, thanks to modern communication technologies, and low administrative costs, thanks to economies of scale (Frenkel and Levich 1975, 1977; Agmon and Barnea 1977). However, banks in search of more substantial margins engaged systematically in maturity transformation – the practice of borrowing at short maturities and lending at longer maturities, usually to nonbank borrowers – provided that the yield curve was positive. By doing so, they did not limit themselves to liquidity distribution but became increasingly involved in liquidity production. To some extent, the development of the interbank market for Eurodollar deposits can be seen as an institutional mechanism aimed at mitigating the risk generated by maturity transformation, as it allowed banks easy access to short-term funding on a global scale (Lewis and Davis 1987, pp. 108–110).

The Interbank Market

A large share of Eurocurrencies were originated by interbank transactions, with a long chain of banks located in the same or in different financial centers acting as intermediary between original depositors and final borrowers. Occasionally the practice of interbank borrowing and lending was termed "deposit pyramiding," a term that evoked banking practices of the National Banking Era in the USA, when country banks used deposits with New York banks as reserves (Mehrling 2002). The disaggregation of BIS data by counterparties allows a more precise estimate of the size of the interbank and its evolution over time. One should bear in mind, however, that for a long period, counterparty information reported only assets and liabilities vis-à-vis nonbanks; as a consequence, bank counterparties could be identified only residually, which might overestimate the size of the interbank market. Moreover, some transactions with nonbanks (for instance, a forward sale of foreign exchange to a commercial company) did not appear on the banks' balance sheets, but generated interbank transactions as the bank hedged the risk through lending and borrowing in the interbank market. This also tended to overestimate the size of the interbank market (BIS 1983, pp. 15–22). In turn, since local positions (i.e., vis-à-vis residents) are not included in the estimated Eurocurrency series, they do not take into account within-border interbank dealings (i.e., interbank transactions between banks located in the same financial center). This may underestimate the actual size of interbank lending. Finally, until the mid-1970s, only European banks reported counterparty

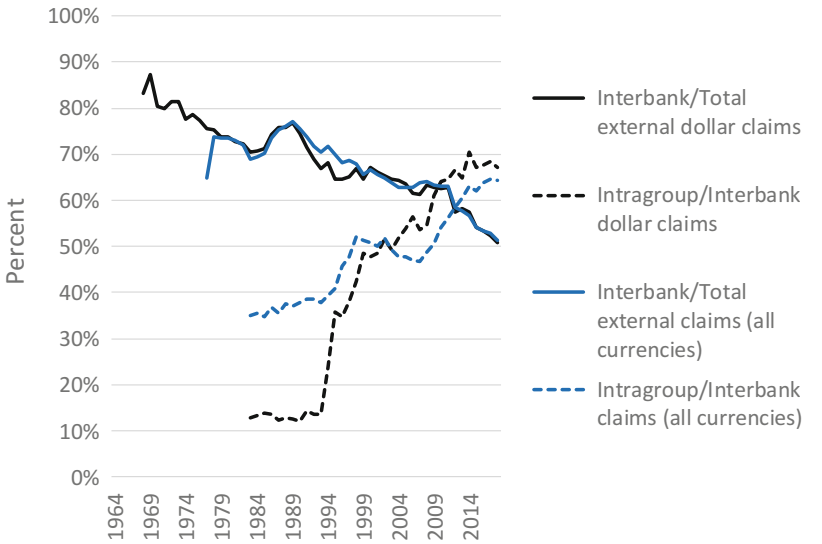


Fig. 6 Eurocurrency and Eurodollar interbank markets. Note. External dollar claims are short-term dollar-denominated assets of banks from BIS reporting countries (including the USA) vis-à-vis foreign residents. Interbank claims are obtained by subtracting from this series the positions vis-à-vis nonresident nonbanks. (Source: 1964–1976, BIS Annual Reports (printed edition); 1977–2018, BIS Statistical Bulletin (online dataset)). Short-term claims of reporting banks (including the USA) vis-à-vis nonresident bank counterparties. (Source: BIS Annual Reports 1964–1976, BIS Annual Report (printed edition); 1977–2018, BIS Statistical Bulletin (online))

information, so that early estimates of the interbank market (for Eurodollars) should be regarded as indicative.

By netting out interbank positions, a more precise estimate of the magnitude of credit generated by Eurocurrencies in favor of nonbank users can be obtained (Mayer 1979; Johnston 1983). In Fig. 6 a downward long-run trend of interbank share of interbank positions can be observed, from 80% in the late 1960s to 60% on the eve of the 2007 crisis, with virtually no differences between Eurocurrencies as a whole and Eurodollars. This is consistent with previous estimates for all Eurocurrencies, which suggested that interbank lending accounted for between 60% and 70% of short-term foreign currency assets throughout the 1980s and 1990s (Ellis 1981; Stigum and Crescenzi 2007, p. 827), although figures for "inside area" positions (i.e., those vis-à-vis banks in major financial centers – the "hard core" of the global interbank market) were probably slightly lower (BIS 1983).

The decline in interbank recycling suggests that, as the Eurocurrency market continued to grow, its capacity to provide credit to nonbank borrowers worldwide increased significantly over time. As will be seen later, since the late 1990s, an increasing share of borrowers from banks in the Eurodollar market were US securities firms and nonbank financial institutions ("shadow banks") raising funds in the repo market. The decline in on-balance sheet interbank positions reflected also the rising popularity of new off-balance sheet instruments – for instance, since the mid-1980s, interbank borrowing on the base of cash Eurodollar deposits was

replaced by FRAs (forward rate agreements). However, shrinking interbank transactions were also caused by rising concerns about credit risk and a more general deterioration of the risk environment facing banks, as growing exposure to derivative and other off-balance sheet instruments and the growing number of participants subject to poor disclosure requirements made it more difficult to ascertain the creditworthiness of bank counterparties. Capital requirements introduced under Basel I also induced banks to reduce interbank exposures that generated low risk-adjusted profits (BIS 1986, 1992; Lewis and Davis 1987, pp. 115–122).

The interbank market performed four fundamental functions in international liquidity (Johnston 1983, pp. 98–103): liquidity smoothing, as the existence of large interbank market allowed banks to economize on liquidity buffer (the stock of liquid assets, cash, and balances held for precautionary purposes); liquidity transfer, by which market makers –usually privileged by primary nonbank investors as original recipient of Eurodeposits (i.e., major US banks) – redistributed excess liquidity to banks of minor standing within the same financial center; currency transfer, the process by which banks matched the currency composition of their assets and liabilities through interbank trading; and global liquidity distribution, as transaction costs between peripheral banks in different countries and the Eurocurrency center, thanks to the existence of branches and subsidiaries, were lower than those between peripheral banks themselves. In general terms, London as the major Eurocurrency center – by performing functions of liquidity and currency transfer which facilitated the distribution of liquidity between banking systems – acted as "a form of global clearing system for currency flows" (Johnston 1983, p. 101).

More recently, "intragroup" transactions (i.e., positions between branches and offices of the same bank) assumed an unprecedented relevance in Eurocurrency interbank business. The internalization of the recycling functions performed by the interbank market was especially relevant for large multinational institutions with branches in different financial centers, such as US and European commercial banks. Data in Fig. 6 suggest that intragroup transactions remained marginal during the 1980s in Eurodollars (slightly above 10% of total external interbank dealings), but were much more important for Eurocurrencies as a whole, which suggest they played a prominent role in non-dollar Eurocurrency markets. Intragroup transactions expanded rapidly since the 1990s and accounted for ca. 60% of total interbank dealings on the eve of the Great Financial Crisis of 2007–2009. This, again, can be related to the "global banking glut" of European banks and reflected their practice of borrowing short-term dollar funds in the US money market through their US branches and shipping them to their European headquarters (Shin 2011, pp. 18–19).

Growth

Regulatory Arbitrage

As anticipated in the Introduction, unregulated Eurocurrencies offered depositors higher yields and borrowers short-term credit at lower costs. As a consequence, regulatory arbitrage is considered in the literature as a major determinant of the early

expansion of offshore deposits (Dufey and Giddy 1978, pp. 133–135; Aliber 1980; Johnston 1983, pp. 86–87; Bryant 1987, pp. 66–68). Regulatory arbitrage is a strategy that generates profits by exploiting functional similarities across financial products or processes (i.e., a time deposits or a short-term loan booked onshore in a US bank and offshore in a non-US bank or a foreign branch of a US bank have a similar functional value for lenders and borrowers, so that they are close although not perfect substitutes) and permanent differences in their treatment under legal and regulatory practices across jurisdictions (Fleischer 2010; Houston et al. 2012). In the case of Eurodollars and Eurocurrencies, by locating their wholesale intermediation offshore in London or other financial centers, banks successfully circumvented domestic regulations that increased intermediation costs and constrained their balance sheet expansion.

Eurodollars are a classical example. In US domestic banking, rates on short-term deposits with commercial banks were regulated under the so-called Regulation Q. Introduced during the Great Depression as part of the Banking Acts of 1933 and 1935, Reg Q applied both to member banks of the Federal Reserve system and to non-member banks insured under FDIC (Federal Deposit Insurance Corporation). It prohibited the payment of interest on demand deposits and gave the Federal Reserve the power to set ceilings on interest rates of savings and time deposits, with the aim of reducing competition for deposits and limit banks' investment in risky assets (Benston 1964; Friedman 1975). When tight monetary policy pushed interest rates in domestic money markets (Treasury Bills, commercial paper) above the ceilings, US commercial banks suffered domestic disintermediation, as commercial companies and institutional investors shifted their liquid balance from bank deposits to alternative money market instrument (Baxter 1966, 1968). The emergence of Eurodollars in London brought the challenge of disintermediation to the international level and reduced their ability to intermediate worldwide dollar liquidity.

US banks responded by innovating and moving. Domestically, in 1961 they started issuing certificates of deposits (CDs), a negotiable money market instrument for which an active secondary market was rapidly organized in New York by specialized dealers. Marketability fully insured large depositors against the risk of liquidity shocks, as they could now manage their position more efficiently by trading CDs in a very liquid market. As a consequence, commercial banks recovered part of their competitive edge. However, the Fed quickly extended its regulatory perimeter by imposing differentiated ceilings on small and large CDs, which would again become binding in periods of very high short-term interest rates (Ruebling 1970; Mayer 1982). The second strategic response was to "swarm" to Europe, mainly by opening branches in London, and to offshore financial centers in the Caribbean and Far East, fully to exploit their natural competitive advantage in the Eurodollar market (Sylla 2002). By the mid-1960s, US banks had already overcome British merchant and overseas banks as the dominant players in the London market (Battilossi 2002b, p. 106). Their international expansion would continue at a sustained pace until the mid-1980s, when the number of US commercial banks with foreign branches peaked at 162, although the subsequent phase

of consolidation and retrenchment reduced their numbers by half (there were only 92 in 1998) (Houpt 1999).

The interaction between bank regulation, monetary policy, and balance-of-payments controls created the conditions for a first cycle of exponential growth in the second half of the 1960s. As monetary policy became restrictive in order to stem domestic inflationary pressures, the Federal Reserve allowed short-term interest rates to exceed by a wide margin Reg Q's ceiling in order to constrain the growth of domestic bank credit. During the credit crunches of 1966 and 1968–1969, large interest differentials in favor of the Eurodollar market attracted a massive inflow of deposits into branches of US banks in London and offshore centers both from US residents and foreign countries. In the same period, voluntary and mandatory capital control programs enforced by the Johnson administration curbed US banks' ability to lend abroad to US multinationals, which turned massively to US banks' branches in London as a source of finance. In order to obviate the effect of the credit crunches, US banks borrowed massively in the Eurodollar market through their London-based and overseas branches and then channeled part of these funds back to their head offices in the USA, part to US multinationals (Kane 1983, pp. 28–50; De Cecco 1987; Woynilower 1980; Meltzer 2009, pp. 739–741). These factors converted for a while the Eurodollar market into a virtual branch of the New York money market (Einzig 1971, p. 143).

Over time, the impact of regulatory arbitrage on the market's growth most likely declined. In 1969 the Federal Reserve, under Regulation D, imposed reserve requirements on net borrowings by US banks in the Eurodollar market, which reduced their incentives to use it in times of domestic tight money (but shifted pressure on the market for bank-related commercial paper, which possibly contributed to the crisis of Penn Central in 1970: Meltzer 2009, pp. 739–741). Constraints on exports of short- and long-term capital were abolished in 1974 by the Nixon administration. Regulation Q ceilings on large time deposits (above 100,000 USD) and CDs were completely removed between 1970 and 1973, while ceilings on smaller time deposits and savings deposits were also gradually relaxed. The threat of disintermediation did not vanish completely – in fact on occasions, high domestic money market rates made regulation binding again, such as in the disinflation period 1979–1982, which led to a new phase of sustained expansion of the Eurodollar market (He and McCauley 2012, p. 39). However, the policy trend favored regulatory convergence between offshore and onshore markets. Reserve requirements on US banks' borrowing in the Eurodollar market were eliminated in 1978. The Depository Institutions Deregulation Act of 1980 set in motion the phasing out of interest rate regulation, which was completed in 1986 (Calem 1985; Koch 2015). In 1981 the Federal Reserve authorized banks located in the USA (either domestic or foreign) to establish International Banking Facilities (IBFs), as already mentioned. In 1990, reserve requirements were finally eliminated on large-denomination domestic time deposits and CDs. Although domestic intermediation costs remained slightly higher than in the external dollar market, due to tougher capital standards, higher deposit insurance premiums, and more active supervision after the wave of banking crises of the 1980s (Goodfriend 1998), the establishment of a domestic deregulated

environment enabled New York to compete successfully with London and, more directly, with Caribbean offshore centers (Key and Terrell 1989). Regulatory convergence between onshore and offshore locations was the main factor behind the sharp contraction of outstanding Eurodollar claims in 1991–1993 – the first ever since its emergence – as US banks shifted a significant amount of their cross-border assets back to US-based offices. The domestic crisis of Japanese banks and the reduction of their international business was also an important contributing factor.

Petrodollar Recycling

In the 1970s, an additional impulse to the Eurodollar market's expansion came from oil shocks. In this period, the rising demand for external finance by corporate and sovereign borrowers in oil-importing countries, and especially in the developing world, was met by the strong preference of private and public institutions in oil-exporting Arab countries for Eurodollar deposits. These were considered as an investment with higher liquidity, shorter maturities, and, importantly, lower exposure to political risk than those offered by banks in the USA. The "recycling" of the large cash surplus of OPEC countries to finance imbalances of oil-importing countries fell largely on the Eurodollar market, which provided the main source of funding for a new market of long-term Euroloans arranged by international bank syndicates (Johnston 1983, pp. 25–27 and 144–159; Altamura 2017, pp. 101–130). In this phase, pure offshore transactions dominated in the market, with non-US banks (mostly European) intermediating between depositors in oil-producing countries and borrowers in oil-importing countries (He and McCauley 2012, pp. 35–36).

The 1970s saw also a rapid internationalization of European banks. As highly regulated and cartelized domestic environments provided modest opportunities for growth in retail banking, European commercial banks turned eagerly to wholesale international banking. By the early 1980s, British, German, French, Dutch, and Swiss banks had already established branches and subsidiaries in all major international financial centers, either directly or through bank alliances and consortium joint ventures (Ross 2002). They also entered massively into the US market (Bryant 1987, pp. 35–45; Grosse and Goldberg 1991) leading to an unprecedented internationalization of the New York financial center (with more than 350 foreign branches and subsidiaries in 1980 – a higher number than in London: Pecchioli 1983, pp. 154–183) and more intimate connections with US money markets. In this period the fast growth of their dollar balance sheets became a major factor in the growth of the Eurodollar market (Battilossi 2002a, b; Altamura 2017, pp. 55–83). Japanese banks showed a similar pattern, with a fast expansion of their international branch networks and their claims vis-à-vis foreigners. In fact, in the mid-1980s, Japanese banks had become the major bank group in London and accounted for 26% of total international banking (measured as total external assets by nationality of ownership), against 23% of US banks and another 22% of French, British, and German banks jointly (Bryant 1987, pp. 34–35 and 49–51).

Although the "recycling" function of the Eurodollar market reached an unprecedented scale in the 1970s, it was not an entirely new phenomenon. Early descriptions of the geographic structure of Eurodollar assets and liabilities (Klopstock 1965; Altman 1967) had already emphasized the role of the market as a mechanism that attracted liquidity from developing peripheries. Depositors from the developing world included central banks (for which the interest earned on their foreign exchange reserve assets was a relevant factor in their portfolio composition), commercial banks, and nonfinancial corporations (whose net liquid resources had no domestic money markets to be invested in). Part of the international liquidity originated in the Middle East, Latin and Central America was channeled to the Eurodollar market in London through Swiss banks, which were large recipients of foreign liquid balances but lacked a sufficiently developed domestic money market to invest them. A similar role was credited to Canadian banks, which held large liquid balances of US corporations seeking yields above domestic regulated deposit rates. While in the 1960s international liquidity from developing countries had been mainly recycled toward industrial economies, in the 1970s the pattern reflected a South-South flow of funds intermediated by US, European, and Japanese banks. In this new circumstance, also Latin American banks expanded rapidly their international presence in New York and London, directly and through consortium ventures, and borrowed extensively in the Eurodollar market to fund domestic business (Álvarez 2015, 2017, 2019).

A "Global Banking Glut"

The last phase of fast growth that started in the second half of the 1990s reversed a trend of secular decline in the rate of expansion of Eurodollars and Eurocurrencies. In this period, characterized by historically low interest rates and inflation (the so-called Great Moderation), European banks played a distinctive role. The explosive expansion of their dollar balance sheets was funded by wholesale deposits in the Eurodollar market or in other Eurocurrencies, which were then converted into dollars in the spot market, hedging exchange risk with swap contracts (a prearranged sale at a forward date and at a specified exchange rate). In many cases, wholesale funds were placed by US residents (including money market funds) with US-based branches of European banks, which then channeled these funds to their headquarters in Europe. These, in turn, invested in dollar-denominated longer-term assets, including structured finance products (i.e., subprime mortgage-backed securities) originated in the US shadow banking system (Bernanke et al. 2011). As a consequence, during the "global banking glut" of the pre-2007 period – in which European banks expanded credit not only to the USA but also to Euro Area countries (Shin 2011) – the Eurodollar market operated mainly as a conduit for offshore round-tripping of funds. This in turn contributed to the easy credit conditions that prevailed in the years before the crisis (McGuire 2004; Allen and Moessner 2011; He and McCauley 2012, pp. 35–40). As assets had usually longer maturities than deposits, banks had to renew periodically their Eurodollar funding. In August 2007, a widespread loss

of confidence in banks' creditworthiness led to a sudden and widespread "freezing" of the market, during which banks lost access to interbank wholesale deposits, especially for longer maturities. The sharp contraction of the outstanding stock of external dollar claims between 2007 and 2012 was the largest ever in the 60 years of history of the market and reflected a massive deleveraging of European banks' balance sheets.

Arbitrage

The Eurodollar Premium

During the last 60 years, the interbank rate on Eurodollars – the London interbank offer rate (LIBOR), a compound average of "offer" rates in the London market – has played the key role of "true global cost of money" (Stigum and Crescenzi 2007, p. 240). Over this very long period, Eurodollar interest rates exhibited two empirical regularities: they commanded a structural premium on comparable domestic deposits and other US money market rates; and they covaried strongly with US domestic interest rates, suggesting a causal relationship between domestic monetary conditions and the global dollar money market. Both characteristics can be observed in Fig. 7, which plots interest rates on 3-month Eurodollar deposits in London against three benchmark money market rates in the USA: the rate on overnight federal funds, 3-month Treasury Bills, and 3-month CDs issued by commercial banks. For Eurodollars, as well as other Eurocurrencies, the positive spread between offshore and onshore deposits is generally explained by sovereign and credit risk (Galpin et al. 2009). The yield pickup that depositors could earn by investing in Eurodollars rather than in US deposits remained large until the early 1980s. This contributes to explain the rising portion of official dollar reserves held by central banks with banks outside the USA. Other factors were litigation risk (i.e., the possibility that US investors demand the seizing of reserves and other assets in case of sovereign debt default) and infrastructure risk (i.e., diversification of reserve holdings across locations in different time zones to guarantee uninterrupted access to liquidity) (McCauley 2005).

The consensus view in the literature is that the main determinant of both premium and covariance lies in the increasing efficiency of the mechanism of international arbitrage (Dufey and Giddy 1978, pp. 48–106; Johnston 1983, pp. 110–143; Gibson 1989, pp. 68–104). This, in turn, can be explained by technological and institutional changes. The impact of new communication technologies on financial services – which include the processing and transmission of information, the confirmation of transactions, electronic accounting, and fund transfer – reduced financial frictions and increased the elasticity with which both lenders and borrowers exploited opportunities for arbitrage across currencies and across borders. At the same time, the unprecedented expansion of multinational banking gave banks direct access to information, arbitrage, and intermediation activities in foreign and international markets (Bryant 1987, pp. 64–66). Arbitrage,

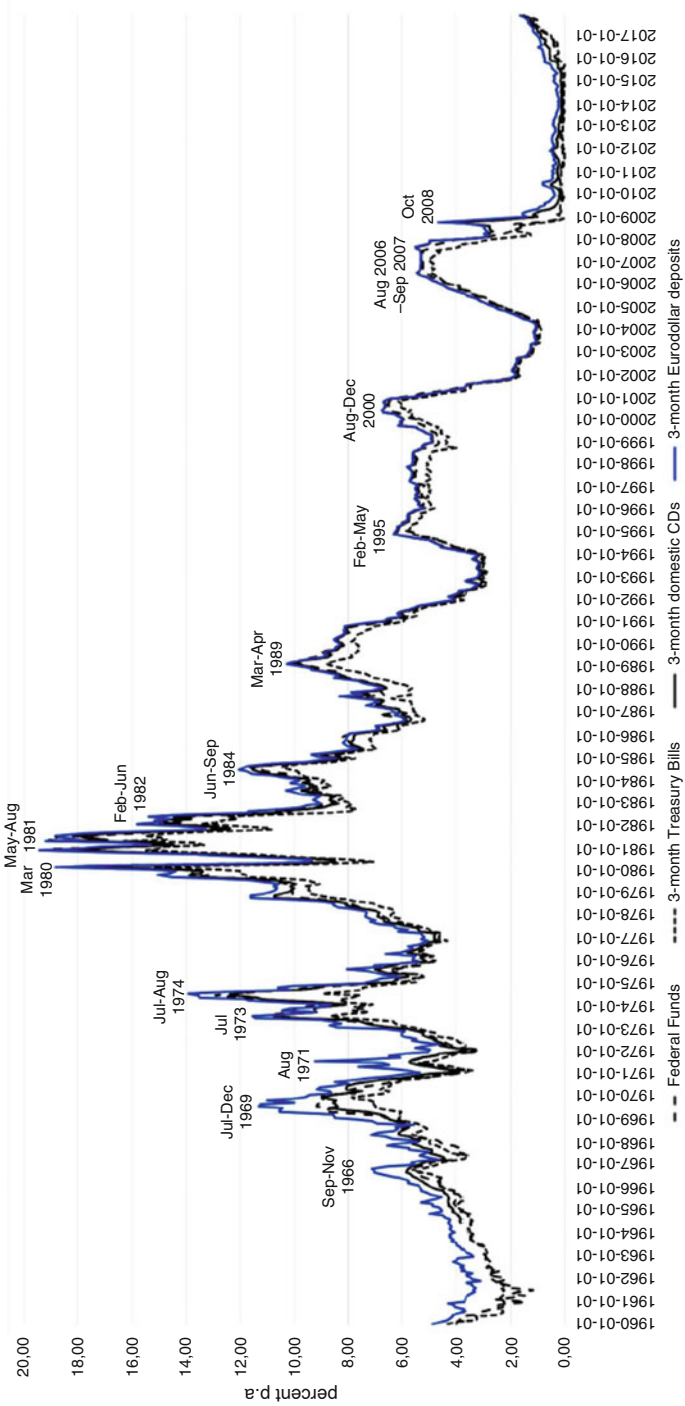


Fig. 7 Eurodollar and US money market rates, 1960–2017. (Data from FRED (Federal Reserve Bank of St. Louis) online datasets)

interest parity, and portfolio adjustments (by depositors, borrowers, and financial intermediaries) represent the building blocks of supply-and-demand models of the international money market (Marston 1974; Dufey and Giddy 1978, pp. 130–135; Gibson 1989, pp. 49–67). The interest parity theory states that, with efficient markets, the differential between interest rates on assets of similar riskiness denominated in different currencies should equal the forward discount/premium, so that arbitrage has no profit opportunities to exploit. In turn the efficiency of arbitrage enhances financial integration – that is, the sensitivity of short-term capital flows and portfolio stock adjustments to changes in interest rates (Gibson 1989, pp. 33–40; Marston 1976, 1997, pp. 70–104).

US Banks' Arbitrage

The role of US banks in the process of arbitrage was especially critical. For them, Eurodollar deposits were close and reserve-free substitutes of domestic deposits. Similar to other innovations – such as money market funds – they allowed banks to minimize their level of reserve holdings and expand their balance sheets without being constrained by the supply of reserves by the Federal Reserve. Their transitory arbitrage (triggered by temporary discrepancies of interest rates in the two markets) was mainly outward: banks borrowed CDs or commercial paper (the latter through bank holding companies) domestically and invested these funds in the "external" market at slightly higher rates, until they reached internally imposed arbitrage constraints (such as perceived risk, capital-to-asset, and return-on-assets ratios) or interest rate convergence removed arbitrage incentives (Frydl 1982, p. 13). In fact the empirical evidence suggests that adjustments in US banks' portfolios – more sensitive to variations in interest rates than individual portfolios – were the main determinant of the strong covariation between Eurodollar and US short-term rates, while the direction of causality ran from the domestic to the external market (Kreicher 1982; Marston 1997 pp. 53–57). This notion was clear to practitioners too; as expressed by an experienced banker in the market: "Rarely does the tail wag the dog. The US money market is the dog, the Eurodollar market, the tail" (Stigum and Crescenzi 2007, p. 860).

The dissemination of theoretical and practical knowledge about how the market worked also may have contributed to reduce information barriers to arbitrage. By the early 1980s, the global scale achieved by the market guaranteed almost perfect access to private liquidity, and the deregulation of financial systems ensured an almost perfect substitutability of domestic and "external" money market assets. Banks and other market participants had completed their learning curve of portfolio adjustment – what Dufey and Giddy (1978, pp. 54–55) termed "information effect." As a consequence, the interest rate incentive required in order to induce depositors and borrowers to switch from domestic to "external" markets experienced a secular decline, and the arbitrage mechanism became more elastic (Kreicher 1982, pp. 10–23; Johnston 1983, pp. 110–142).

Risk

Market Spreads

Figure 8 shows the spread between Eurodollar interest rates and two benchmark rates in the US money market: the rate on CDs issued by commercial banks and the rate on Treasury Bills. At first sight, the large spreads on CDs from the mid-1960s to the early 1980s – which on occasions greatly exceeded the differential allowed by the cost of reserve requirements on CDs in the USA – may cast doubts on the efficiency of US banks' arbitrage to maintain interest parity. However, the spread can be considered a measure of how effective were regulations and capital controls in keeping money market relatively segmented. As seen earlier, in the late 1960s, CDs rates were rigid under Regulation Q, and the prohibition to lend abroad shifted credit demand from the domestic to the "external" market (Formuzis 1973; Marston 1974). As soon as constraints were removed and the onshore regulatory environment became much more similar to the deregulated offshore environment, interest rate linkages were strengthened, US banks became net lenders, and the efficiency of arbitrage increased. This led to a secular reduction of the Eurodollar-CD spread, which remained close to zero since the late 1980s. In a similar fashion, the removal in 1979 of the British controls that restricted UK residents from investing in Eurosterling deposits in Paris led to a structural fall of the spread of Eurosterling rates on domestic deposit rates (Marston 1997, pp. 45–57).

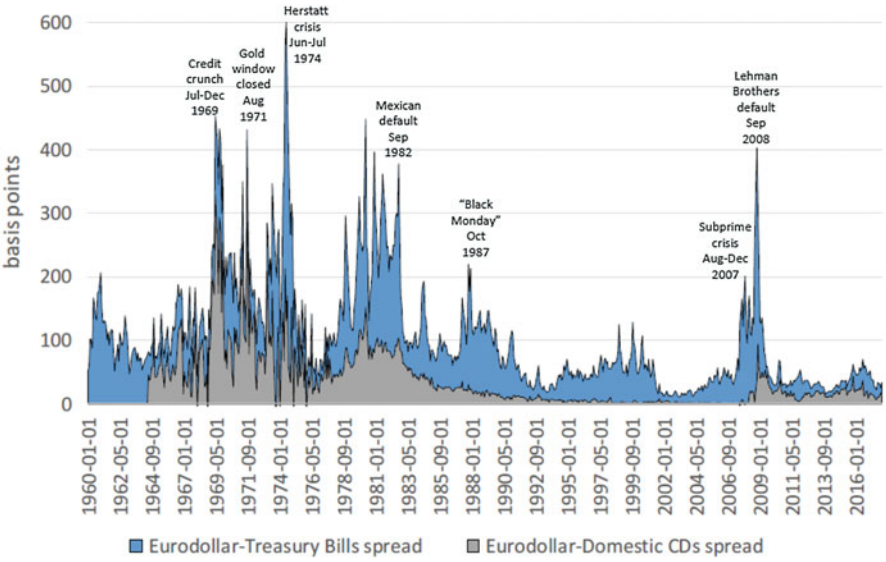


Fig. 8 LIBOR spreads over domestic CDs and US Treasury Bills. (Data from FRED (Federal Reserve Bank of St. Louis) online datasets)

However, spreads reflected also differences in riskiness. These were determined in part by structural factors – for instance, the fact that offshore dollar deposits were not only exposed to host-country political risk but also were not uniformly subject to US law, which generated legal uncertainty. Until recently, in lawsuits involving the expropriation or freezing of Eurodollar deposits due to political shocks, courts failed to agree on a single legal system ordering and regulating Eurodollar bank liabilities (Windecker 1993; Comizio and Chiachiere 2014). Shifts in perceived risk were another important factor. For instance, Marston (1997, pp. 56–57) explains the persistence of high Eurodollar-CDs spreads in 1974–1975 and 1980–1983 (which created scope for profitable outward arbitrage) with a higher risk premium demanded by depositors on Eurodollars due to greater default risk.

Another widely used indicator of perceived risk in the Eurodollar interbank market is the spread between LIBOR and the yield on US Treasury Bills – better known by its acronym TED. As already explained, money markets are exposed to counterparty risk, that is, the possibility that some of the debt-issuing banks might default on their contractual obligations. As a consequence, banks that rely massively on risk-intolerant market funding are especially vulnerable to liquidity risk. For money markets to develop a high level of liquidity, debt securities must be unanimously regarded by participants as safe assets – that is, close substitutes for money that can be taken at face value "with no questions asked." In terms of modern financial economics, a safe asset does not require investors to produce any information about its underlying value – it is "information insensitive." Bank debt (deposits) is an example: they are privately produced, money-like securities widely accepted as means of payments and short-term store of value "without extensive and costly verification of its value" in spite of the fact that its underlying collateral (the bank's loan portfolio) is risky. This is possible because bank debt has short maturities and must be rolled over regularly, which allow investors to check frequently the quantity and quality of the underlying assets, thus decreasing risk. Another factor is banks' opaqueness – the value of their loan portfolio is difficult to observe, so that outsiders have no incentive to produce private information. By reducing information asymmetries, opaqueness makes trading in bank money very liquid (Gorton 2017; Dang et al. 2017).

Even safe assets – such as bank debt – however are vulnerable to runs when rational doubts arise about the quality of the underlying collateral. Then, near money suddenly becomes information-sensitive and loses its money-like characteristics, the convertibility option loses its credibility, and holders scramble for "real" cash. A run on a systemic player in the market can be propagated to the rest of participants when the interbank portion of the money market is large – that is, when in a large share of transactions, banks act not only as intermediaries but also as borrowers from and lenders to other banks. Given the dense web of financial contracts negotiated in the interbank market, shocks that generate widespread uncertainty about the solvency of some systemically relevant counterparty or the quality of the collateral behind the securities traded can lead to a complete "freezing" of the interbank market, in which banks hoard liquidity and wholesale funds are rationed. Then, banks find themselves unable to roll over their short-term debt and cannot borrow even against

good-quality collateral – a situation of funding illiquidity in which the debt capacity of an asset falls to a small part of its fundamental value (Acharya et al. 2011) – nor can liquidate assets quickly with only a minor price impact (a situation of market illiquidity) (Brunnermeier and Pedersen 2009). In such circumstance, the TED spread increases as rates in interbank market remain high and a large number of investors switch their liquid funds into Treasury Bills – the safe asset par excellence – bringing their yields down.

Risk Factors

The TED spread suggests that the average level of perceived risk increased significantly in the 1970s compared to the previous decade. As a rule, the risk that US and foreign investors attached to Eurodollars was significantly higher than that of equivalent US deposits (with the exception of very safe deposits placed with foreign branches of top US banks). Deposits in the offshore market were not insured, and many of them were issued by foreign banks with no access to the lender-of-last-resort facilities of the Federal Reserve (Dufey and Giddy 1984, pp. 577–588). An additional source of perceived risk was maturity transformation. In its initial stages, Eurodollar transactions were relatively simple business that banks could carry out from their traditional foreign exchange offices, normally small departments offering trade finance services and dealing with correspondent banks. As the market matured, banks began to use interbank transactions also for domestic purposes, that is, to adjust their reserve position in domestic currency, to back loans to corporate customers (both in national or foreign currency), to support and make less dependent from national regulations their traditional foreign-exchange banking activities, and to undertake covered interest arbitrage in the foreign exchange market (i.e., covering in the Eurodollar market forward transactions undertaken on behalf of corporate customers). Thanks to the relatively stable nominal exchange rates of the 1960s, currency risk was low. Similarly, liquidity risk remained modest, as the share of interbank business increased and transactions with nonbank borrowers maintained their traditional short-time self-liquidating characteristics (McKinnon 1977, pp. 17–18; Davis 1979, pp. 82–86).

An upward shift in riskiness took place in the 1970s. Risk perceptions could be heightened by recurrent policy discussions about the introduction of unilateral or multilateral regulations and controls, the alarm raised by an increased frequency of bank failures, the persistent weakness of the US dollar during and after the breaking of the Bretton Woods system, and repeatedly voiced concerns about the possibility of a sudden contraction or collapse of the market. Volatile exchange rates increased currency risk and created scope for substantial losses from foreign exchange exposures. Higher and more volatile interest rates increased interest risk, as sudden fluctuations in the cost of short-term funding could cause losses on fixed rate loans with long maturities. Banks reacted by adopting marginal pricing in roll-over lending at floating rates, which adjusted assets to potential liabilities and separated interest risk from liquidity risk. However, this implied that interest risk was now

passed on to borrowers, thus harming their ability to service debts and adding to banks' credit and default risk for banks (Harrington 1987, pp. 46–48). At the same time, keener competition narrowed lending margins, and banks sought actively to expand profit margins by engaging in more maturity transformation and net liquidity creation. This could be achieved by funding loans of longer contractual maturity through short-term deposits, as well as by mismatching or short-funding roll-over deposit maturities in order to increase the profit yielded, which increased funding risk (Heinevetter 1979; Kane 1983, pp. 101–103). Data published by the Bank of England about the maturity mismatch of London banks in their Eurocurrency business confirm the existence of an upward trend in riskiness. Between 1973 and 1985, their share of foreign currency claims with maturities longer than 1 year increased from 13% to 24% of total claims, whereas liabilities with the same maturities had remained stable around 5% of total liabilities (Gibson 1989, pp. 226–233).

In this period the Eurodollar market developed also important links with the Eurobond market, where international bank syndicates managed, underwrote, and placed bonds issued by corporate and sovereign borrowers. Here, investment banks, securities firms, and commercial banks financed part of their underwriting commitments by borrowing in the Eurodollar and other Eurocurrency markets, especially when low short-term interest rates created large profit opportunities on Eurobond investments. After 1973, the fast growth of medium-term dollar lending to developing countries through syndicated loans pushed in the same direction. However, falling interest spreads on syndicated loans between 1977 and 1982 suggest that banks failed to fully appreciate the risk taken, possibly because of their increasing confidence in the depth and resilience of the Eurodollar market as a source of funding (Gibson 1989, pp. 160–195) or because implicit guarantees provided by governments of industrial countries and multilateral organizations such as the IMF generated a major problem of moral hazard (Altamura 2017, pp. 109–120).

The Crisis of Herstatt Bank 1974

The crisis of Herstatt Bank (a relatively small German bank involved in heavy foreign exchange speculation) in June 1974 is considered the first episode that brought to light the vulnerability of the Eurodollar interbank market to sudden shift in confidence. The liquidation of the troubled German bank caused a sudden contraction of liquidity, a sharp increase in interbank rates, and a widening of spreads across tiers (Busch 2009, pp. 100–102; Sarver 1988, pp. 28–30). As shown in Fig. 8, the TED spread jumped to 600 basis points in June and July 1974, signaling very severe money market pressure. The crisis also revealed the global range of interbank linkages through which contagion could be transmitted and showed the serious impact that relatively minor funding shocks could have on larger counterparties in key markets such as New York and London (Guttentag and Herring 1985). In the UK and West Germany, the crisis also exposed the serious limitations of traditional approaches to bank supervision, based on informality and mutual trust, when

confronted with the complex regulatory and jurisdictional issues raised by the growth of international money markets (Schenk 2014; Mourlon-Druol 2015). Yet, its effect was short-lived and did not represent any significant setback for the expansion of the market in the medium run; in fact, maturity mismatch in the market (the risk stemming from funding long-term assets on the base of short-term liabilities) increased in the years after the crisis (Gibson 1989, pp. 142–159).

This can be explained by the fact that national monetary authorities of G10 countries arranged emergency facilities in favor of domestic banks and collectively assured market participants that, without assuming formal responsibility for the market, they would be prepared and willing to provide lender-of-last-resort assistance "if and when necessary" (as stated in the Basle communiqué of September 1974) (Altamura 2017, p. 134). In the aftermath of the crisis, market participants upgraded internal checks and control mechanisms, and US banks extended the principle of corporate liability to assume legal responsibility for the exposure of their foreign branches and subsidiaries (Heininger 1979). With the Basle Concordat of 1975, central banks simply committed to improve monitoring and prudential supervision (Dufey et al. 1979; OECD 1985, pp. 48–72). However, as the recycling toward developing countries gained momentum in the second half of the 1970s, new alarms were raised about banks' overexposure to sovereign borrowers, the unsecured nature of syndicated lending, the rising degree of maturity transformation, the generalization of unsound banking practices, the risk of contagion through the interbank market, and the possibility of a "disorderly scramble for liquidity in Eurocurrency markets" (Bryant 1983, p. 29) in the absence of an international lender of last resort. New initiatives for a multilateral regulation of Eurodollar and Eurocurrency banking emerged in this period (see the next section), motivated both by policy and prudential issues, but failed to produce sufficient international consensus (Greenberg 1983; Hawley 1984).

The Debt Crisis of 1982

The outbreak of the debt crisis of developing economies materialized all the concerns of the previous years. The crisis, initiated by the Mexican moratorium of August 1982, brought US and European banks on the verge of collapse. The growth of the Eurocurrency market had been slowing down since 1980 and went into standstill in the second half of 1982, with the usual symptoms of rising market illiquidity: a strong contraction in interbank positions, shortening of interbank lending maturities, widening spreads across tiers, and a generalized upsurge in interest rates. In this occasion, the TED spread approached 400 basis points – a level already hit in various occasions since 1980. Latin American banks, which had heavily borrowed in the US and Eurocurrency interbank markets at short-term floating rates to fund longer-term loans at fixed rate, lost access to the market, and their default was avoided by freezing their interbank positions at the pre-crisis level for almost a decade as part of the process of negotiations on debt rescheduling.

The crisis also triggered a controversial policy discussion about who should bear the responsibility of supporting illiquid or insolvent banks from developing countries with branches operating in the New York and London money markets (Álvarez 2015, 2017). As already in 1974, the potentially systemic consequences of a crisis located in relatively peripheral corners of the international money market were brought clearly to the fore, but no widespread contagion was observed (Guttentag and Herring 1985). The response of national authorities was similar to what had been seen 10 years earlier: a new commitment to cooperate on the supervision of international banks, formalized by the Basel Committee on Banking Supervision in 1983, but with less practical consequences as far as sharing information among national supervisors was concerned (Schenk 2014, p. 1156).

The Liquidity Crisis of 2007–2008

The approval in 1988 of a set of new prudential regulations under the so-called Basel I agreement, together with the stable macroeconomic environment that characterized the Great Moderation, brought perceived risk in the Eurodollar market to its historical minimum in the 1990s and the early 2000s. Possibly for these reasons, the sudden freezing of the interbank Eurodollar market (and other international markets for short-term bank debt, such as repos and commercial paper) in 2007–2008 was perceived as a "black swan" – an extremely rare and totally unexpected event (Taylor and Williams 2009). This apparent insensitivity to risk factors is especially surprising in light of a growing evidence that structural changes in international banking had caused a major upward shift in riskiness. International reports (BIS 1986, 1992) warned that banks' growing exposure to off-balance sheet instruments had diminished their transparency and increased counterparty risk in interbank positions. Linkages between different sectors of the financing industry as well as between banks and nonfinancial firms had been strengthened, blurring the separation between interbank and general wholesale intermediation. Markets had become more concentrated on a relatively small group of highly rated banks, whose role as dealers and market makers had become more prominent, while at the same time increasing the dependence of smaller banks. Funding liquidity and market liquidity were problematic for a number of participants. Banks found pricing credit risk increasingly difficult, which led to a generalized used of quantity rationing (i.e., credit limits) and a lower propensity to support troubled counterparties. These trends spelt trouble in case of sudden shocks to confidence. Banks with complex positions in derivatives relied on a "presumption of liquidity in a number of markets" – a presumption that could easily become "illusory in time of stress, with a consequent impact on other markets" (BIS 1992, pp. 2–3).

Many of these elements emerged in the liquidity crisis of 2007–2008, the most recent example of money market panic. Twenty years earlier, Guttentag and Herring (1986, pp. 16–20) correctly anticipated that liability management, by increasing banks' dependence on the international interbank market, had dramatically increased both their vulnerability to funding shocks in case of bad news and the probability of a

"contagious loss of confidence" that could threaten the stability of the international banking system. The surprising absence of contagion in the debt crisis of 1982 however had allowed banks to maintain their "operating assumption … that the positive benefits from liability management outweigh[ed] the dangers." The crisis of 2007–2008 made this assumption obsolete. The sudden evaporation of global liquidity kept interest rates on interbank short-term lending unusually high and volatile for a long period, challenging established approaches to monetary theory and policy (Taylor and Williams 2009). As the "freezing" travelled through all international markets for interbank funding in dollars, the Fed was called upon to mitigate an incipient global "dollar shortage" by acting as an international lender of last resort. Foreign banks in the Eurodollar market used balances with their US correspondents as reserves; these in turn added pressure on the market for Federal Funds. In order to block this transmission channel, the Fed arranged foreign exchange swap lines with the most important foreign central banks, which allowed them to supply dollar liquidity to money market dealers and banks operating in their jurisdictions (McGuire and von Peter 2009; Fleming and Clagge 2010). The crisis of 2007–2008 brought to light the fact that the safety of the global dollar money market depends on the Fed's unconstrained ability to create both domestic and international dollars, so that "the whole world treats dollar deposits at the Fed not only as good as dollar currency, but also as the ultimate world reserve in a time of crisis" (Mehrling 2011, p. 29).

Political Economy

Origins of Eurodollars and the British State

The origins of the practice of "borrowing" and "lending" dollar deposits by British banks in the 1950s were initially surrounded by an aura of mystery. According to Paul Einzig, one of the most influential financial analysts of the time, its existence was initially "hidden from economists and other readers of the financial press by a remarkable conspiracy of silence," which banking circles in the city were "emphatically" fond of preserving (Einzig 1965, pp. vi-vii). This fear of publicity is not surprising in a context, such as the second half of the 1950s, dominated by a continuous pressure on a weak Sterling in foreign exchange markets. In fact recent research based on archival records show that initial concerns by British authorities, as they tried to grasp the nature and implications of this new practice, focused on its possible consequences for the external value of the pound.

The attitude of the Bank of England and the British Treasury toward the emerging Eurodollar, and more generally the relationship between the British state and the city, is the subject of a lively debate. Some interpretations contend that British authorities played an active role in promoting and encouraging the development of Eurodollar banking as a "conscious act of policy" to enhance London's ability to compete with New York as an international financial center. Special relevance is attributed to the outward-looking regulatory tradition that prevailed at the Bank of England. In this

perspective, they allowed Eurocurrency business of foreign and British banks (with the exception of domestic clearing banks) to grow free of binding regulations in order to allow the city to revive its historical role as an entrepôt center for international finance, although detached from the international role of the pound (Forsyth 1987, pp. 144–149; Helleiner 1994, pp. 83–84; Palan 1998, pp. 632–633; Schenk 2002; Saadma and Vaubel 2014).

Alternative interpretations assign British authorities a more passive role, as they simply tolerated the development of a new line of business that market actors had discovered and promoted by exploiting loopholes in the existing regulatory framework. Their permissive attitude was conditional to the absence of negative spillovers from Eurocurrency business to their overarching policy objectives. British clearing banks were subject to a severe regime of financial repression (Allen 2014); this included a cartel arrangement that capped deposit and loan interest rates and binding constraints (both qualitative and quantitative) on commercial banks' sterling asset portfolios. UK residents were also subject to binding capital controls that strongly limited their direct and portfolio investment abroad as well as their holdings of foreign currency deposits. Controls also prevented UK banks from lending in foreign currencies to residents. The purpose of controls was to prevent speculative pressures on the pound (a weak currency under continuous threat of depreciation in the 1950s and 1960s) and British international reserve, especially in light of the large sterling balances held by sterling area governments that could be converted into dollars (James 1996, pp. 185–186; Schenk 1994, pp. 33–35). Therefore the Eurodollar market in London was allowed to grow insofar as its developments did not threaten exchange rate stability, did not spill over on domestic credit conditions, or did not generate negative externalities for the role of the pound as an international currency (Capie 2010, pp. 182–185; Schenk 2010).

In its relationship with the banking system, the Bank of England continued to follow a supervisory approach based on gentlemen's agreements and moral suasion, to which the city usually responded with a cooperative attitude (Schenk 2004). Internal debates about its possible regulation through reserve requirements in the early 1960s led to a renewal of a vote of confidence in the city's ability to self-regulate (Schenk 1998, pp. 233–234; Burn 1999, pp. 240–243). The wave of bank failures of the mid-1970s in the USA and Western Europe, often caused by fraudulent behavior, shook mutual confidence between bankers and regulators. Their effects of contagion, transmitted through the global network of interbank linkages created by the Eurocurrency market, posed complex issues of supervision, regulation, and jurisdiction. However, the Bank of England remained wary of switching to statutory regulations in its relationship with the banking sector – a step it gave only at the end of the 1970s (Schenk 2014, pp. 1154–1155).

In a political economy perspective, Burn (2006, pp. 170–190) suggests that the debate about the ultimate responsibility for the development of Eurocurrencies – whether it was driven by markets in their attempt to escape the regulatory grip of the state or by a conscious policy action by the British authorities – is misleading. In the UK, the state-market boundaries were blurred, and the Bank of England acted as an interface between them ("poacher and gamekeeper" at one time). In fact, the

existence of the Eurodollar market revived an institutional structure that strongly resembled the state-bank-city nexus that had characterized the heyday of London as "the center of world liquidity" until 1931. Others (Green 2016) however argue that this interpretation is too inward-looking and neglects the deeper consequences of the Anglo-American financial and institutional interactions that made the post-1945 period profoundly different from the past. In this alternative view, the Eurocurrency market is seen as the "construction of an offshore political-economic space" led by British merchant banks and the Bank of England. In this space, the convergence between the "regulatory embrace" of British authorities and the "regulatory escape" of US financial interests weakened national monetary regimes and regulatory orders both in the UK and USA and strengthened Anglo-American institutional interdependence. The "Anglo-American financial synthesis" produced by the Eurodollar market in the city is seen here as the main driver of the reconfiguration of the international regulatory order promoted since the 1970s by the Basel Committee, which provided the bedrock of the more recent expansion of global finance.

The Failure of International Regulatory Coordination

The debate of the 1970s among central bankers over the regulation of Eurocurrencies sheds light on important aspects of the process through which this new international regulatory order was forged. As already discussed, the absence of regulation was a key determinant of the initial success of cross-border wholesale banking. However, whether or not it should remain unregulated soon became a major source of controversies. In the second half of the 1960s, Eurocurrencies were vehicles of large international flows of short-term capital, which made it more difficult for monetary authorities to achieve their objectives. We already mentioned how US banks successfully used the Eurodollar market to counteract the effects of the domestic credit crunches enforced by the Fed in 1966 and 1968–1969. In Europe, countries with strong currencies – such as West Germany – also found it increasingly difficult to pursue restrictive monetary policy, as high interest rates induced offsetting capital inflows (Porter 1972; Toniolo 2005, pp. 464–465). On the other hand, countries with weak currencies –such as Italy – found it more difficult to implement expansionary monetary policy without suffering capital outflows (Argy and Hodjera 1973, Argy and Kouri 1974).

Under the Bretton Woods system of adjustable pegs, the compatibility between fixed exchange rates and a degree of monetary autonomy – that is, the ability of central banks to gear monetary policy toward domestic objectives (for instance, to set interest and inflation rates at variance with those prevailing internationally) – rested on the effectiveness of capital control in reducing capital mobility (Obstfeld 1993, pp. 215–220). As capital flows' sensitivity to arbitrage and speculative opportunities increased, and responsiveness of stock adjustment to changes in information and expectations improved, the scope for national monetary independence was eroded (Obstfeld 1980; Obstfeld and Taylor 1998, pp. 391–393). Monetary control proved particularly difficult for small open economies, whose monetary

conditions were substantially affected by changes in monetary conditions in reserve countries, such as the USA, rapidly transmitted to the international money market, in which interest rate fluctuated sharply (Gibson 1989, pp. 34–35).

Policymakers viewed in the revival of short-term capital flows a proof that the international money market had turned into "an unregulated juggernaut 'out of control'." Central bankers complained that its "unbridled expansion and contraction" undermined their ability to control monetary and credit aggregates, caused instability in exchange rates and domestic interest rates, and enhanced the international transmission of inflation. As a consequence, they felt urged to "do something about the unregulated Eurocurrency market" (Bryant 1983, pp. 8–9 and 15). Their first response between the late 1960s and the mid-1970s had been the introduction of unilateral controls as a shelter against "disequilibrating short-term capital flows" (Argy 1971; Bordo 1993, pp. 55–60; Eichengreen 1996, pp. 120–122). Attempts to isolate domestic monetary and credit conditions from international disturbances took the form of direct regulation of banks' net external position (to limit outflows) or the prohibition of paying interest on nonresident bank deposits and the application of reserve requirements and mandatory cash deposits (*Bardepot*) on external borrowing (to limit inflows, as in the case of West Germany, Switzerland, and Japan, whose governments tried to limit the international use of their currencies) (Mills 1972; Marston 1997, pp. 45–69). The effectiveness of these measures was limited. In the West German case, for example, controls on banks were offset by nonbank transactions (as German companies resorted to massive external borrowing), while banks circumvented controls by transferring lending to German residents to their offshore branches and subsidiaries in Luxembourg. At that time the Bundesbank complained that the Eurocurrency system had turned into a sort of "substitute central bank" (*Monetäre Nebenregierung*) (Hewson and Sakakibara 1975, pp. 57–58; Emminger 1977; Obstfeld 1980, p. 22).

During their meetings at the BIS, G10 central bankers had been discussing the growth of the Eurodollar market and its possible implications for capital mobility and financial stability since the mid-1960s within the framework of multilateral surveillance, but the general sentiment was of benign neglect. The BIS also intervened to moderate the level and volatility of interest rates by drawing from its dollar swap lines with the Federal Reserve to increase the supply of deposits in the market in periods of strong demand. However, after the market's explosive growth of 1966–1969, concerns escalated, and the idea that coordinated actions were necessary in order to limit the impact of international banking activities on domestic monetary conditions gained consensus (Toniolo 2005, pp. 452–462; Yago 2013, pp. 163–172). Proposals for more forceful interventions were presented and discussed in the Standing Committee on the Euro-currency Market, created in April 1971 by G-10 central banks' governors. Its meetings became the main forum in which proposals of coordinated controls were debated.

Between 1971 and 1973, the struggle to preserve the Bretton Woods system of adjustable pegs was the background of the committee's discussions. The debate proved ultimately inconclusive as no agreement could be reached between conflicting stances of central banks (Toniolo 2005, pp. 465–469; Schenk 2010;

Altamura 2017, pp. 89–97). Initially, the Bundesbank and the Nederlandsche Bank (both presiding over strong currencies that attracted large capital inflows) pushed for a coordinated use of reserve requirements to constrain the growth of the Eurocurrency "monster." The Bundesbank also insisted on the need to better coordinate monetary policies if controls on the Eurocurrency market were to be effective. The Banque de France and the Bank of Italy supported the view that surveillance should be tightened and controls "at both ends" should be introduced. On the other side, the Bank of England denied the necessity of regulatory interventions, emphasized the efficiency of the system, reaffirmed its confidence on the "judgment and self-discipline" of market actors, and insisted that the introduction of coordinated regulations at G-10 level would simply displace the market from London to offshore centers. US authorities were initially sympathetic with the Bundesbank (in 1969 the Fed had introduced reserve requirements on US banks' borrowing from their foreign branches in the Eurodollar market), but in the end supported the British view that coordinated controls were unnecessary and possibly counterproductive. The Committee's discussions had reached a standstill, when the currency crisis of March 1973 and the generalized transition to floating exchange rates took most of the heat out of the issue. A few months later, the first oil shock and the huge volume of international liquidity necessary to finance the large balance-of-payments deficits of oil-importing countries made the Eurocurrency market suddenly look like a resource rather than a problem.

The issue of coordinate intervention on the Eurocurrency markets re-emerged in the late 1970s in a different macroeconomic context. While economists generally agreed that it played a relatively marginal role in the creation of new money and credit, with modest inflationary consequences (Niehans and Hewson 1976; Mayer 1979; Grabbe 1982), some central bankers switched from benign neglect to a more energetic regulatory and supervisory approach. In the USA, the fast expansion of the Eurodollar market was regarded as a key factor behind the confidence crisis of the dollar in 1978 (Hawley 1984, pp. 145–154). One year later, with the monetarist reorientation of monetary policy, US authorities switched from a Keynesian demand-management approach to an explicit anti-inflationary approach (Hetzel 2008, pp. 150–171; Meltzer 2009, pp. 1025–1033). In order to restore its credibility in the fight against inflationary expectations and the pursuing of price stability, the Federal Reserve abandoned interest rate targeting and began to use monetary aggregates as intermediate targets – an approach followed also by the German Bundesbank and the Swiss National Bank in Europe. Here, again, the Eurodollar market complicated its task. Similarly to other money market innovations of the period (such as repurchase agreements and money market funds), Eurodollar deposits allowed investors to place funds outside the domestic banking system without limiting its capacity to generate deposits. Being free from reserve requirements (which had been introduced in 1969 but lifted in 1974, since they were easily circumvented), Eurodollars, by decreasing the effective reserve ratio, provided an additional "avenue for the growth of credit" not directly constrained by the supply of federal reserves (Aliber 1980; He and McCauley 2010, p. 9). As already shown (Fig. 5), by then the volume of external dollar claims vis-à-vis nonbanks – a measure

of the credit generated by the Eurodollar market – was already equivalent to more than 10% of the US money stock as measured by M3. As dollars held abroad were not counted in the US monetary aggregates, unpredictable changes in Eurodollars reduced the usefulness of conventionally defined money stock targets (Frydl 1982). Moreover, in the 1970s the number of branches of foreign banks in the USA had expanded enormously, to the extent that in 1979, they held 40% of commercial and industrial loans in the balance sheets of banks in New York and California. Yet, most of them operated only in the wholesale and money markets, and their dominant source of funding was the global interbank market (Buttrill White 1982). The International Banking Act of 1978 had brought foreign banks under the supervision of the Fed, but banks and nonfinancial institutions still enjoyed large scope to exploit "ambiguities in the domain of surveillance by different central banks and governments," which made the effects of monetary policy less predictable (Hester 1981, pp. 153–155).

To fix the problem of monetary control, the Fed contemplated alternative options. One was taking Eurodollars into account in the targeting of monetary aggregates, which however met technical difficulties. The alternative was to put the external dollar sector on an equal foot with the domestic sector; this would eliminate the risk-adjusted differential in favor of Eurodollars and the incentive for US residents to use them, so that the growth rate of the domestic and external component of money would converge. Since unilateral reserve requirements had failed to achieve this objective, the idea of coordinated controls was brought back to life. In May 1979 the Fed and the Bundesbank launched their joint proposal of "a uniform Euro-reserve requirement" on Eurocurrency intermediation (both in dollars and other currencies, to limit scope for further regulatory arbitrage), which kicked off an intense debate among different committees at G10 level (Johnston 1983, pp. 248–278; Clement and Maes 2015). At the same time, a Eurocurrency Market Control Act was introduced in the US Congress, which envisioned a system of temporary Eurocurrency reserves elaborated in coordination with the Fed, to be gradually phased out in 3 years (Helleiner 1994, pp. 135–139; Saadma and Vaubel 2014, pp. 339–357; Altamura 2017, pp. 199–203). In parallel, the Monetary Control Act (finally passed in 1980), which aimed at extending reserve requirements to all depository institutions (including savings and loans banks and credit unions, which were not members of the Federal Reserve System), was being discussed in Congress. By then, the Fed had also agreed on the plan proposed by New York bankers for the establishment of International Banking Facilities (previously discussed) as a way to attract back onshore part of the recycling mechanism that until then had been intermediated through London and other offshore centers. This, the Fed hoped, would give US representatives an upper hand in the international negotiations with Britain (Hawley 1984, p. 156; Helleiner 1994, p. 138). However, the proposal met strong opposition domestically from American bankers, who lobbied energetically against it, and internationally from central banks of countries hosting large financial centers (the UK, Switzerland, and Luxembourg). Eventually the idea of an international regulatory cartel was abandoned (Hawley 1984, pp. 155–160). As an alternative to the regulatory approach promoted by the Fed and the Bundesbank, the Bank of England

proposed a macroprudential approach aiming at slowing down the growth of international bank lending through the adoption of prudential measures, such as limits on banks' maturity transformation and exposure to exchange and country risk, the adoption of special balance sheet provisions for involuntarily rescheduled loans, and capital ratios. This approach did not meet sufficient support either (Clement and Maes 2015). In the end, the mountain of G10 central bankers gave birth to a mouse in the form of a Euromarkets committee in charge of "regular and systematic monitoring of international banking developments" (Johnston 1983, p. 278).

The failure of international attempts to introduce coordinated controls on Eurocurrencies has been explained as a consequence of conjunctural factors – i.e., shifting policy priorities in the wake of the two oil shocks – and structural changes, i.e., a profound redefinition of the relationships between markets and the state and a rising confidence on the markets' ability to self-regulate. The defeat of the Eurocurrency regulatory proposals of the early and late 1970s is also given a prominent role not only as a locus of the clash between opposed visions of international finance but also as a turning point in the competition between states on deregulation and the rise of financial globalization (Saadma and Vaubel 2014; Altamura 2017, pp. 247–254). However, a realistic assessment should acknowledge the nature of public good of international cooperation and its exposure to free riding by individual nations, which prevents them from pursuing "cooperative responses to systemic global problems" (Bryant 1983, p. 35). It should acknowledge the "formidable" technical problems that the implementation of the reserve requirements scheme would have encountered (such as the uniformation of bank reporting or jurisdictional conflicts on the regulation of overseas subsidiaries and consortium banks) and the disintermediation effects that noninterest-bearing reserves would generate for G10 banks (which would require a global and uniform scheme) (Johnston 1983, p. 277). It should also consider the degree of coercion required to impose cooperation on reluctant participants, the incentives of governments and banks to break from the agreement which undermined its credibility, and the difficulties to impose sanctions on countries that failed to abide by its terms. Against this background, "it [was] extremely unlikely that an international agreement [was] a feasible solution to the regulatory dilemma created by Eurodollars" (Greenberg 1983, p. 1510).

Moreover, Eurocurrencies were just one of the manifold manifestations of the increasing financial interdependence among national economies. As such, they reflected, but were not at the origins of deeper macroeconomic forces that affected a globalizing economy, such as shifts in investors' confidence in different currencies, differences in monetary policy objectives and preferences, and adjustments in asset portfolios driven by an expanding set of investment opportunities. As Bryant (1983, pp. 15 and 36) posited, "those problems [would] not go away regardless of what might conceivably be done to regulate the Eurocurrency market," as "inhibiting one channel … [was], by itself, merely likely to force the interdependence into other channels." As a consequence, even a successful implementation of Eurocurrency regulations would "generate false expectations about what could and should be accomplished." The Eurocurrency market "acted primarily as an international

transmission mechanism"; although it might "blunt the effectiveness of certain domestic monetary policy instruments" or "exacerbate conflicts between external and domestic requirements," the only feasible solution to these problems was "greater international coordination of policies and instruments" (Mayer 1979, p. 65). As the post-2007–2009 policy debate on regulatory reforms of international finance revived the interest in possible multilateral schemes for Eurocurrency reserve requirements (Fowler 2014, pp. 856–859), a critical assessment of the experience of the 1970s would be instructive.

Conclusions

In 1959 an obscure new kind of business in which a bunch of British banks in the city of London "placed" and "accepted" short-term dollar deposits spawned for the first time the interest of the financial press. In the following 60 years, Eurodollars and other Eurocurrencies continued to hit the headlines as powerful agents of structural changes in international finance. They promoted the adoption of an entirely new approach to bank intermediation based on wholesale banking and liability management. This created unprecedented scope for leverage, which allowed banks to expand balance sheets faster than ever, but also immensely increased their vulnerability to shocks. Eurocurrencies favored the distribution and creation of international liquidity while at the same time creating new channels for the international transmission of monetary and financial disturbances. They allowed the financing of enormous balance-of-payments imbalances but also generated moral hazard on global scale. They were praised as a paradigm of efficient markets and blamed as a source of systemic instability. Their history is a privileged prism through which the virtues and vices of financial globalization can be observed and assessed.

Cross-References

- ► International Currencies in the Lens of History
- ► International Monetary Regimes: The Bretton Woods System
- ► The Asian Dollar Market
- ► The Evolution of Monetary Policy (Goals and Targets) in Western Europe
- ► The Evolution of US Monetary Policy

References

Accominotti O, Ugolini S (2019) International trade finance from the origins to the present: market structures, regulation and governance. CEPR discussion paper DP 13661. Centre for Economic Policy Research, London

Acharya V, Gale D, Yorulmazer T (2011) Rollover risk and market freezes. J Financ 66(4):1177–1209

Agmon T, Barnea A (1977) Transaction costs and marketability services in the Eurocurrency money market. J Monet Econ 3(3):359–366
Aliber R (1980) The integration of the offshore and domestic banking system. J Monet Econ 6(4):509–526
Allen WA (2014) Monetary policy and financial repression in Britain 1951–1959. Palgrave Macmillan, Basingstoke
Allen WA, Moessner R (2011) The international liquidity crisis of 2008–2009. World Econ 12(2):183–198
Altamura CE (2017) European banks and the rise of international finance: the post-Bretton Woods era. Routledge, London
Altman OL (1961) Foreign markets for dollars, sterling, and other currencies. IMF Staff Pap 8(3):313–352
Altman OL (1963) Recent developments in foreign markets for dollars and other currencies. IMF Staff Pap 10(1):48–96
Altman OL (1967) Eurodollars. Fund Bank Rev 4:9–16
Álvarez S (2015) The Mexican debt crisis redux: international interbank markets and financial crisis, 1977–1982. Financ Hist Rev 22(1):79–105
Álvarez S (2017) International money markets, liquidity risk and financial cooperation in the wake of the Latin American debt crisis of 1982. UPIER working paper no 17/1. Department of History, University of Oxford
Álvarez S (2019) Mexican banks and foreign finance: from internationalization to financial crisis, 1973–82. Palgrave Macmillan, Basingstoke
Argy V (1971) Monetary policy and internal and external balance. IMF Staff Pap 3:508–525
Argy V, Hodjera Z (1973) Financial integration and interest rate linkages in industrial countries, 1958–1971. IMF Staff Pap 20(1):1–73
Argy V, Kouri P (1974) Sterilisation policies and the volatility in international reserves. In: Aliber RZ (ed) National monetary policies and the international financial system. The University of Chicago Press, Chicago, pp 209–230
Battilossi S (2000) Financial innovation and the golden ages of international banking:1890–1931 and 1958–1981. Financial History Review 7(2):141–175
Battilossi S (2002a) International banking and the American challenge in historical perspective. In: Battilossi S, Cassis Y (eds) European banks and the American challenge. Cooperation and competition in international banking under Bretton Woods. Oxford University Press, Oxford, pp 1–35
Battilossi S (2002b) Banking with multinationals. British clearing banks and the Euromarket challenge, 1958–1974. In: Battilossi S, Cassis Y (eds) European banks and the American challenge. Cooperation and competition in international banking under Bretton Woods. Oxford University Press, Oxford, pp 103–134
Battilossi S (2010) The Eurodollar revolution in financial technology. Deregulation, innovation and structural change in Western Banking in the 1960s–70s. In: Ksyrtis A (ed) Financial markets and organizational technologies. System architectures, practices and risks in the era of deregulation. Palgrave Macmillan, Basingstoke, pp 29–63
Baxter ND (1966) Commercial banks and the new generation of corporate treasurers. The Banker Magazine (Winter)
Baxter ND (1968) Marketability, default risk and yields on money market instruments. J Financ Quant Anal 3(1):75–85
Benston GJ (1964) Interest payments on demand deposits and bank investment behavior. J Polit Econ 72(5):431–449
Bernanke B, Bertaut C, DeMarco L, Kamin S (2011) International capital flows and the returns to safe assets in the United States, 2003–2007. Federal Reserve Board International Finance discussion papers no 1014. Board of Governors of the Federal Reserve, Washington D.C.
Bignon V, Flandreau M, Ugolini S (2012) Bagehot for beginners: the making of lending-of-last-resort operations in the mid-19th century. Econ Hist Rev 65(2):580–608

BIS (1983) The international interbank market. A descriptive study. BIS economic papers no 8. BIS, Basel
BIS (1986) Recent innovations in international banking. BIS, Basel
BIS (1992) Recent developments in international interbank relations. BIS, Basel
BIS (2015) Introduction to BIS statistics. BIS Q Rev (September):35–51
Bordo MD (1993) The Bretton Woods international monetary system: a historical overview. In: Bordo M, Eichengreen B (eds) A retrospective on the Bretton Woods system. Lessons for international monetary reform. The University of Chicago Press, Chicago, pp 3–98
Bordo MD, McCauley RN (2017) Triffin: dilemma or myth? BIS working papers no 364. BIS, Basel
Borio C, Disyatat P (2011) Global imbalances and the financial crisis: link or no link? BIS working paper no 346. BIS, Basel
Brunnermeier M, Pedersen L (2009) Market liquidity and funding liquidity. Rev Financ Stud 22(60):2201–2238
Bryant RC (1983) Eurocurrency banking: alarmist concerns and genuine issues. OECD Econ Stud (1) 1(1):7–41
Bryant RC (1987) International financial intermediation. The Brookings Institution, Washington, DC
Burn G (1999) The state, the city and the Euromarkets. Rev Int Polit Econ 6(2):225–261
Burn G (2006) The re-emergence of global finance. Palgrave Macmillan, Basingstoke
Busch A (2009) Banking regulation and globalization. Oxford University Press, Oxford/New York
Buttrill White B (1982) Foreign banking in the United States: a regulatory and supervisory perspective. Federal Reserve Bank of New York Q Rev 7(Summer):48–58
Calem P (1985) The new bank deposit markets: goodbye to Regulation Q. Federal Reserve Bank of Philadelphia Bus Rev (November/December):11:19–29
Capie F (2010) The Bank of England: 1950s to 1979. Cambridge University Press, Cambridge
Cassard M (1994) The role of offshore centers in international financial intermediation. IMF working paper 94/107. International Monetary Fund, Washington, DC
Clement P, Maes I (2015) Peripheral visions of economic development. New frontiers in development economics and the history of economic thought. In: Garcia-Molina M, Trautwein H-M (eds) The BIS and the Latin American debt crisis of the 1980s. Routledge, London, pp 203–227
Coates BA (2018) The secret life of statutes: a century of the Trading with the Enemy Act. Mod Am Hist 1(2):151–172
Comizio VG, Chiachiere R (2014) "Ringfencing" US bank foreign branch deposits: working toward a clearer understanding of where deposits are payable in the midst of chaos. Am Univ Bus Law Rev 3(2):249–275
Crockett AD (1976) The Eurocurrency market: an attempt to clarify some basic issues. IMF Staff Pap 23(2):375–386
Dang TV, Gorton G, Holmström B, Ordoñez G (2017) Banks as secret keepers. Am Econ Rev 107(4):1005–1029
Davis SI (1979) The management function in international banking. Macmillan, London
De Cecco M (1987) Inflation and structural change in the Euro-Dollar market. In: De Cecco M, Fitoussi JP (eds) Monetary theory and economic institutions. Macmillan, London, pp 182–208
Dufey G, Giddy IH (1978) The international money market. Prentice Hall, Englewood Cliffs
Dufey G, Giddy IH, Min S (1979) Interest rates in the US and Eurodollar markets. Weltwirtsch Archiv 115:51–67
Dufey G, Giddy IH (1984) Eurocurrency deposit risk. Journal of Banking and Finance 8(4): 567–589
Eichengreen B (1996) Globalizing capital. A history of the international monetary system. Princeton University Press, Princeton
Eichengreen B (2011) Exorbitant privilege. The rise and fall of the dollar and the future of the international monetary system. Oxford University Press, Oxford/New York
Eichengreen B (2012) International liquidity in a multipolar world. Am Econ Rev 102(3):207–212

Eichengreen B, Chitu L, Mehl A (2016) Stability or upheaval? The currency composition of international reserves in the long run. IMF Econ Rev 64(2):354–380
Einzig P (1960) Dollar deposits in London. Banker 110:23–27
Einzig P (1964) The Euro-dollar system. Practice and theory of international interest rates. Macmillan/St. Martin's Press, London/New York
Einzig P (1965) Foreign dollar loans in Europe. Macmillan/St. Martin's Press, London/New York
Einzig P (1971) Parallel money markets. V.1 The new markets in London. Macmillan/St. Martin's Press, London/New York
Ellis JC (1981) Eurobanks and the interbank market. Bank of England Q Bull 21(3):351–364
Emminger O (1977) The D-Mark in the conflict between internal and external equilibrium. Essays in international finance, no 122. Princeton University, Princeton
Flandreau M, Jobst C (2005) The ties that divide: a network analysis of the international monetary system, 1890–1910. J Econ Hist 65(4):977–1007
Flandreau M, Ugolini S (2013) Where it all began: lending of last resort and Bank of England monitoring during the Overend-Gurney panic of 1866. In: Bordo MD, Roberds W (eds) The origins, history and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, New York, pp 113–161
Fleischer V (2010) Regulatory arbitrage. Tex Law Rev 89:227–289
Fleming MJ, Clagge NJ (2010) The Federal Reserve's foreign exchange swap lines. Federal Reserve Bank of New York Curr Issues Econ Financ 16(4):1–7
Formuzis P (1973) The demand for Euro-dollar and the optimum stock of bank liabilities. J Money Credit Bank 5(3):806–818
Forsyth JH (1987) Financial innovation in Britain. In: De Cecco M (ed) Changing money. Financial innovation in developed countries. Basil Blackwell, London, pp 141–157
Fowler SA (2014) The monetary fifth column: the Eurodollar threat to financial stability and economic sovereignty. Vanderbilt J Transnatl Law 47:825–860
Fratianni M, Savona P (1972) The international monetary base and the Eurodollar market. Kredit und Kapital (Special Issue) n.1, Proceedings of the First Konstanzer Seminar on Monetary Theory and Monetary Policy, pp. 347–409
Frenkel JA, Levich RM (1975) Covered interest arbitrage: unexploited profits? J Polit Econ 83(2):325–338
Frenkel JA, Levich RM (1977) Transaction costs and interest arbitrage: tranquil versus turbulent periods. J Polit Econ 85(6):1209–1226
Friedman M (1971) The Euro-dollar market: some first principles. Review of the Federal Reserve Bank of St. Louis 53 (July): 16–24. Federal Reserve Bank of St. Louis MO.
Friedman BJ (1975) Regulation Q and the commercial loan market in the 1960s. J Money Credit Bank 7(3):277–296
Frydl EJ (1982) The Eurodollar conundrum. Federal Reserve Bank of New York Q Rev 7(1):12–21
Galpin W, Resnick B, Shoesmith GL (2009) Eurocurrency risk premia. Int J Bus 14(3):199–220
Gibson HD (1989) The Euro-currency markets, domestic financial policy and international instability. Macmillan, London
Goodfriend M (1981) Eurodollars. Correct reference: Federal Reserve Bank of Richmond Economic Review 67(3): 12–18. Federal Reserve Bank of Richmond, VA.
Goodfriend M (1998) Eurodollars. In: Instruments of the money market. Federal Reserve Bank of Richmond, Richmond, pp 48–58
Gorton G (2017) The history and economics of safe assets. Annu Rev Econ 9:547–586
Grabbe JO (1982) Liquidity creation and maturity transformation in the Eurodollar market. J Monet Econ 10(1):39–72
Green J (2016) Anglo-American development, the Euromarkets, and the deeper origins of neoliberal deregulation. Rev Int Stud 42:425–449
Greenberg RD (1983) The Eurodollar market: the case for disclosure. Calif Law Rev 71(5): 1492–1515

Grosse LG, Goldberg R (1991) Foreign bank activity in the United States: an analysis by country of origin. J Bank Financ 15:1093–1112
Guttentag JM, Herring RJ (1985) Funding risk in the international interbank market. In: Ethier WJ, Marston RC (eds) International financial markets and capital movements. Essays in international finance no 157. Princeton University, Princeton, pp 19–32
Guttentag JM, Herring RJ (1986) Disaster myopia in international banking. Essays in international finance no 164. Princeton University, Princeton
Harrington R (1987) Asset and liability management by banks. OECD, Paris
Hawley JP (1984) Protecting capital from itself: US attempts to regulate the Eurocurrency system. Int Organ 38(1):131–165
He D, McCauley RN (2010) Offshore markets for the domestic currency: monetary and financial stability issues. BIS working paper no 320. BIS, Basel
He D, McCauley RN (2012) Eurodollar banking and currency internationalization. BIS Q Rev (June):33–46
Heinevetter B (1979) Liquidity creation in the Euromarkets. J Money Credit Bank 11(2):231–234
Heininger P (1979) Liability of US Banks for deposits placed in their foreign branches. Law Policy Int Bus 11:903–1034
Helleiner E (1994) States and the reemergence of global finance. From Bretton Woods to the 1990s. Cornell University Press, Ithaca
Hester D (1981) Innovations and monetary control. Brook Pap Econ Act 1981(1):141–189
Hetzel R (2008) The monetary policy of the Federal Reserve: a history. Cambridge University Press, Cambridge
Hewson J, Sakakibara E (1975) The Eurocurrency markets and their implications. Lexington Books, London
Holmes AR, Klopstock FH (1960) The market for dollar deposits in Europe. Federal Reserve Bank of New York Mon Rev 42(11):197–202
Houpt JV (1999) International activities of U.S. banks and in U.S. banking markets. Fed Reserv Bull 85(9):599–615
Houston JF, Lin C, Ma Y (2012) Regulatory arbitrage and international bank flows. J Financ 67(5):1845–1895
James H (1996) International monetary cooperation since Bretton Woods. Oxford University Press/International Monetary Fund, Oxford/New York/Washington, DC
Johnston RB (1981) Theories of the growth of the Eurocurrency market: a review of the Eurocurrency deposit multiplier. BIS economic papers, BIS, Basel, no 4. pp 8–27
Johnston RB (1983) The economics of the Euro-market. History, theory and policy. Macmillan, Basingstoke/London
Kane EJ (1979) The three faces of commercial bank liability management. In: Dooley MP et al (eds) The political economy of policy-making. Sage, Beverly Hills/London, pp 149–174
Kane DR (1983) The Euro-dollar market and the years of crisis. Croom Helm, London
Key SJ, Terrell HS (1989) The development of International Banking Facilities. In: Park YS et al (eds) International banking and financial centers. Kluwer, Dordrecht, pp 191–217
Klopstock F (1965) The international money market: structure, scope and instruments. J Financ 20(2):182–208
Klopstock F (1970) Money creation in the Euro-dollar market: a note on Prof. Friedman's view. Federal Reserve Bank of New York Mon Rev 8(1):12–15
Koch C (2015) Deposit interest rate ceilings as credit supply shifters: bank level evidence on the effects of Regulation Q. J Bank Financ 61:316–326
Kreicher LL (1982) Eurodollar arbitrage. Federal Reserve Bank of New York Q Rev 7(2):10–23
Lewis MK, Davis KT (1987) Domestic and international banking. Philip Allan, Oxford
Machlup F (1970) Euro-dollar creation: a mystery story. Banca Nazionale del Lavoro Q Rev 25(94):219–260
Makin JH (1972) Demand and supply functions for stock of Eurodollar deposits: an empirical study. Rev Econ Stat 2:381–391

Marston RC (1974) American monetary policy and the structure of the Euro-dollar market. Princeton studies in international finance, no 34. Princeton University, Princeton
Marston RC (1976) Interest arbitrage in the Eurocurrency markets. Eur Econ Rev 7(1):1–13
Marston RC (1997) International financial integration. Cambridge University Press, Cambridge
Masera R (1972) Deposit creation, multiplication and the Euro-dollar market. Ente Einaudi Quaderni di Ricerche no 11. Ente Einaudi, Rome, pp 129–189
Mayer HW (1979) Credit and liquidity creation in the international banking sector. BIS economic papers no 1. BIS, Basel
Mayer T (1982) A case study of Federal Reserve policymaking: Regulation Q in 1966. J Monet Econ 10(2):259–271
Mayer HW (1985) Interaction between the Euro-currency markets and the exchange markets. BIS economic papers, no 15. BIS, Basel
McCauley RN (2005) Distinguishing global dollar reserves from official holdings in the United States. BIS Q Rev 57–72
McGuire P (2004) A shift in London's Eurodollar market. BIS Q Rev 67–71
McGuire P, von Peter G (2009) The US dollar shortage in global banking and the international policy response. BIS working paper no 291. BIS, Basel
McKinnon RI (1977) The Euro-currency market. Essays in international finance no 125. Department of Economics, Princeton University, Princeton
Mehrling P (2002) Economists and the Fed: beginnings. J Econ Perspect 16(4):207–218
Mehrling P (2011) The new Lombard Street. How the Fed became the dealer of last resort. Princeton University Press, Princeton
Meltzer AH (2009) A history of the Federal Reserve. Vol 2, Book 1 1951–1969; and Book 2 1970–1986. The University of Chicago Press, Chicago/London
Mills RH (1972) The regulation of short term capital movements in major industrial countries. Board of Governors of the Federal Reserve System staff economic studies, no 74. Board of Governors of the Federal Reserve System, Washington D.C.
Mourlon-Druol E (2015) 'Trust is good, control is better': the 1974 Herstatt Bank crisis and its implications for international regulatory reform. Bus Hist 57(2):311–334
Niehans J (1982) Innovation in monetary policy. Challenge and response. J Bank Financ 6(1):9–28
Niehans J (1984) International monetary economics. Johns Hopkins University Press, Baltimore/London
Niehans J, Hewson J (1976) The Eurodollar market and monetary theory. J Money Credit Bank 8(1):1–27
Obstfeld M (1980) Sterilisation and offsetting capital movements: evidence from West Germany, 1960–70. NBER working paper no 494. NBER, Cambridge MA
Obstfeld M (1993) The adjustment mechanism. In: Bordo M, Eichengreen B (eds) A retrospective on the Bretton Woods system. Lessons for international monetary reform. The University of Chicago Press, Chicago/London, pp 201–268
Obstfeld M, Taylor A (1998) The Great Depression as a watershed: international capital mobility over the long run. In: Bordo M et al (eds) The defining moment. The Great Depression and the American economy in the twentieth century. The University of Chicago Press, Chicago/London, pp 353–402
OECD (1985) Trends in banking in OECD countries. Report to the Committee on Financial Markets. OECD, Paris
Palan R (1998) Trying to have your cake and eating it: how and why the state system has created offshore. Int Stud Q 42(4):625–643
Pecchioli RM (1983) The internationalisation of banking. OECD, Paris
Porter M (1972) Capital flows as an offset to monetary policy: the German experience. IMF Staff Pap 19(2):395–424
Ross D (2002) Clubs and consortia: European banking groups as strategic alliances. In: Battilossi S, Cassis Y (eds) European banks and the American challenge. Cooperation and competition in international banking under Bretton Woods. Oxford University Press, Oxford, pp 135–160

Ruebling CE (1970) The administration of Regulation Q. Federal Reserve Bank of St. Louis (February). Federal Reserve Bank of St. Louis MO, pp 29–40
Saadma T, Vaubel R (2014) The emergence and innovations of the Eurodollar money and bond market: the role of openness and competition between states. In: Bernholz P, Vaubel R (eds) Explaining monetary and financial innovation. A historical analysis. Springer, New York/Dordrecht/London
Sarver E (1988) The Eurocurrency market handbook. New York Institute of Finance-Prentice Hall, New York
Schenk CR (1994) Britain and the sterling area. From devaluation to convertibility in the 1950s. Routledge, Abingdon/New York
Schenk CR (1998) The origins of the Eurodollar market in London: 1955–1963. Explor Econ Hist 35:221–238
Schenk CR (2002) International financial centres 1958–1971: competitiveness and complementarity. In: Battilossi S, Cassis Y (eds) European banks and the American challenge. Cooperation and competition in international banking under Bretton Woods. Oxford University Press, Oxford, pp 74–102
Schenk CR (2004) The new city and the state in the 1960s. In: Michie R, Williamson P (eds) The British government and the City of London in the twentieth century. Cambridge University Press, Cambridge, pp 322–339
Schenk CR (2010) The decline of sterling: managing the retreat of an international currency 1945–1992. Cambridge University Press, Cambridge
Schenk CR (2014) Summer in the city: banking failures of 1974 and the development of international banking supervision. Engl Hist Rev 129(540):1129–1156
Shaw ER (1978) The London money market. Heinemann, London
Shin HS (2011) Global banking glut and loan risk premium. Jacques Polak annual research conference, November 2011. https://www.imf.org/external/np/res/seminars/2011/arc/pdf/hss.pdf
Stigum M, Crescenzi A (2007) Stigum's money market. McGraw-Hill, New York
Swoboda AK (1968) The Euro-dollar market: an interpretation. Essays in international finance, no 64. Department of Economics, Princeton University, Princeton
Sylla R (2002) US banks and Europe: strategy and attitudes. In: Battilossi S, Cassis Y (eds) European banks and the American challenge. Cooperation and competition in international banking under Bretton Woods. Oxford University Press, Oxford, pp 53–73
Taylor JB, Williams JC (2009) A black swan in the money market. Am Econ J Macroecon 1(1):58–83
Toniolo G (2005) Central bank cooperation at the Bank for International Settlements, 1930–1973. Cambridge University Press, Cambridge
Werner RA (2016) A lost century in economics: three theories of banking and the conclusive evidence. Int Rev Financ Anal 46:361–379
Windecker GH (1993) The Eurodollar deposit market: strategies for regulation. Am Univ Int Law Rev 9(1):357–384
Woynilower A (1980) The central role of credit crunches in recent financial history. Brook Pap Econ Act 1980(2):277–339
Yago K (2013) The financial history of the Bank for International Settlements. Routledge, London

The Asian Dollar Market

12

Seung Woo Kim

Contents

Abstract

The Asian Dollar Market functions as a regional center for Eurodollars in Asia-Pacific. From its establishment in 1968, it mobilized the surplus US dollars in the region and then served the demand for capital from developing countries and national and multinational corporations in Asia. As a result, Singapore emerged as an international financial center. This chapter introduces the history of the Asian Dollar Market and analyzes factors behind the success of

S. W. Kim (✉)
Graduate Institute of International and Development Studies,
Geneva, Switzerland
e-mail: kim.seungwoo@graduateinstitute.ch

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_56

Singapore. It sees the emergence of Asian dollars from the hybridization of the following elements: the colonial experience of Singapore as an entrepôt for Southeast Asia under the British Empire, the independence of Singapore and the making of a nation-state with an international financial center, and the role of foreign bankers who brought the idea of the Asian Dollar Market. It pays attention to the deliberate efforts by the government of Singapore, particularly the Monetary Authority of Singapore, as a host agent to invite foreign banks and Eurodollars in Asia. The authority continuously and incrementally lifted regulations concerning foreign exchanges and introduced tax structure favorable to international capital. Then this chapter examines the evolution of the Asian Dollar Market – the Asian Bond Market – and the negotiable certificate of deposits denominated in Asian dollars. It concludes with the impact of the Asian Dollar Market on the economy of Singapore and the Asia-Pacific region and the market's importance in the history of international finance of the twentieth century.

Keywords
Asian Bond Market · Asian Currency Unit · Asian Dollar Market · Bank of America · International financial center · Monetary Authority of Singapore · Offshore market · Singapore

Introduction

The Asian Dollar Market, by definition, is a regional market for Eurodollars, expatriate US dollars, located in Singapore. It includes various Eurodollar-denominated financial instruments such as Asian bonds of long-term loans and Asian dollar certificate of deposits of midterm maturities. The scope of the market went beyond the US dollar-denominated financial instruments later, increasingly including other Eurocurrencies – deposits of currencies held outside their home market. From its establishment in 1968 by the decision of the government of Singapore to mobilize Eurodollars in Southeast Asia, the Asian Dollar Market aimed to be a funding center of the region. It also connected Singapore with other international financial centers, the City of London, Tokyo, and San Francisco, making the global network of Eurodollar business.

As a regional center for Eurodollars, the history and evolution of Asian Dollar Market are comparable to and a part of the history of international financial centers, notably that of the City of London – the main market for Eurodollar business. The Asian Dollar Market also shares distinctive features in Eurodollar transactions such as the unknown end user and double-counting, which hinders the calculation of precise market size (Kim 2019). The term "Asian," in this regard, merely indicates the market's location. However, the Asian Dollar Market presents different features of the European case. In the market's development, the government of Singapore, particularly the Monetary Authority of Singapore, offered various measures favorable to foreign capital which was in search of a

safe haven. It also progressively lifted regulations on foreign exchange market. Furthermore, the political stability of Singapore attracted international banks and multinational corporations. In return, foreign banks with expertise in the Eurodollar market brought innovative practices into the nascent market. As a result, Singapore became an offshore banking center in the Asia-Pacific region.

Therefore, the origin of the Asian Dollar Market is rather clear than that of the Eurodollar market in the City of London, which is under contestation. For the newly independent small city-state of Singapore, the making of an international financial center was a part of the project for economic development and the nation-building process. In the history of the offshore market in the late twentieth century, Singapore presents the tradition of the British Empire, and the English common law and a part of offshore archipelagos emerged in the Global South (Ogle 2017). And as a local funding source of Eurodollars, the Asian Dollar Market has been not only in a complementary position with other centers by facilitating capital flows but also in competition with them – the City of London and Hong Kong. In 1976, less than a decade, the number of banks in the Asian Dollar Market reached 69, whose total assets and liabilities exceeded 17 billion US dollars. (Hodjera 1978, p. 226). About a half-century later from the establishment, its continuous growth indicates the success by reaching 1.2 trillion US dollars (Burner 2016, p. 120) (Table 1).

This chapter introduces the history of the Asian Dollar Market until the late 1970s. From its launching in 1968, the market received attention from practitioners, monetary authorities, international organizations, and academics. In the early days, they examined the regulatory circumstances which led to the development and expansion of the Asian Dollar Market and the rise of Singapore as an international financial market in Asia (Singapore International Merchant Bankers Ltd 1973; Tan 1982; Lessard 1994). The approach not only compared the two Eurodollar markets in London and Singapore but also examined the economic growth of Singapore as a city-state in the 1970s. The scholarship of international political economy considers the Asian Dollar Market within the context of the emergence of tax havens and offshore banking centers in the former British colonies (Park 1982; Shaxon 2011; Palan 2010; Palan et al. 2013). More recently, economic historians offer archive-

Table 1 Comparison of the net size of the Eurocurrency market and Asian Currency Units assets/liabilities, 1968–1975 (millions of US dollars) (Source: Bhattacharya 1977, p. 10)

Year	Net Eurocurrency market size	Asian Currency Unit assets and liabilities	Asian Currency Unit as a percentage of the Euromarket
1968	30,000	31	0.10
1969	44,000	123	0.28
1970	57,000	390	0.68
1971	71,000	1,063	1.50
1972	92,000	2,976	3.20
1973	132,000	6,277	4.80
1974	177,000	10,357	5.60
1975	205,000	12,597	6.10

based accounts which include the following: the competition of Singapore with Hong Kong in the making of an international financial center with Asian dollars and political history of the Asian Dollar Market within the context of the Cold War and postcolonialism (Schenk 2017; Kim 2018). This chapter is divided into two parts: it charts the historical development of the Asian Dollar Market with an emphasis on the role of the government of Singapore and evaluates the market's impact upon the domestic and the Asia-Pacific economy.

A Note on Sources

It is worthwhile to note a common feature in the literature – the paucity of sources for further research. Early studies on the Asian Dollar Market were based on interviews with practitioners and monetary officials. The official archives, particularly that of the Monetary Authority of Singapore, the chief regulator of the Asian Dollar Market, are not available. Instead, the National Archives of Singapore offers the online oral history interviews whose collection includes interviews with practitioners and monetary officials on the economic development of Singapore such as Yap Siong Eu and Joseph H Greene. Their eyewitness accounts provide motifs, rationale, and inside stories behind the making of the Asian Dollar Market in Singapore. Also, documents acquired from the US National Archives and Records Administrations housed in the national archives are of valuable assistance (Kim 2018). Other than the oral records, the following materials provide relevant sources and observations: contemporary periodicals (e.g., *The Banker*, *Euromoney*, *The Bankers Magazine*, *Far Eastern Economic Review*, *Malayan Economic* Review), bank reviews (e.g., Bank of America, First National City Bank, Moscow Narodny Bank), and research papers by academic institutions (Lee 1979), monetary authorities, and international organizations. For example, the Organization for Economic Cooperation and Development (OECD) and the International Monetary Fund (IMF) published comprehensive studies on the Asian Dollar Market (Wellons 1977; Hodjera 1978). Recently, scholars have introduced a multi-archival study by examining original documents from commercial banks, central banks, and international institutions which took part in the development of the Eurodollar market in Asia. Documents from the following archives have been used: central banks (Federal Reserve Bank of New York, Bank of England, and Bank of Korea), commercial banks (Credit Lyonnais and HSBC), and international organizations (IMF, Bank for International Settlements, and OECD).

The Legacy of the British Empire: An Entrepôt for Southeast Asia

One of origins of the rise of Singapore as an international financial center with Asian dollar transactions is the historical experience as a trading post under the British Empire in the nineteenth century. Owing to its geographical location

in Southeast Asia, it functioned as a focal point for trade routes in the region. It effectively bridged economic activities and cultures across Europe and Asia. In 1819, Stamford Raffles, a British statesman, founded Singapore as an international trading post of the British East India Company to secure the control over the Straits of Malacca. Since then, Singapore had been a gateway to China from the West (Burner 2016, p. 113). And the status of a crown colony under the British Empire in 1867 further strengthened the port city's standing of an entrepôt for commercial activities in the region, which provided a foundation for financial operations in Singapore. In this regard, the legacy of the British Empire, agglomeration of banking skills, English common law, and the language, is attributable to the city-state's emergence as an internationally oriented financial market. The historical ties with the City of London and international bankers provided not only access to a pool of capital for financing trade but more importantly transferred banking skills and techniques. The tradition of English common law which had protected property rights and contract enforcement strengthened the competitive position of Singapore in the region after the demise of British Empire (Burner 2016, p. 112). Furthermore, it proved to be flexible for continued innovations by the banking sector. The English language was undoubtedly an advantage in hosting international trade and facilitating cross-border financial services. As a result, despite the changing status of Singapore under the British rule from a crown colony with the autonomy to the further granting of the status of a state under the British Commonwealth, the imperial link with Britain remained intact and attracted Western banks, particularly Americans, to Singapore. Palan considers that Singapore is one of former British colonies such as Hong Kong, the Channel Islands, and the Bahamas which share features of offshore markets – city-state status, under the British control, and the local financial sector run indirectly from the City of London (Palan 2010, p. 171). A characteristic feature of these small jurisdictions, with a lesser degree in Hong Kong, is that while they maintain strong ties with British banks, the legal status of independence or crown dependency allows them to exercise discretion over international financial transactions beyond the control of the Bank of England. These institutional and cultural legacies culminated into favorable conditions for Singapore to emerge as a regional financial center. Still, as explained in detail below, these factors are short of comprehensively explaining the introduction of business in Eurodollars in Singapore in the late 1960s. In fact, Hong Kong was the international financial center for the Pacific Basin with more presence of foreign banks. Then a question arises concerning the competition between the two small city-states in Asia. As Cassis suggests in his seminal work *Capitals of Capital*, Singapore's evolution into an international financial center was accompanied by a systematic effort by the national government of Singapore, after its separation from Malaysia in 1965 (Cassis 2006, p. 276). The political project of making a newly independent nation in the context of geopolitical tensions in Southeast Asia lies behind the development of the Asian Dollar Market.

The Independence of Singapore and the Making of an International Financial Center

After the Japanese occupation during the Second World War from 1942 to 1945, the United Kingdom granted Singapore the status of a crown colony under the British Commonwealth. Except for the matters of defense and external relations, the government of Singapore exercised self-governing autonomy. In 1963, it merged with the Federation of Malaysia but underwent a violent conflict within the union for the fear of a Communist take-over of the island. On 9 August 1965, the Parliament of Malaysia voted to expel Singapore from the Federation, and immediately the Parliament of Singapore declared the foundation of a new sovereign state by passing the Republic of Singapore Independence Act. Singapore, now a new country, faced a task to build a nation. In his famous broadcast press conference to announce the secession of his state from Malaysia, Prime Minister Lee Kuan Lee emphasized to build an independent nation-state (Kim 2018). Beyond the racial and ideological tensions in the multicultural community of Singapore, another essential task for survival was to establish a thriving economy when the economic ties with Malaysia were lost.

However, Singapore lacked the means to achieve economic growth independent from Malaysia. It did not possess hinterland for industry, and natural and human resources for manufacturing were absent. Moreover, the geopolitical factors threatened the stability of Singapore. The threat of Indonesia, which opposed the creation of the Federation of Malaysia, demanded the withdrawal of the Western military in Southeast Asia. The conflict in Vietnam was already escalating. Furthermore, the British government announced the scheduled withdrawal of its troops from Singapore. The new nation lacked a military force to protect itself. Prime Minister Lee returned from his tour to London in 1968 only with a bitter experience (Josey 1968).

The economic ties with Britain, furthermore, were in trouble with the decline of the sterling area. Under the monetary bloc, Singapore pegged its currency to the sterling pound and had been a satisfactory member by holding 90% of its reserves in sterling (Kim 2018). However, Singapore lost faith when recurring sterling crises questioned the currency's stability and reliability as a reserve currency. In 1967, according to an interview record, the Bank of America, on behalf of the government of Singapore, profited from the 1967 sterling devaluation in the foreign exchange market (Schenk 2017, p. 5). This event marked the changing preference of Singapore from sterling to the US dollar as a reserve currency and encouraged local traders to hold the US currency.

In such geopolitical and economic circumstances, the government of Singapore identified trading and financial services sector a key driver of economic development in the real economy and a growth industry in its own right (Woo 2016, p. 26). In other words, Singapore saw a vision of it to become a global city which would invite as many countries as possible and intermediate their economic activities regardless of political differences. A global city for international trade and finance, by connecting interests of ideological belligerents, was expected to grant Singapore a status of neutrality and provide measures of protection. In this sense, the making of

an international financial center had political implications for a small city-state (Kim 2018). Many political leaders such as Sinnathamby Rajaratnam, the country's first Minister for Foreign Affairs, often expressed the globalist approach (Woo 2016, p. 26). Features of Singapore satisfied the requirements for the market: the free-enterprise, pro-management orientation of the People's Action Party led by Prime Minister Lee Kuan Yew. Lee was not only outspoken about the globalist approach but also took a strategy to emulate Switzerland, a neural state in Europe with an international financial center (Kim 2018). Accordingly, as explained in detail below, the government of Singapore introduced various measures to attract foreign capital or relaxed regulations. The strategy was successful; in its 1973 report, the First National City Bank, an American bank and one of the early participants of the Asian Dollar Market, depicted the development of Singapore as the Zurich of the East (Burner 2016, p. 119).

The Role of Foreign Bankers: The Bank of America and the Idea of the Asian Dollar Market

It took more than the willingness of the government of Singapore to make an international financial center: experienced foreign banks in international finance introduced the idea of Eurodollars from the City of London. The square-mile financial district in London incubated the offshore market for US dollars in the late 1950s (Schenk 1998; Burn 1999, 2006). Despite continued suspicion on its speculative nature and future prospect by practitioners and central bankers, the Bank of England protected the nascent market from attempts to introduce regulations by monetary authorities (Kim 2019). As the regulator of the City, the British central bank saw the Eurodollar market developed in its jurisdiction would facilitate the revival of the City as the international financial center (Palan 2010, p. 162). By the mid-1960s, the Eurodollar market was "here to stay" and further attracted foreign banks, particularly Americans, into the banking district in London (see Battilossi and Cassis 2005). In addition to the growth in the market's volume, City bankers introduced innovative techniques to extend maturities of short-term Eurodollars into Eurobonds (long-term) and Euro-credits (midterm). Throughout the decade, the global financial market under no single monetary authority's jurisdiction facilitated the cross-border capital movements under the Bretton Woods of the international monetary system. And the time was ripe for the innovative idea of Eurodollars to reach other regions in the world by international bankers.

The Bank of America, one of the leading American banks with branches in London, saw the potential of a regional market for Eurodollars in Asia. In the 1960s, many economies of countries in the region began to flourish. The demand for capital and a funding center to finance trade and economic development increased across the region. Overseas Chinese merchants already held a substantial amount of US dollars. Furthermore, owing to the entry of the American military into the conflict in the Indochina War, more expatriate US dollars were in circulation. Despite the question on the standing of US dollar as a reserve currency in the Bretton

Woods system, many governments preferred to hold it in their central banks' vaults; sterling was in retreat, and there were no alternatives (Kim 2018).

It was not Singapore, however, but Hong Kong that the Bank of America contacted with the idea of Eurodollars. The rationale was quite simple; it was the international financial center of Asia with virtues of a connection with the City of London, the British legacies of common law and English language, and a small city. It had enjoyed a long history of international banking with the backing of the Bank of England. The American bank proposed the idea of the Asian Eurodollar market to the colonial government of Hong Kong. Yet, surprisingly the local authority declined the offer. From the 1950s, it had imposed restrictions on the financial sector and refused to lift the interest withholding tax on foreign currency deposits, a condition to effectively attract foreign capital (Palan 2010, p. 172). Schenk also documents that the Financial Secretary J. J. Cowperthwaite of Hong Kong feared of its jurisdiction to become tax havens and saw the danger that the new foreign currency market would drain liquidity from the domestic market (Schenk 2017, p. 12). Moreover, unlike Singapore, Hong Kong was under the direct control of Britain with a strong commitment to the sterling area (Kim 2018).

On the contrary, the government of Singapore was willing to accept a plan to make an international financial center in its jurisdiction. Two international bankers substantiated the idea – J. D. van Oenen of the Bank of America and Albert Winsemius, a Dutch economist on a mission by the United Nations in the early 1960s, who had been an economic advisor to the government of Singapore. In the early 1960s, there emerged business opportunities for foreign exchange operations in the Southeast Asia region from the Indonesia-Malaysia Confrontation of 1963, a violent conflict under which Indonesia opposed the creation of Malaysia. Holders of Indonesian rupiah who were in search of safe assets against the currency's devaluation converted local currencies into Singapore or US dollars or sterling pound to be deposited in Singapore. The Bank of America conducted the operation for its customers with its Eurodollar scheme. And it took substantial funds from overseas Chinese in the region when the interest rate of US dollar increased. Interestingly, the geopolitical origins of the Asian Dollar show some similarity with the Eurodollar; the Communist China transferred its US dollar assets to a Soviet-owned bank in Paris during the Korean War; the Soviet Union took the similar step in the 1950s, particularly after its invasion of Hungary in 1956 (Kim 2018). van Oenen, who had been in a foreign exchange clerk and supervised the Eurodollar business for the Bank of America branch in London, saw a business opportunity in the refugee funds in Southeast Asia. He was promoted to the leadership of the Bank of America in Singapore and established a good relationship with the government. For example, as mentioned above, during the Sterling Crisis of 1967, his bank made a fortune for the Singapore government by running a sterling overdraft in London. Then the international banker asked the government for permission to collect dollar deposits from Southeast Asia and suggested that the government should set up an Asian currency market and make Singapore the center. In 1971, a banker from the Overseas-Chinese Banking Corporation credited that the Asian Dollar Market was the brainchild of van Oenen (Loen 1971, p. 80).

Another important foreign expert was Albert Winsemius, who inspired the government to develop Singapore as an international financial center. When he contacted van Oenen for advice, the vice president of the Bank of America in Singapore presented a vision of Singapore a regional hub for Eurodollars. With its time zone advantage, van Oenen explained the city-state would facilitate 24-h banking around the globe. In this scheme, Singapore could not only bridge the gap between East and West but also connect international financial centers – London, Hong Kong, Tokyo, and San Francisco. For example, when the London market closes, Singapore opens, and during its operation hours, one may make deals with Japan. By the end of the day in Singapore's financial market, the US begins in San Francisco, for example, which can also reach New York during its opening hours (Burner 2016, p. 116).

The Launch of a Regional Eurodollar Market in Singapore: The Asian Currency Unit and the Asian Dollar Market

The hybridization of three factors – the legacy of the British Empire, the willingness of the government of Singapore to attract foreign capital, and the banking expertise from foreigners – culminated into a new regional center for Eurodollar business. By 1967, moreover, the domestic unrest in Hong Kong made Singapore attractive for flight capital. The standing of Singapore as an independent state was another advantage for it allowed the government's discretionary action to introduce measures favorable to foreign capital. Also, Prime Minister Lee made a political decision to take the risk of the disquiet of the Bank of England, which concerned the new foreign exchange business in Singapore would precipitate the disintegration of the sterling bloc (Lee 2000, p. 89). The conditions for the promotion of the Asian Dollar Market in Singapore were ready.

In 1967, the Parliament of Singapore passed the Currency Act of 1967 to establish the Board of Commissioners of Currency and take steps to create Singapore's central bank – the Monetary Authority of Singapore – in 1971. In November 1968, the Asian Dollar Market was instituted when the government of Singapore began to license a branch of the Bank of America to set up a special department to handle nonresident currency transactions. The license granted the American bank to operate the Asian Currency Unit – a special department conducting Eurodollar operations in Singapore with the following conditions:

(a) They may accept deposits from any source other than residents of Singapore without the prior approval of the authorities.
(b) They may lend to companies or individuals outside the scheduled territories without the prior approval of the authorities.
(c) They may lend to companies or individuals within the scheduled territories with the approval of the exchange control of the country concerned.
(d) They may not lend to residents of Singapore without the prior approval of the Singapore authorities (Attwood 1971, p. 52).

It is possible to identify two intentions by the government of Singapore in the Asian Currency Unit – the desire for a full-fledged financial center and the need for a responsive domestic banking system (Wellons 1977, p. 354). The exemption of Asian Currency Unit from exchange controls stimulated offshore banking, which would mobilize the spare US dollars held by Asians (Euromoney 1969, p. 29). The authority also abolished the 10% withholding tax on interest from nonresident foreign deposits in 1968. On the other hand, it separated local and foreign business by residents of Singapore from nonresident transactions by setting up a separate accounting unit for authorized banks (Hodjera 1978, p. 223). It sought to maintain the price stability of the domestic economy from capital inflows and prevent the channeling of domestic savings outward. The authority was selective in granting license to foreign banks; it preferred giants in international banking to reduce the likelihood of liquidity crisis. Licensed banks were required to open branches, rather than subsidiaries, so that the bank's headquarters and home country would take the ultimate responsibility (Wellons 1977, p. 359). In this regard, the government of Singapore carefully devised the Asian Currency Unit to create a controlled offshore market in its jurisdiction.

By the end of 1969, in addition to Singaporean banks, many international banks received the Asian Currency Unit license – Chartered Bank, Bank of Tokyo, First National City Bank, Chase Manhattan, Hong Kong and Shanghai Bank, and General Bank of the Netherlands. They mobilized untapped Eurodollars hitherto in the region whose volume was reported to have reached a few hundred million US dollars, and American banks took the lion's share of the market. However, in its early days, the future of the Asian Dollar Market received skepticism about its independence from the Eurocurrency market and its viability. It was still regarded as an appendage of the Eurodollar market in London. Contemporary commentators accused that the new market was another device to siphon local savings to Europe or North America owing to the higher interest rates in the Eurodollar market. Others claimed that the offshore market merely attracts hot money or encourage the evasion of currency control regulations by wealth investors. In other words, they problematized the term "Asian" by questioning the contribution of the nascent market to the local and regional economy (Kim 2018).

In the early 1970s, the influx of international bankers into Singapore continued; they captured the imagination of international finance in Singapore. The initial flow of funds from Asia to the Eurocurrency market reversed, and savings from Hong Kong, Indonesia, Philippines, and Taiwan were channeled into borrowers in the Pacific Basin such as Australia and South Korea. Deposits from Europe and the Middle East increased, which were then supplied to "Asian" customers. The Asian Dollar Market was an attractive source of capital for developing countries in Asia such as South Korea which had to endure higher interest rates in the City of London. Countries without a giant international bank or a bank with a less global reputation opened representative offices in Singapore instead. Multinational and national corporations in the petroleum, shipbuilding, and basic metals industries entered into the Asian Dollar Market not only to enjoy a geographical position that provided contacts with the rest of Asia but also tap into the pool of

liquid capital of Asian dollars. The presence of new customers and foreign banks conferred legitimacy of Singapore and the Asian Dollar Market as a funding center for productive industries of the region, not that for speculations, and strengthened the "Asian" characteristic of the market. Tables 2 and 3 show the list of foreign banks and representative offices of foreign banks in Singapore in 1972.

Furthermore, the government of Singapore granted the entry of banks from the communist bloc in the financial district. For example, the Moscow Narodny Bank

Table 2 Foreign banks in Singapore in 1972 (Source: Attwood 1972, p. 6)

Country of origin	Number of offices (in total)	Names
China	6	Bank of China, Kwangtung Provincial Bank
France	2	Banque de l'Indochine, Banque Nationale de Paris
Germany	2	Asien-Pazifik Bank, Dresdner Bank
Holland	1	Algemene Bank Nederland NV
Hong Kong	13	Bank of Canton Ltd., Bank of East Asia Ltd., Hong Kong & Shanghai Banking Corp., Mercantile Bank Ltd.
India	6	Bank of India, Indian Bank, Indian Overseas Bank, United Commercial Bank
Italy	1	Banca Commerciale Italiana
Japan	3	Bank of Tokyo Ltd., Mitsui Bank Ltd.
Malaysia	31	Ban Hin Lee Bank Bhd, Kwong Lee Bank Bhd, Malayan Bank Bhd, United Malayan Banking Corp.
Pakistan	1	Habib Bank Ltd.
United Kingdom	21	Chartered Bank, Moscow Narodny Bank
United States	11	American Express International Banking Corporation, Bank of America NT & SA, Chase Manhattan Bank, First National Bank of Chicago, First National City Bank
Total	98	

Table 3 Representative offices in foreign banks in Singapore in 1972 (Source: Attwood 1972, p. 7)

Country of origin	Name
Australia	Bank of New South Wales, National Bank of Australia Ltd.
Indonesia	Bank of Indonesia
Italy	Banca Nazionale del Lavoro
Japan	Dai-Ichi Kangyo Bank Ltd., Sanwa Bank Ltd., Sumitomo Bank
Korea (South)	Korea Exchange Bank
Philippines	Philippine National Bank
Switzerland	Swiss Bank Corporation, Swiss Credit Bank, Union Bank of Switzerland
United States	First National Bank of Dallas, First Pennsylvania Banking & Trust Company, Irving Trust Co., Marine Midland Bank, National Bank of Commerce of Seattle, Republic National Bank of Dallas, United California Bank

opened a branch in Singapore in November 1971 for trade financing between the USSR and other socialist countries in Southeast Asia. The Soviet-owned but registered as a British bank in the City of London had been one of the pioneering practitioners in the Eurodollar market. It not only connected the communist bloc and the international financial market in the West but also contributed to the innovation of Euro-credit facilities in the mid-1960s (Kim 2018). It is not surprising that the communist bank received the Asian Currency Unit license and participated in the Asian Dollar Market. By 1974, it successfully competed with American banks and became one of the major foreign banks. The coexistence of banks from belligerents of the Cold War symbolically represented the success of the political leadership of Singapore – the globalist approach and the neutrality with international financial services (Kim 2018).

The Asian Dollar Market dealt the completely homogenous commodity of offshore dollars; the only difference was the location of the market. Accordingly, the Eurodollar market in London exerts influence on its regional market in Singapore (Jang 1972, p. 15). The interest rates of these two markets proved their correlations. The benchmark interest rate of the Asian Dollar Market, the Singapore Inter-Bank Offered Rate (SIBOR), followed the London Inter-Bank Offered Rate (LIBOR), taking the former as a base with a fractional difference and an occasional intersection. The two markets also demonstrate parallel growth, despite the faster rate of the Asian Dollar Market than the Eurodollar Market (Wellons 1977, p. 362).

Efforts by the Singapore authorities to develop a regional offshore market for US dollars continued. It progressively lifted regulations to render the entry into the Asian Dollar Market by residents and companies. It exempted stamp duties on financial instruments to improve the competition position of Singapore against other centers in Asia such as Hong Kong. Further relaxations on the foreign exchange market were introduced, as Table 4 indicates. The Parliament of Singapore legislated the Banking Act of Singapore in 1970 and permitted banks to use numbered accounts to allow foreign depositors to hide their identity. It was regarded to have been a help to the Asian Dollar Market (Lee 1971, p. 48). From 1971, a principal regulator with ambition for the Asian Dollar Market took the responsibility – the Monetary Authority of Singapore. The regulator aimed to create the right climate for private sectors to contribute to the development of financial center in Singapore; it introduced legislation and offered administrative monitoring (Burner 2016, p. 119; Tan 1982, p. 35).

Established by the Currency Act of 1967, the Parliament of Singapore authorized the Monetary Authority of Singapore to be the banker and financial agent of the government and a promoter of monetary policy and conditions conducive to the economic growth (Effros 1985). It also took over regulatory authorities from different government agencies of the Commissioner of Banking, the Accountant General's Office, and the Controller of Foreign Exchange and consolidated them under a single agency. Also, it assumed roles for a host central bank to conduct prudential supervision on the operations by foreign financial institutions in Singapore (Bryant 1989, p. 348). With regard to the Asian Dollar Market, it

Table 4 List of key developments by Singapore authorities relevant to the development of the Asian Dollar Market (Sources: Tan 1982, p. 42; Schenk 2017, p. 11)

August 1968	Abolition of the withholding tax on interest income of nonresidents, laying the precondition for the establishment of the Asian Dollar Market
November 1968	Bank of America was given approval and commences operating the first ACU
September 1970	License was given to a foreign bank to operate in Singapore, the first in 6 years. This marks the shift in policy regarding the entry of new foreign banks and financial institutions
January 1971	Establishment of the Monetary Authority of Singapore, thus centralizing various monetary functions, other than currency issue, under one body
June 1971	Elimination of preferential exchange control among sterling area countries
December 1971	The issue of the first Asian dollar bond by the Development Bank of Singapore
January 1972	Abolition of the 20% liquidity ratio which ACUs were required to maintain against deposits. This brings the competitive position of ACUs more in line with banks operating in the Eurocurrency market
November 1972	Passing of the Monetary Authority of Singapore (Amendment) Act by Parliament which enables the Monetary Authority to provide lender of last resort facilities
1973	Reduction of income tax on interest earnings from offshore loans from 40% to 10%
March 1973	Offshore banking licenses issued to 7 foreign banks
July 1973	Singapore dollar floated and the cartel system for foreign exchange quotes removed. Banks were allowed to freely quote their own rates for foreign exchange
November 1977	Issue of floating rate US dollar negotiable certificates of deposit by two Japanese banks in the Asian Dollar Market
January 1978	Twenty-five banks authorized to issue fixed rate US dollar negotiable certificates of deposit
June 1978	Exchange control is completely liberalized

continuously relaxed relevant regulations to encourage international banks' entry. For example, in June 1972, it lifted the 20% liquidity requirements for Asian Currency Units for smaller international operators (Tan 1982, p. 36). By regulating the issuing of Asian Currency Unit license, it was able to exert influence over and led the orderly stimulation of the Asian Dollar Market. In order to avoid undue competition, it categorized the license and limited banks' access to the market. Only the first 13 foreign commercial banks opened before 1970 was given the full license. From 1971 to 1973, restricted licenses were granted to new entrants, which allowed their Asian Currency Unit operations but not in domestic banking (Hodjera 1978, p. 224). In March 1973, it introduced the concept of offshore banks to facilitate the entry of international banks which would bring foreign funds to nonresidents or local industry only under the explicit approval by the Monetary Authority of Singapore (Schenk 2017, p. 7). These regulatory measures stimulated the growth in the number of banks with Asian Currency Unit license and the size of the Asian Dollar Market (Asian Wall Street Journal 1978, p. 9).

In the same year, the Monetary Authority of Singapore granted license to merchant banks. In doing so, the authority hoped to bring their expertise and contacts into Singapore (Tan 1982, p. 35). And the regulatory relaxation such as the favorable tax structure on incomes from the offshore banking culminated in the abolition of exchange controls in 1978, which liberalized the foreign exchange market (Table 3).

The Development of Asian Dollars: The Asian Bond Market and the Negotiable Certificate of Deposits

A Capital Market for Asia: Asian Bond Market

As in the case of the Euromarket in the City of London, the next step for Singapore as a regional center for "Asian" dollars was the formation of a capital market to meet local demands for long-term bonds. The Eurobond market, a market for long-term obligation denominated in Eurodollars, was invented by innovative bankers in the City, such as Siegmund Warburg (Ferguson 2009). However, not surprisingly, the government of Singapore exhibited the state-led finance development strategy to encourage the formation of the Asian Bond Market. In December 1971, Sui Sen, the Finance Minister of Singapore, announced a plan to mobilize Asian dollars for long-term capital demands. The move was based on the confidence that the Asian Dollar Market could attract funds from Europe and North America rather than the reverse. Still, the growth of the Asian Bond Market was sluggish in the early 1970s, as Table 5 shows. Once again, unlike the Eurobond market, the government ushered the issues by having the Development Bank of Singapore to float the Asian Currency Unit bond in December 1971. The government owned substantial share of the bank, and its guarantee to the issue led to the success marked by the oversubscription by international and regional financial institutions. With the Development Bank of Singapore, the government guaranteed another Asian bond for the Private Investment Company for Asia, a

Table 5 Asian bond issues, 1971–1980 (in million) (Source: Lee 1986, p. 215)

Year	Number of issues	Total amount	Year	Number of issues	Total amount
1971	1	US$ 10	1976	9	US$ 247 DM 50
1972	2	US$ 20 DM 100	1977	14	US$ 315 DM 100 $A 10
1973	3	US$ 100	1978	12	US$ 220 SDR 25 ¥ 15,000 $A15
1974	–	–	1979	8	US$ 315 DM 70
1975	3	US$ 47	1980	18	US $659

multinational company for capital investment in Asia. With more issues floated, the Eurodollar market in London now supported lendings in the Asian Bond Market (Kim 2018). By 1972, the main customers of the Asian Bond Market were from Asia and Australia. Furthermore, the government introduced another measure to promote the long-term capital market – the exemption of withholding tax on nonresidents' interest from Asian dollar deposits.

In its early days, the government of Singapore took the lion's share of the capital market for Asian dollars. Later, it appealed countries in Asia in search of funding sources such as South Korea which attempted to issue Asian bonds in 1973. There was no issue in 1974 owing to foreign exchange losses and the failure of Herstatt Bank of West Germany. The abolition of capital controls and the interest equalization tax in the United States, which allowed US corporations' borrowing in the domestic capital market, was expected to shrink the size of the Asian Bond Market. In the following year, Michael Wong Pakshong, managing director of the Monetary Authority of Singapore, however, expressed his optimism on the future of the Asian Dollar Market from the continued entry of American banks and new inflows of petrodollars. The director considered that his agency's strict supervision on foreign exchange markets and the control over the Asian Currency Unit keep the market from the upheaval (Pakshong 1975: pp. 101–102). The Authority arranged loans for Japanese and Philippine banks in difficulties of securing US dollars from Europe. Yet, the attempt to attract petrodollars by Singapore officials was limited for the oil-producing countries in the Middle East still preferred London to Singapore (Bhattacharya 1977, p. 80). Since 1976, the Asian Bond Market returned in consonance with the Eurobond market and commonly offered floating interest rates on the basis of LIBOR or SIBOR (Lee 1986, pp. 214–15). As noted in a contemporary periodical, the Asian currency market had a life of its own (Far Eastern Economic Review 1974, p. 42). By the late 1970s, the Asian Bond Market acquired a more Asian character, with issues entirely managed by Asian institutions. Also, Asian managers and underwriters earned experiences and expertise to handle bond issues for Asia (Asian Wall Street Journal 1978, p. 11).

Midterm Credit of Asian Dollars

Another financial innovation in the Asian Dollar Market was the medium-term credit of Eurodollars to bridge the gap between the short-term and long-term Asian dollars. Like Asian bonds, international bankers in Singapore could easily emulate the financial instrument from the City of London, the Eurodollar certificate of deposits, which became popular in the Eurodollar market for its liquidity, flexibility, and security. By the end of 1972, the Monetary Authority of Singapore had already taken steps to encourage the formation of a market in negotiable certificates of deposits denominated in Asian dollars with the abolition of stamp duty (Pakshong 1972, p. 3). The international bankers in Singapore further expanded the array of financial instruments – the Singapore dollar negotiable certificate of

deposits in 1975, the US dollar negotiable certificate of deposits in 1977, and the Special Drawing Rights deposits in Asian Currency Units in 1979 (Woo 2016, p. 30). Japanese banks were quite active since the issuance of floating rates certificate of deposits in 1977. These diversified means of mobilizing funds attracted nonbank customers, including multinationals and other corporations (Asian Wall Street Journal 1978, p. 10).

By the end of the 1970s, although it still occupied a small share, the Asian Dollar Market in Singapore functioned as a regional market for the Eurocurrency market with an Asian character. It developed a structure which resembled the Eurocurrency market in the City of London and offered identical financial instruments to customers in Asia. The emerging economies in the Pacific Basin, the ASEAN countries, Hong Kong, South Korea, Taiwan, and Japan and their multinational and national corporations, tapped the local source of capital for their economic development. In the tenth anniversary report on the Asian Dollar Market in November 1978, the Asian Wall Street Journal forecasted its stable growth in the next decade in parallel with the economic development of Asia and placed Singapore as an international financial center along with New York, Zurich, and other offshore funding centers such as Hong Kong, Bahrain, Panama, and the Bahamas, which demonstrated the complementarity with the Asian Dollar Market (Asian Wall Street Journal 1978, p. 10–12).

Impact of the Asian Dollar Market

Singapore

The Asian Dollar Market certainly benefited Singapore in many ways. First of all, the city-state without natural resources garnered the standing of an international financial center for Asia less than two decades from its independence from Malaysia. One might find out from the full of foreign banks in the banking district of Singapore and its thriving economy. Wellons points out the following economic benefits: the surplus from financial services covered the country's consistent trade deficits; the offshore market offered a source of foreign credit for financing trade and industries of Singapore residents within the limit imposed by the government. The Development Bank of Singapore played a significant role in mobilizing Asian dollars in projects for Singapore as well. Of course, there were harmful effects of excessive liquidity and inflation in 1972 and 1973 owing to speculative inflows from the Asian Dollar Market to the domestic property and stock markets. The Monetary Authority of Singapore responded by imposing a higher reserve requirement and then the special deposits on net foreign interbank liabilities (Wellons 1977, pp. 363–369). Politically, the Asian Dollar Market was an invaluable success: it substantiated the globalist approach and future envisioned by the political leadership and helped Singapore to keep its neutrality and political stability in Southeast Asia, which further attracted

international banks. The trading port of Southeast Asia in the nineteenth century successfully transformed itself into an entrepôt for capital in the next century.

The Asia-Pacific Region

As noted above, in the early days, the Asian Dollar Market was contested as observers challenged its "Asian" character; the funds flowed from Asia to London. When the capital flow reversed, Singapore increasingly served to meet the regional demand for funds. The offshore market for Asian dollars lured investors and clients in the Pacific Basin area. In early 1974, the Monetary Authority of Singapore reported that 76% of Asian Currency Unit funds were taken by Asia (Wellons 1977, p. 372). By the mid-1970s, loans from nonbanks from Asia increased. These entities unknown to Europe or America had suffered higher interest rates in the Eurocurrency market. However, as Bhattacharya explains, the contribution of the regional Eurodollar market to the economic development of developing countries was controversial. The United Nations Conference on Trade and Development, an assembly of developing countries, had been complaining about the selective lending operations in the Eurocurrency market – the concentration of credits to five countries (Brazil, Mexico, Peru, South Korea, and the Philippines) (Bhattacharya 1977, p. 71). And the Asian Dollar Market was not free from the contestation.

Another important impact of Singapore as an international financial center was its competition and complementarity with the other countries. Its preeminence galvanized regulatory competition such as favorable tax treatment and lifting of restrictive regulations by other authorities. For example, Singapore reduced the tax on offshore loans in expectation of Hong Kong's reduction of income tax (Wellons 1977, p. 371). Later, Hong Kong changed its tax structure in the early 1980s to respond to the competition with Singapore (Schenk 2017, p. 15). The Philippines, which heavily relied its borrowing on Singapore, saw the lucrative business opportunities in the Asian Dollar Market and sought to make Manila a competing financial center (Wellons 1977, pp. 372–373). By the end of 1986, Japan joined the trend of offering an offshore market by opening the Japan Offshore Market in Tokyo. On the other hand, Wellons observed the complementary relationship between Singapore and other offshore markets around the globe particularly with Hong Kong via operations of international banks (Wellons 1977, p. 371).

Conclusions

This chapter concludes with the importance of the Asian Dollar Market in the history of international finance of the twentieth century. First, it exhibits a case study of the making of an international financial center by way of deliberate efforts by the local government. The government of Singapore was explicit in its

plan and was open to foreign experts. Second, one may point out the legacy of the British Empire in comparative research on offshore markets around the globe in the late twentieth century. These archipelagos in the non-Western region share features of the link with the City of London, the legal system, and the language inherited from Britain. Their success in the making of offshore banking centers and complementary relationship with Singapore contributed to the re-emergence of global finance. Third, the Asian Dollar Market shows the way in which the newly independent state of Singapore appropriated the financial instruments of Eurodollars from the West. From a vantage point of the emerging research on the global history, Singapore suggests the agency of the Global South and opens opportunities to integrate the history of international finance and international financial center with the Cold War and postcolonialism. In this way, the impact of offshore markets in non-Western world or developing countries upon the international monetary and financial system particularly the making of unfettered global finance in the late twentieth century and the rise of neoliberalism can be studied in the future.

Cross-References

- ▶ International Currencies in the Lens of History
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Money Markets: Eurocurrencies
- ▶ The Sterling Area 1945–1972

References

Asian Wall Street Journal (1978) Tenth anniversary report: the Asian Dollar Market. Asian Wall Street J 8–12

Attwood TM (1971, November) Singapore: the first two years of the Asian Dollar Market. Euromoney, p 52

Attwood TM (1972, September) Foreign banks and international banking. Euromoney, pp 5–7

Battilossi S, Cassis Y (2005) European banks and the American challenge: competition and cooperation in international banking under Bretton Woods. Oxford University Press, Oxford

Bhattacharya AK (1977) The Asian Dollar Market: international offshore financing. Praeger, New York/London

Bryant RC (1989) The evolution of Singapore as a financial centre. In: Sandhu KS, Wheatley P (eds) Management of success: the moulding of modern Singapore. Westview, Boulder, pp 337–372

Burn G (1999) The State, the City and the Euromarkets. Rev Int Polit Econ 6(2):225–261

Burn G (2006) The re-emergence of global finance. Palgrave Macmillan, Basingstoke

Burner CM (2016) Re-imagining offshore finance: market-dominant small jurisdictions in a globalizing financial world. Oxford University Press, Oxford

Cassis Y, Colier J (trans) (2006) Capitals of capital: a history of international financial centres, 1780–2005. Cambridge University Press, Cambridge

Effros RC (1985) Emerging financial centers: legal and institutional framework. International Monetary Fund, Washington, DC

Euromoney (1969, August) Far East: the Un-Asian Dollar. Euromoney, pp 29–30
Far Eastern Economic Review (1974) Asiandollars: a sign of independence. Far East Econ Rev 86(50):42–43
Ferguson N (2009) Siegmund Warburg, the City of London and the financial roots of European integration. Bus Hist 51(3):364–382
Hodjera Z (1978) The Asian Currency Market: Singapore as a regional financial center. IMF Staff Papers 25(2):221–253
Jang WN (1972, September) The growth of the Asian Dollar Market. Euromoney, pp 15–16
Josey A (1968) Lee Kuan Yew in London. D Moore, Singapore
Kim SW (2018) The Euromarket and the making of the transnational network of finance, 1959–1979. PhD dissertation, University of Cambridge, Cambridge
Kim SW (2019, Forthcoming) "Has the Euro-Dollar a future?" The production of knowledge, contestation and authority in the Eurodollar Market, 1959–1964. In: Hoppit J et al (eds) Money and markets: essays in honour of Martin Daunton. Boydell and Brewer, Suffolk
Lee S (1979) Recent development in Asian Currency Market and Asian Bond Market. Occasional paper/technical report series Institute of Economics and Business Studies, Nanyang University No 32
Lee LK (2000) From third world to first – the Singapore story: 1965–2000. Marshall Cavendish, Singapore
Lee LK (2015) The papers of Lee Kuan Yew: speeches, interviews and dialogues, vol 5, 1969–1971. The National Archives of Singapore, Singapore
Lee SY (1971) The Asian Dollar Market in Singapore. Malayan Econ Rev 16:46–56
Lee SY (1986) Developing Asian financial centres. In: Tan, Augustine HH et al (eds) Pacific growth and financial interdependence. London: Allen and Unwin, pp 205–36, 378–81
Lessard DR (1994) Singapore as an international financial centre. In: Roberts R (ed) Offshore financial centres. Edward Elgar, Aldershot, pp 200–235
Loen TK (1971, June) Singapore: financial entrepôt for Asia. Euromoney, p 80
Ogle V (2017) Archipelago capitalism: tax havens, offshore money, and the state, 1950s–1970s. AM Hist Rev 122(5):1431–1458
Pakshong MW (1972, September) The future for international banking. Euromoney, pp 2–3
Pakshong MW (1975, June) The Asian Dollar Market now. Euromoney, pp 101–102
Palan R (2010) International financial centers: the British-Empire, city-states and commercially oriented politics. Theor Inq Law 11(1):149–176
Palan R et al (2013) Tax havens: how globalization really works. Cornell University Press, Ithaca
Park YS (1982) The economics of offshore financial centers. Columbia J World Bus 17(4):31–35
Schenk CR (1998) The origins of the Eurodollar market in London, 1955–1963. Exp Econ Hist 35:221–238
Schenk CR (2010) The decline of sterling: managing the retreat of an international currency, 1945–1992. Cambridge University Press, Cambridge
Schenk CR (2017) Regulatory competition and complementarity in an offshore financial centre: the Asia Dollar Market in Singapore and Hong Kong 1968–86. UPIER Working Paper 17(2)
Shaxon N (2011) Treasure islands: tax havens and the men who stole the world. Bodley Head, London
Singapore International Merchant Bankers Ltd (1973) The Asian Dollar. Singapore International Merchant Bankers, Singapore
Tan CH (1982) Singapore as an international financial centre. In: Grub PD et al (eds) East Asia dimensions of international business. Prentice Hall, Englewood Cliffs, pp 29–44
Wellons PA (1977) Borrowing by developing countries on the Euro-currency market. Development Centre of the Organisation for Economic Co-operation and Development, Paris
Woo JJ (2016) Singapore as an international financial centre: history, policy and politics. Palgrave Macmillan, London

International Currencies in the Lens of History

13

Barry Eichengreen

Contents

Abstract

What makes for an international currency? What explains the predominant use of one or more national units in cross-border transactions? What explains changes in their popularity in absolute terms and relative to rivals? In the second half of the twentieth century, when the US dollar was the world's international currency par excellence, these questions appeared to have obvious answers. The United States was far and away the largest economy in the world, and it engaged in the largest volume of international transactions. Doing international business in dollars was logical and attractive insofar as the dollar was stable, and the United States had the largest and most liquid financial markets in the world. Finally, the United States had the capacity to project military and diplomatic power. As a country with a strong military, it was less vulnerable to attack from abroad of a sort that can destabilize its finances and economy and undermine confidence in its

B. Eichengreen (✉)
George C. Pardee and Helen N. Pardee Professor of Economics and Political Science, University of California, Berkeley, Berkeley, CA, USA
e-mail: eichengr@econ.berkeley.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_27

currency. Other countries held its currency as reserves as a way of signaling their allegiance – equivalently – of offering hostages. But does this same answer to the question of what makes for an international currency – size, stability, liquidity, and ability to project military power – also explain the rise and fall of international currencies earlier in world history? This chapter reviews almost two millennia of international monetary history and concludes that the answer to this further question is, to a surprising extent, yes.

Keywords
International currency · Reserve currency · Gold standard · Dollar

Introduction

What makes for an international currency? What explains the predominant use of one or more national units by individuals, firms, and governments undertaking cross-border transactions? What explains their rise and fall – that is to say, changes in their popularity in absolute terms and relative to rivals? And how many are there likely to be? Is international currency status a natural monopoly, in which a single leading national unit will always dominate international transactions, or can multiple currencies simultaneously play consequential roles in the international domain?

In the second half of the twentieth century, when the US dollar was the world's international currency par excellence, these questions appeared to have obvious answers. The United States was far and away the largest economy in the world, and it engaged in the largest volume of international transactions. It was logical and convenient for American banks, firms, and individuals, when engaged in cross-border transactions, to expect payment to be made in dollars. And what made sense for them made sense equally for foreign banks, firms, and individuals seeking to attract their business.

Doing international business in dollars was logical and attractive insofar as the dollar was stable. Under the Bretton Woods System of the 1950s and 1960s, the dollar was pegged to gold at a fixed price of $35 an ounce, and other currencies were effectively pegged to the dollar. Subsequently the dollar fluctuated more widely on foreign exchange markets, but without obvious trend. Aside from a brief period in the late 1970s, the United States did not experience the kind of chronic high inflation that tends to undermine confidence in a currency. Nominal stability in the United States was not absolute, but it was impressive relative to that now evident in countries whose currencies had been widely used in international transactions in earlier periods, Britain and France, for example.

Utilizing dollars in international transactions was also attractive because the United States had the largest and most liquid financial markets in the world. Dollars could be bought and sold at low cost, subject to minimal spreads. US banks with extensive foreign operations could make payments and extend loans to counterparties in virtually every corner of the globe. The liquidity of the market meant that dollars could be bought and sold in substantial quantities without moving

prices against the investor initiating the transaction. And the fact that US treasury bonds were traded in the single largest and most liquid financial market made dollar securities the logical form for central banks to hold their foreign reserves and financial and nonfinancial firms to hold their working balances.

A fourth and final prerequisite for international currency status is the capacity to project military and diplomatic power. A country with a strong military will be less vulnerable to attack from abroad of a sort that can destabilize its finances and economy and undermine confidence in its currency. Other countries will want to hold its currency as reserves as a way of signaling their allegiance – equivalently – of offering hostages.

These answers to the question of what makes for an international currency – size, stability, liquidity, and the ability to project military and diplomatic power – seemed obvious in the second half of the twentieth century, when the dollar dominated international transactions and only the United States possessed these attributes in abundance, and in some cases at all. Moreover, that the dollar far and away dominated in international transactions, in virtually all parts of the world, strongly pointed to the conclusion that international currency status is a natural monopoly and that only one national unit will play a consequential international role at a point in time.

Today in the twenty-first century, the answers to these questions are less obvious. America no longer accounts for as large a share of global GDP as in the dollar's heyday after World War II. The United States has been overtaken by China as the world's largest exporter. China has negotiated currency swap arrangements with foreign central banks and designated official renminbi clearing banks for financial centers around the world. It is prepared to challenge the United States in the geopolitical sphere, through foreign investments by its state-owned enterprises in Africa and by the Asian Infrastructure Investment Bank in East, South, and Central Asia (foreign investments, it should be noted, that are denominated in Chinese renminbi) and by building islands, complete with runways, in the South China Sea. Last but not least, the global crisis of 2007–2008 centered on the United States understandably raised questions about the stability and liquidity of US financial markets.

But with what implications is unclear. For the moment, the dollar still remains the dominant currency in the international monetary and financial sphere. Perhaps this is an indication that the network ("it pays to do what everyone else is doing") effects supporting dollar dominance in the past are not just powerful but persistent. Perhaps it means that large shocks like the two world wars that caused the dollar to finally supplant the pound sterling as the leading international currency, half a century and more after the United States became the leading exporter and overtook Britain as the world's largest economy, will be required for the baton to again be passed. Alternatively, perhaps as we navigate the transition from a post-World War II era dominated by the United States to a more multipolar world in which economic leadership is provided simultaneously by several powers, we will similarly experience a transition to a more multipolar monetary and financial world where several national currencies play consequential international roles. We may just have to wait a bit longer to see it.

These are hypotheses and questions on which history can presumably shed light. The rise first of sterling and then of the dollar and the transition between them is the most immediate such history. But there is also a longer prehistory of international currencies, prior to the period of sterling and dollar dominance, on which informed observers might usefully draw.

A Byzantine Arrangement

The silver drachma coined in ancient Athens in the fifth century B.C. is sometimes referred to as the first international currency since it circulated beyond the borders of the Athenian Empire (Chown 1994; Dwyer and Lothian 2002). Its successors, the Roman gold aureus and silver denarius, circulated more widely, reflecting the greater geographical scope and greater military and administrative capacity of the Roman Empire. In practice the silver drachma continued to circulate alongside these Roman units for an extended period, an observation consistent with the "new view" of international currency status that multiple international currencies can coexist (Cohen 1998; Eichengreen et al. 2017).

These gold and silver coins were used by the Romans to pay their legions. Their soldiers of course had to transform those high-value gold and silver coins into smaller units practical for use in everyday transactions. It followed that the gold and silver coins passed into other hands and came to be used in a range of transactions in a range of places.

Acceptance of the aureus and especially the silver denarius declined from the first to fourth centuries AD with the mounting financial challenges and declining power of the Roman Empire. These problems were met in part by the imperial authorities through traditional methods, namely, currency debasement and inflation. The aureus was then supplanted by the Byzantine solidus introduced by Constantine the Great as part of a set of administrative, financial, and economic reforms starting in 306 AD, which were designed to combat inflation and strengthen the Eastern Roman Empire, what came to be known as Byzantium. The solidus was actually introduced by Diocletian in 301, but minted only on a small scale, and entered into widespread circulation 11 years later under Constantine. It was also Diocletian who had divided the increasingly ungovernable Roman Empire into western and eastern halves, the latter of which was inherited by Constantine.

The solidus, like the aureus and denarius before it, was used to pay imperial soldiers. Indeed the word "soldier" derives from solidus, referring to the solidi with which solders were paid. In the fifth century, a soldier received an allowance of four solidi a year instead of the food that his predecessors received when taxes had been paid in kind (Spufford 1988, p. 8). The solidus was utilized, necessarily, in large-value transactions; its gold content, at a purity or fineness of 23 carats (95.8%), is today worth about US$185. Solidi were the heaviest gold coins circulating anywhere in the world, despite the fact that they were only the size of a modern American dime. As with the aureus and denarius before it, soldiers paid in solidi exchanged them for smaller monetary units, food, and other goods. In practice this meant that

the coin was used mainly by merchants and aristocrats. Cipolla (1967, p. 26) refers to the solidus as an "aristocratic coin."

The solidus circulated through much Europe and Asia for the better part of a millennium. Hoards have been found in Central Europe, Russia, Georgia, Syria, and other Arabic countries. That the imperial seat was moved from Rome to Byzantium, later Constantinople, elevated the stature of the Eastern Empire and encouraged use of the solidus in the Near East. Cipolla (1967, p. 16) writes that, from the fifth to seventh centuries, the coin circulated in the rich towns of the Near East, in the markets of North Africa, in the ports of Italy, and around the monasteries and castles of France and Spain. He quotes Kosmas Indicopleustes, a Greek monk and traveler, to the effect that the solidus was accepted "from end to end of the Earth." Lopez (1951) summons archeological and numismatic evidence to conclude that the solidus was used in transactions everywhere from England to India.

With which hypothesis or hypotheses about international currency status is this historical experience consistent? The answer, it would appear, is all of them. First, international use of the solidus was encouraged by the relatively large economic size of the Byzantine Empire and its substantial volume of its international transactions. Lopez (1951, p. 224) argues that Byzantium had "one of the highest average standards of living in the early Middle Ages." It was the richest and most cultured part of the Roman Empire once this was split into western and eastern halves by Diocletian. It engaged in a significant amount of foreign trade, which Lopez goes on to describe as "fairly large and well balanced," at least prior to the eighth century, when that trade, especially that following land routes, went into steep decline (Brownworth 2009). Other authors like Herrin (2007) portray the attitude of the Byzantine aristocracy toward trade in less positive terms. The Byzantines put together a collection of regulations governing naval contracts, so-called Rhodian Sea Law, to guarantee local merchant compensation for damage or loss from shipowners (Herrin 2007, p. 150). It is logical that their currency and not that of the rump Roman Empire, now descending into the Dark Ages, should have been the successor to the aureus and the denarius.

Second, historians emphasize the stability of the unit as key to its attractiveness as an international currency. The gold content of the solidus remained the same from its introduction in the fourth century until well into the tenth, a striking long period of time. Lopez writes that "its record [in stability and intrinsic value] has never been equaled or even approached by any other currency." The coin was minted mainly in Constantinople, under close watch of the emperor. Subsidiary mints in other cities such as Rome, Milan, Syracuse, Alexandria, and Jerusalem were, by contrast, subject to problems of quality control.

Byzantium, and Constantinople in particular, was able to avoid currency debasement because it maintained a balanced budget, seemingly for more than 300 years. Starting with Diocletian's own reforms, the government built up an impressive administrative apparatus with which to raise revenues through taxes on land, persons, and trade. And even when the budget was not balanced, prudent emperors accumulated a reserve of coin in good times on which they could draw in bad times (in military or political emergencies).

These policies were rooted in public support. The Byzantine Empire was characterized by a powerful landed aristocracy that supported the emperor. A landed aristocracy that depended on fixed dues paid by serfs and tenants would not have been pleased by inflation. The emperor, on the other hand, was largely indifferent to the interests of those whose obligations would have been eroded by currency debasement.

At a symbolic level, the solidus was identified with the power of the emperor, and this was one factor encouraging the government to maintain its value. If the solidus was stable and strong, thinking went, then the empire would be regarded as stable and strong. As Herrin (2007) puts it, the coin had "propaganda value."

Finally, international use of the solidus was supported by Byzantium's strong military. The solidus was the currency of an empire that for centuries after Constantine succeeded in controlling large swathes of territory and repelling invaders. The Byzantines' ability to mobilize tax revenues had more than a little to do with this.

But the solidus, like the drachma, had rivals, again consistent with the so-called new view of international currency competition. From the late seventh century, cross-border transactions were also undertaken using the dinar, a solidus-like gold coin introduced by Abd al-Malek, the fifth caliph of the Syrian Umayyad dynasty, who sought to unify the Moslem lands. The Byzantines were forced to import most of the gold from which their coins were struck. Al-Malek had the advantage of an indigenous supply of gold, obtained from the Upper Nile, which was an advantage for overcoming entry barriers.

Al-Malek's motivation was to put to an end the monetary disorder that prevailed in the Moslem lands, where myriad different coins of different weights and fineness circulated, causing confusion and complicating economic development. In modern terminology, he sought to create a "uniform currency." In addition, this was a period of discord over the merits of Islam and Christianity between the Umayyads and Byzantines. The Umayyads were less than pleased by the circulation of coins bearing Christian religious symbols. Where the Byzantine cross appeared on the front of the solidus, al-Malek substituted a column placed on three steps topped by a sphere and the phrase "In the name of Gold, there is no deity but Gold; He is One; Muhammad is the messenger of God." Justinian II's response was to strike a new solidus with the head of Christ on the front and himself, robbed and holding a cross, on the back (Ali 1999). The result was rising tensions between the Byzantine and Umayyad empires, culminating in war.

When al-Malek defeated Byzantine forces at the Battle of Sebastopolis in 692, the dinar became the sole circulating medium in the Moslem world, in a clear example of the association of international currency status with military power. It circulated not just in Syrian lands but in North Africa and Spain when these came under Umayyad control. Al-Malik issued a decree requiring that all Byzantine coins circulating in Umayyad territory be handed over to the treasury, which would melt them down and restrike them. Anyone failing to comply was subject to the death penalty. This, evidently, is one way that an existing international currency can be rapidly replaced by another.

Initially, the dinar had only a slightly lower gold content than the solidus (20 karats or 4.0 g) and was therefore used mainly in large-value, long-distance transactions. Al-Malek's successors sought to generalize his success by introducing one-half and one-third dinar coins. The dinar's gold content remained unchanged for fully two centuries following its introduction, again underscoring the association between currency stability and international use.

The subsequent period provided proof by counterexample, as both Byzantium and the Umayyads encountered problems of imperial overstretch not unlike those of the Greeks and Romans before them. Fiscal strains led to money creation and inflation achieved through debasement (in the Byzantine case mainly in the period ending around 1080). Reductions in the weight of the solidus (also now known as the nomisma) started in the late tenth century and continued into the eleventh. These reductions in size and weight, which were the principal method of debasement initially, made relatively little difference for acceptance, since in larger transactions coins were weighed rather than counted (Lopez 1951). Reductions in fineness, in contrast, were harder to detect and more corrosive of confidence. These followed in the late eleventh and twelfth centuries. "Byzantium's prestige plummeted as international merchants abandoned the worthless coins" (Brownworth 2009, p. 221). For the Byzantine Empire, the arrival of the Fourth Crusade at the beginning of the thirteenth century (organized by the Republic of Venice, more on which below) was then a political and economic disaster. The empire was dismembered and forced to pay a costly tribute to the victorious Crusaders (Harris 2014).

The Italian Job

The principal coins to circulate internationally in the subsequent period were the Genoese genoin and Florentine florin introduced in the mid-thirteenth century. Genoa, which went first in 1252, was on "bad terms" (Lopez 1951, p. 213) with the emperors of Nicaea who presided over what remained of Byzantine Empire and continued to mint gold coins. It therefore sought a circulating medium of its own. Venice followed, minting the ducat in the image of the florin in the fourteenth century, prompted by the continued debasement of the solidus by Emperor Andronicus II. (Andronicus was in a battle to the death with the Ottomans, entailing costly expenditures on mercenaries. The Venetians for their part continued to conduct a good deal of trade and business with Byzantium, making for their reliance on the solidus and now their creation of an alternative.) These 24 carat gold coins were state of the art; they were as pure as the technology of the time allowed, encouraging their acceptance. They circulated side-by-side and alongside token coinage and silver coins used for smaller-value, mainly local, transactions, again consistent with the so-called new view.

The three Italian city-states had important mercantile connections; Genoa and Venice were entrepôt centers between the Levant and Western Europe. Florence had links with the Champagne fairs and traded with the east using ships leased from the Genoese and Venetians (Najemy 2006). The Venetian ducat in particular acquired

and maintained its international role on this basis. Genoa and Venice had significant military prowess and were able to maintain enclaves abroad useful for their mercantile activities. They possessed a degree of natural protection by virtue of their geography, which made for security and political stability. They had relatively advanced fiscal administrations capable of raising the revenues required to finance the military operations needed to defend the state. They were technologically advanced by the standards of the day: Venetian merchant ships could carry as much as 700 tons of cargo already in the twelfth century (Madden 2012), helping to explain the Venetians' prominence not only in commerce but also the Crusades.

But as city-states they lacked the scale one might expect of the issuer of a leading international currency. For a time they controlled far-flung empires, but even then their territorial scope was limited. One can argue that what mattered more were trade connections, as opposed to productive capacity, in a period when there was an absence of other unitary states with extensive industrial, agricultural, or even pastoral potential. And like other international currency issuers before them, their units benefited from having a stable value. The florin was particularly stable, with no change in its design or metallic content for nearly 300 years from the mid-thirteenth through the mid-sixteenth centuries. It was a high-value coin containing 3.5 g of gold of relatively (99% plus) high purity, meaning that it was used mainly for large commercial and financial transactions.

The "financial" part of this last observation is key. What mattered most, surely, for the three Italian currencies' international roles was their development of large and liquid financial sectors, by the standards of the time. Venice and Genoa developed elaborate contracts, supported by a body of contract law, to facilitate collective finance of expensive commercial voyages (Epstein 1996). Venice invented double-entry bookkeeping. Florence and Venice were the inventors of deposit banking, which the Venetians claim grew out of the activities of their money changers at Rialto. The state required bankers to obtain a license, and as a condition for a license, the banker had to deposit with the state a sum of money that would be used to pay off depositors if the bank failed (see Madden 2012, p. 207 on the Venetian case, although the point is more general). These, then, were the early modern equivalent of capital requirements. They were an indication that for a financial system to acquire the depth and liquidity needed to support an international currency, it required more than the spontaneous stimulus of commerce. It depended in addition on the existence of a strong state capable of adequate regulation.

In time, these bankers became major providers and brokers of bills of exchange and letters of credit that were in practice denominated in and convertible into gold ducats, florin, and genoin. This class of credit instruments was first issued and changed hands among merchants and then others at the Champagne fairs, subsequently in settlements between Champagne and the banking houses of Tuscany and Venice, and ultimately more widely. They could be reliably converted into gold at a fixed rate of exchange. When large imbalances in bills of exchange built up in Florence, Genoa, or Venice, they were settled through transfers of bullion and coin. In this way bills of exchange became interchangeable, from the point of view of merchants and bankers, with bullion and coin (Spufford 1988, pp. 254–255).

Thus, bills on Venice, Genoa, and Florence substituted for and supplemented gold coins issued by these city-states. Effectively, Venice, Genoa, and Florence met the needs of international trade and finance with more than just coin, just as the United States today meets those needs with more than physical dollar bills. The Italian city-states pioneered the large-scale provision of financial services, first to local merchants, then to the fairs where large amounts of mercantile activity took place, and finally more broadly. Because Florentine banks had branches and did business throughout Europe (the Peruzzi company's bank had branches everywhere from Rhodes to Paris), the florin in particular became the dominant trade coin and unit of denomination for large-scale transactions across the western part of the continent. At the beginning of the fifteenth century, Florence then experienced some monetary and financial troubles, and the florin was debased. Before long the Venetian ducat had overtaken the florin as the leading international currency (Cipolla 1967, p. 21).

It is interesting to contrast the widespread use of the three Italian currencies in international trade and finance with much more limited use of the Spanish dollar, especially toward the end of the period. Spain accumulated massive troves of gold, silver, and copper when it acquired its colonial possessions in the Western Hemisphere starting in the fifteenth century. Mints in Seville and Burgos began minting high-quality silver and copper coins as early as 1505. Spanish coins were then minted in the Americas, starting in Mexico City and later in Santo Domingo, Lima Panama City, Potosi, Cartagena, and Bogota. These coins circulated throughout the Western Hemisphere; the Spanish dollar and its sub-units ("pieces of eight") were the dominant form of coin in the British North American colonies for two centuries prior to the Revolutionary War, since the British Parliament prohibited minting by the colonies themselves, accounting for their importation and reliance on Spanish dollars. The coins in question were supplemented by paper money issued by colonial governments (Rothbard (2002) provides additional detail). These New World coins were also exported to Spain, where they were used as collateral and in settlement for the crown's debts and were circulated in Florence, Genoa, the Low Countries, and the Baltics. They were accepted in payment for imports from China, India, and other parts of Asia. As Stein and Stein (2000, p. 201) put it, "Europe's internal exchanges were multiplying [in the period after 1720], matched by trade expansion to South India and on to the South China coast. Both were based on the reliable, mutually acceptable monetary foundation that only Spanish American silver *pesos fuertes* provided." (*Pesos fuertes* were pesos imported into Asia via the Spanish East Indies (the Philippine Islands in particular).)

Spanish money does not appear to have been used as widely as Florentine, Venetian, or Genoese coinage, however, or for as extended a period. One explanation for this may be questions about stability and uniformity. From the seventeenth century, there were complaints about the weight and fineness of newly minted Spanish coins. An investigation organized under Philip IV in the mid-seventeenth century documented several decades of fraud and abuse at the Potosi mint. In the early eighteenth century, there were then widespread complaints about the silver content of coins produced by the Mexico City mint, where the business was farmed out to a group of wealthy merchants, who in turn bankrolled local mine

owners. Although the substandard coins were recalled and, in the Mexico City case, the state assumed direct control of the mint, still the reputational damage was done.

In addition, the chronic fiscal difficulties of the crown, met in part by forced levies on taxpayers and intermittent debt restructuring (forced conversion of the monarch's liabilities), slowed the development of Spanish financial markets (Drelichman and Voth 2014). Finally, it is argued that the very foundations of the dollar's international currency status, namely, the silver "invasion," created Dutch disease problems for the economy, slowing the growth of merchandise exports. These Dutch disease problems are discussed by Kindleberger (1996), who argues that Spanish industry and commerce did not revive until the eighteenth century. A recent empirical treatment is Drelichman (2003), who documents a strong and persistent increase in the relative price of non-traded goods, implying a decline in the production of exportables, following the silver discoveries.

Evidently, three of the key foundations that might have led to wider international use of the Spanish dollar – stability, extensive mercantile links, and financial development – were missing or at least inadequately provided. Thus, there is evidence of Spanish silver dollars being used in China from the sixteenth century, mostly in coastal provinces like Fukien and Kwangtung disproportionately engaged in cross-border transactions (Yang 1952, p. 48). But those coins circulated by weight rather than face value, indicating problems of standardization, reflecting the aforementioned political and also technical problems at the mint leading to lack of uniformity.

Dutch Treat

Widespread international use of the three Italian currencies rested on a high level of mercantile activity – in modern terms, on the complementarity between trade and finance. It followed as commercial leadership shifted from the Mediterranean to the Low Countries, Italian monies were replaced by the Dutch guilder (also referred to as the florin, reflecting the positive reputation and circulation there of Florentine coins) as the leading international currency. Starting in the early seventeenth century, the Dutch Republic became not just the leading commercial power but also the leading source of credit and international finance for trade-related activity. Once it acquired this role, the guilder remained the leading currency used in cross-border transactions for the balance of the seventeenth and eighteenth centuries, reflecting the substantial value of the Dutch trade and financial transactions and the Dutch Republic's retention of this commercial predominance.

A further factor, once again, was the currency's stable value: there were no debasements of the guilder for more than 150 years after 1630. The guilder banco, deposit entries on the books of the Bank of Amsterdam (more on which below), was the prevailing unit of account. The physical guilder coin was the medium of exchange, at least initially. The guilder banco was traded at a slight and strikingly stable premium, known as the agio, relative to the physical guilder.

Here too financial innovation and development were key. The Bank of Amsterdam, established in 1609 by the governing council of the city, was central to the operation of this financial infrastructure. The Bank of Amsterdam is sometimes regarded as a proto-central bank (Quinn and Roberds 2005). It provided clearing and settlement services; converted specie into bank deposits at stable, standardized rates of exchange (after first deducting a small management fee, the agio); and held both gold and silver coin and bullion as reserves. It accepted and converted foreign as well as domestic coin, thereby supporting the growth of Amsterdam's international financial connections (Dehing and t'Hart 1997, pp. 46–47). It paid large bills drawn on Amsterdam in bank money, that is, through the transfer of bank deposits. It came to dominate the market, both in Amsterdam where city council regulation required all large bills of exchange to be settled through transfers of Bank of Amsterdam balances and elsewhere on the basis of reputation.

In the prior period, a wide variety of different coins, foreign as well as domestic, clipped and worn as well as full bodied, had circulated in the Dutch Republic. This created a reluctance on the part of foreigners to accept payment in coin, or for that matter to settle transactions in Amsterdam, since they were uncertain in which coin exactly they would be paid. With the substitution of bank money for this heterogeneous circulation, this confidence problem was solved. From the establishment of the Bank of Amsterdam and especially after 1700, bills on Amsterdam were accepted by merchants and bankers throughout the Baltic region and in Russia. In the same way, it has been argued that the establishment of the Federal Reserve System in 1913 and its subsequent support for the development of a market in trade credits was a key condition for the subsequent emergence of the US dollar as an international currency (Eichengreen 2011); the establishment of the Bank of Amsterdam was critical for wide international acceptance of the guilder.

This emphasis on finance is not to deny the importance of commerce. The Dutch Republic was the commercial superpower of the time. Technological advances in shipbuilding – design of the "fluyt" as a dedicated cargo vessel and new industrial methods for its construction – supported the explosive growth of the Dutch merchant marine, which accounted for fully half of all European shipping tonnage by the middle of the seventeenth century. The lion's share of the merchandise they carried passed through Dutch ports, encouraging use of the guilder. Still, it can be argued that the Dutch case, coming on the heels of the Italian city-states, illustrates the growing importance of financial as opposed to commercial prerequisites for international currency status, just as today US financial development dominates China's export prowess as a factor in the rivalry between the dollar and the renminbi.

In addition, the stability of the guilder, like the stability of the Italian currencies before it, raises interesting questions of political economy with relevance to modern experience. In Byzantium, as noted, the stability of the unit and avoidance of debasement were rooted in the support that the emperor derived from large landowners, who as creditors dependent on rents and dues fixed in nominal terms had a natural aversion to inflation. Support for the governing institutions of the Dutch Republic derived not so much from landed interests as from bankers and merchants. For those bankers and merchants, what was important was not so much avoiding

modest inflation as avoiding serious volatility, which might be a source of high uncertainty that disrupted trade and finance. In fact, the guilder was allowed to depreciate modestly, notably in the period before 1850. What were successfully avoided were major outbreaks of volatility that would be bad for banking and trade. There is an obvious parallel with the dollar in the second half of the twentieth century: the currency could decline at least modestly on the foreign exchange market without eroding its international currency status, but only so long as serious spikes in volatility were avoided.

Conductor of the International Orchestra

Sterling's rise to prominence as international currency primus inter pares can similarly be dated from the establishment of the Bank of England in 1694. In contrast to the Federal Reserve, in whose case it has been argued that the desire to internationalize the dollar was one of the motivations for founding the central bank (Broz 1997), enhancing sterling's international role was not one of the immediate objectives of the founders of the Bank of England. Rather, the bank was established in the wake of England's defeat by France in a series of naval encounters ending in 1690. William III had exhausted his credit in the unsuccessful campaign, leaving him few resources with which to rebuild the navy. The solution, following a plan devised by William Patterson, a Scottish merchant banker, was a bank to organize a loan. The Bank of England would manage balances generated through investor subscriptions, at some cost to itself, in return for specified monopoly privileges, notably the exclusive right, as a joint-stock company, to issue bank notes.

This was a hybrid public/private institution, like the Bank of Amsterdam before it and like the Swedish Riksbank established in 1688. It was also an institution, like the Bank of Amsterdam, whose responsibilities evolved over time. The Bank of England's duties as debt manager expanded over the course of the eighteenth century, mirroring the expansion of the English national debt. Its role in overseeing the operation of both the gold standard and the British financial system was acknowledged by the Bank Charter Act of 1844, which divided the bank into an Issue Department responsible for the convertibility of the currency and a Banking Department responsible for the stability and operation of the financial system. When liquidity grew scarce, other banks could now turn to the Bank of England to rediscount their bills. The bank intervened as a lender of last resort starting in the Overend, Gurney Crisis of 1866 (Flandreau and Ugolini 2011), reassuring international investors that the market in sterling-denominated claims would remain liquid. It could adjust its discount rate to attract gold from abroad (7%, according to the popular maxim, was enough to "draw gold from the moon"), stabilizing its reserve and buttressing confidence in the sterling parity and the gold standard generally. Operations like these were what led Keynes (1930) to dub the bank "conductor of the international orchestra."

Entire books have been written about the early history of the Bank of England. But this is enough to establish the essential point that the existence of a central bank

to provide liquidity to the market was a necessary but not sufficient condition for sterling's rise to international prominence.

"Necessary but not sufficient" because, historical experience suggests, other elements, plausibly four in number, had to fall into place to cement sterling's international role. First, the financial system had to develop further to enhance the stability and liquidity of the market. The middle decades of the nineteenth century, in particular, were a period of rapid growth and structural change in the banking system. A formerly fragmented financial sector underwent significant consolidation, creating a more stable and confidence-inspiring banking system. Overseas lending expanded, increasing the exposure of foreign investors to sterling and raising London's profile as an international financial center. The period after 1870 also saw growing circulation of treasury bills and bonds. These provided the banking system with a liquid asset in which to invest and the Bank of England with a convenient instrument with which to conduct open market operations, further enhancing the stability and attractions of the London market. Increasingly, the Bank of England bought and sold treasury securities on the market as a way of stabilizing interest rates at desired levels (Bloomfield 1959, p. 45). Others like Wood (1939) and Clapham (1944) note that the Bank of England engaged in sporadic open market purchases and sales in earlier periods, such as the 1830s, but such operations only became commonplace toward the end of the nineteenth century. Be this as it may, with interest rates and security prices relatively stable, treasury securities now became an increasingly attractive alternative to bank deposits for foreign governments and central banks seeking safe and liquid foreign assets.

Second, the economy had to develop so as to stimulate the volume of cross-border transactions. Britain's position as the first industrial nation was intimately related to sterling's position as the leading international currency. Britain was the world's leading exporter throughout the nineteenth century, where 80% of those exports were of manufactured goods at the century's end. Liverpool was a leading entrepôt center for the import and reexport of raw materials, starting with the cotton that was an essential input into the textile industry at the center of the first Industrial Revolution but extending eventually to a wide range of other commodities. London was the leading gold market, where a majority of the world's newly mined gold was priced and traded. Britain had the world's largest merchant fleet, a status it maintained well into the twentieth century. These developments in the real economy made for a large volume of overseas transactions (overseas rather than foreign because the Empire and Commonwealth were intimately involved), and a substantial fraction was naturally denominated in sterling.

Third, there had to be a consensus favoring currency stability. Some will trace this to the Glorious Revolution of 1688, which strengthened the political position of the large landowners who were the natural opponents of inflation and debasement (North and Weingast 1989). From these political changes limiting the arbitrary power of the crown and eliminating confiscatory government, it is argued, flowed the need to create the Bank of England, the deepening of British financial markets, and support for maintenance of the gold parity, established first by the Master of the

Mint under royal authority in 1717 and then by Parliamentary Act in 1816. Gold convertibility might still be suspended under duress, as during the Napoleonic Wars and the financial crises of 1847, 1857, and 1866. But on each occasion it was restored subsequently at the earlier statutory price. The maintenance of convertibility was supported not just by the landed interests but by bankers, who saw it as central to London's status as an international financial center, and by merchants and industrialists, who saw it as critical to Britain's success as an exporter. By the late nineteenth century, as Frank Fetter (1965) put it, sterling's fixed gold parity had attained almost constitutional status. "A suspension of the gold standard would have been almost inconceivable in Britain except under the most extraordinary circumstances," as David Glasner (1989, p. 42) writes. And foreigners knew it.

Fourth and finally, the country had to be militarily secure. Napoleon had contemplated invading the British Isles, but this maneuver was beyond even his very considerable reach. Not only did Britain enjoy the natural protection of the channel, but Britannia ruled the waves: its extensive merchant marine was complemented by an equally extensive naval fleet. Eventually, in the run-up to World War I, its naval preeminence, like its economic preeminence, would be challenged by other rising powers, notably Germany. This in turn raised questions in the minds of contemporaries about Britain's capacity to anchor the international gold standard (de Cecco 1974). But for much of the nineteenth century, sterling's heyday as an international currency, these questions were remote.

The preceding makes it seem all but inevitable that sterling should have been the leading international currency in the second half of the nineteenth century, when these five forces so fortuitously combined. It is important therefore to emphasize that it did not monopolize this function. Both the French franc and German mark were consequential rivals, especially toward the end of the period. Data gathered by Flandreau and Jobst (2009) show that while sterling was quoted and actively traded on every foreign exchange market worldwide circa 1900, the French franc was also traded on 80% of those markets and the German mark on 60%. While sterling accounted for half of global foreign exchange reserves at the turn of the century, the French franc accounted for fully 30% and the German mark 15%, according to the estimates of Lindert (1969). The franc was backed by the Bank of France, established in 1800, and by the creation of important new deposit banks starting in the 1860s, including big banks that engaged in a considerable volume of foreign lending. The international role of the mark was supported by the Reichsbank, founded in 1876, 5 years after the creation of Imperial Germany, and a rapidly developing financial sector, including some institutions like Deutsche Bank, founded in 1870 with the express purpose of financing foreign trade ("to promote trade relations between Germany, other European countries and overseas markets" as stated in its 1870 statute).

To be sure, sterling had a head start as an international currency. The relevant political changes (the French Revolution, German unification) came later than the Glorious Revolution. The Industrial Revolution took time to diffuse from England and Wales to the European continent. But by the end of the nineteenth century, it is clear that sterling had nothing resembling a natural monopoly.

Interwar Interregnum

The currency that is prominent by its absence from this list is, of course, the US dollar. The United States had already overtaken the United Kingdom as the single largest economy by the 1870s. It overtook the United Kingdom as the single largest exporter on the eve of World War I. And as the war made clear, it now had the most powerful military, backed by the largest industrial sector, of any country.

At first sight, it thus seems paradoxical that the dollar played essentially no role as a currency in which to invoice and settle export and import transactions, as a unit for denominating international bonds, and as a form in which central banks and governments held their foreign reserves. On closer look, however, the paradox dissolves. The United States lacked a central bank to act as lender and liquidity provider of last resort prior to the establishment of the Federal Reserve System in 1913. In the absence of an elastic currency, the US financial system was prone to periodic bouts of financial stringency and crisis, a pattern that did nothing to attract foreign business to New York or enhance the attractions of the dollar. US banks were essentially prohibited from branching abroad under the National Banking Act put in place during the American Civil War. That internecine conflict was hardly a shining example of political stability and solidarity of the sort needed to engender confidence on the part of foreign investors. And the country's commitment to the gold standard was a perennial question, at least prior to William Jennings Bryan's defeat in the 1896 presidential election and passage of the Gold Standard Act of 1900, given its system of universal male suffrage, which extended the vote to small farmers and other debtors, and a Populist Movement that associated gold with deflation and hardship.

As in the case of the Bank of England before it, the creation of a central bank was a necessary condition for altering these conditions. A variety of factors and considerations came together to prompt the establishment first of the National Monetary Commission and then passage of the Federal Reserve Act. The most prominent was surely the perceived need for an "elastic currency," as the concept was put in the act, provided by a central bank that could utilize its discount rate, in the manner of the Bank of England, to modify money and credit conditions as needed to prevent the seasonal spikes in interest rates that were a source of financial dislocation and distress.

But an important subsidiary motive was to internationalize the dollar: to create an institutional framework in which America's currency could play a larger international role. Under the provisions of the Federal Reserve Act, US banks were permitted to open foreign branches and originate foreign business. The reserve banks were encouraged to discount and purchase trade acceptances – essentially, promissory notes financing import and export transactions – denominated in dollars, both on their own account and on behalf of foreign central banks. Paul Warburg, the German-American financier who testified before the National Monetary Commission and subsequently became a founding member of the Federal Reserve Board, was intimately familiar with the advantages accruing to European exporters from the existence of markets in trade acceptances denominated in European currencies.

He actively sought the same advantages for his adopted country. While Warburg's stint on the Federal Reserve Board was short, his influence was enduring. The System was quick to enter the market in trade acceptances, on which, for the better part of two decades, it was the dominant player.

The impact of the establishment of the Fed on the international role of the dollar is hard to pinpoint because the new central bank's opening for business coincided with the outbreak of World War I. The European belligerents embargoed gold exports and placed financial flows under government control. Credit that had previously financed import and export transactions in third countries was now directed exclusively toward domestic needs. Previously stable European exchange rates began to fluctuate. In some cases they would fluctuate even more violently after the war, when administrative controls on capital flows were removed but normalcy in its other aspects was not yet restored.

From this monetary and financial turmoil, the United States stood apart, just as it stood apart from the war itself until 1917. Only the dollar exchange rate against gold did not move. Only US gold exports were officially free of embargo. US banks, having been liberated by the Federal Reserve Act, could step in and fill the vacuum created by the absence of European banks. Already during the war, they established branches in Latin America and Asia, followed after 1919 by branches in Europe itself. They were now well positioned to originate trade credits and underwrite foreign loans, all denominated in dollars.

It was not as if London and sterling were prepared to abdicate the throne. Maintaining London's position as an international financial center was an important factor in Winston Churchill's decision, as Chancellor of the Exchequer, to restore the prewar gold parity and $4.86 exchange rate against the dollar in 1925. Following earlier suspensions, resumption had always been at the earlier gold parity, as we saw above. To do otherwise now would diminish confidence in sterling and undermine the international position of the City (Moggridge 1971; Boyce 2004).

That decision was and remains controversial. It did not succeed in rejuvenating the British economy, if anything having the opposite effect. It did not even succeed in stabilizing the exchange rate for more than a brief period; Britain was forced to abandon the gold standard and once again allow the currency to depreciate in 1931. But it did allow sterling to regain its position as the leading international currency and to maintain it until after World War II, or so it is widely asserted (see, e.g., Chinn and Frankel 2007). This historical experience, so interpreted, thus lends support to the traditional interpretation of international currency status in network-effect and natural-monopoly terms.

Yet the fact that the sterling parity restored in 1925 was now referred to in dollar terms ("the Norman conquest of $4.86" after Montagu Norman, governor of the Bank of England), terms of reference that would not have been used before World War I, sits uneasily with this traditional interpretation. Those terms of reference are a clear indication that something had changed, that something being a more prominent dollar. By the mid-1920s, central banks already held as many reserves in dollars as sterling (Eichengreen and Flandreau 2009). Trade acceptances denominated in dollars were already as important as trade credits denominated in sterling

(Eichengreen and Flandreau 2012). More international bonds were denominated in dollars than sterling, reflecting the rise of New York as an international financial center and the impact of controls and moral suasion utilized in an effort to limit long-term foreign investment by Britons and strengthen the country's balance of payments (Eichengreen et al. 2013). All this is more easily reconciled with the "new view" questioning the power of network effects and increasing returns and suggesting that multiple international currencies can coexist.

Two explanations suggest themselves for why these facts were not better appreciated until recently. One is data limitations. Data on the currency composition of foreign exchange reserves in the 1920s and 1930s were only fragmentary until new evidence was extracted from the archives in the last decade. And data on the composition of reserves in the second half of the 1940s seemingly suggested – misleadingly, it will be argued below – that sterling remained the dominant reserve currency even in the aftermath of World War II.

The other explanation is the retreat of the dollar in the 1930s – its declining share in various international markets, as mentioned above. In part this reflected the retreat of international currencies more generally. Having suffered capital losses on their foreign balances due to currency devaluation, central banks liquidated their foreign deposits and bonds, shifting instead into gold. The collapse of world trade in the depression of the 1930s reduced the demand for trade credit whether sourced in New York, London, or other financial centers. Sovereign defaults demoralized the market and depressed the value of new bond issues marketed to international investors, independent of currency of denomination. The proliferation of capital controls repressed cross-border financial flows more generally.

But the decline in international use of the dollar was even more pronounced than the decline in the use of sterling on several of these metrics. The liquidation of dollar reserves was more complete. There was no dollar analog to the Sterling Area, a group of countries that continued to peg their currencies to sterling, with varying degrees of rigidity, and to hold the bulk of their foreign exchange reserves in London. The cohesion of the Sterling Area reflected ties of Commonwealth and Empire; for reasons of tradition and affiliation, London was the logical place to maintain reserves for countries like Australia and New Zealand. The Ottawa Agreements and Imperial Preference gave these and other countries preferential access to the British market, again making London the logical place to do international financial business. When the United States took its protectionist turn in the 1930s, it did not extend analogous preferences.

In addition, sovereign default rates were higher on dollar bonds than sterling bonds, discouraging new dollar issuance. The US banking and financial crisis was more severe, causing American banks to disproportionately curtail their security-market operations. Even Federal Reserve support for the market in trade acceptances became a liability when the central bank, preoccupied by other matters, withdrew from the market starting in 1931, revealing the extent to which its active acceptance purchase and discounting program had slowed the entry of other investors.

The 1930s, a period of high turbulence and economic and financial crisis for the United States, thus marked a pause in the dollar's rise to the status of leading

international and reserve currency. But there would be no analogous pause after World War II.

Dollar Dominance

The second half of the twentieth century was the period of dollar dominance. The United States dominated the post-World War II international monetary system, just as it dominated the wartime Bretton Woods Conference where the institutional contours of that system were forged. The United States was the world's sole monetary superpower. The Soviet Union and the Republic of China were present at Bretton Woods, but following the conclusion of World War II, Stalinist Russia withdrew from the international economy, taking the rest of the Soviet Bloc with it. Chiang Kai-shek's Republican government withdrew to the island of Taiwan, leaving the Chinese mainland in Communist hands.

Of the countries that remained active participants in the international system, only the United States possessed the geopolitical clout associated with widespread international use of its currency for trade, investment, and reserve-holding purposes. It had the largest GDP of any country, and for a brief period after World War II, it accounted for fully half of the free world's industrial production. It was far and away the leading exporter. Other countries, in particular European countries grappling with the challenges of postwar reconstruction, were desperate to get their hands on US-produced capital goods, resulting in chronic US trade surpluses. This made dollars a precious commodity. US financial markets survived the war intact, not something that could be said of many other countries. The United States was the leading source of foreign investment, mainly direct foreign investment in the immediate postwar years. It was the leading source of foreign aid, most famously extended through the Marshall Plan. The US investment and aid in question, along with eventually their own exports, enabled other countries to accumulate dollar balances. And the heavy weight of the United States in global current- and capital-account transactions, not to mention the security umbrella it provided its allies, gave other countries an incentive not just to hold dollars but to use them in their international transactions.

The mirror image of US economic and financial strength was weakness abroad. Germany and Japan had both demonstrated considerable industrial muscle during the war, and Germany if not Japan, had been a significant international monetary player in an earlier era, as argued in the previous sections of this chapter. But both countries were now under foreign occupation, and Germany was partitioned. They were reluctant to relax the exchange and capital controls whose removal was the sine qua non for widespread international use of a currency. Germany only restored full convertibility for transactions on current account at the end of 1958, while Japan delayed that step until 1964. Japan only finally removed the last of its capital controls in the 1980s. Germany moved more quickly to restore capital-account convertibility, but it was not afraid to reintroduce controls when free international capital mobility and domestic policy priorities proved to be at odds, as it was the case in the early

1970s. Free capital mobility was incompatible with the industrial policy operated by the Japanese Ministry of International Trade and Industry (MITI), which directed financial resources to priority uses. It was incompatible with the West German desire for a competitive exchange rate to fuel the postwar growth miracle (the *Wirtschaftswunder*), and the now deep-seated German aversion to inflation, insofar as the combination of a pegged exchange rate and capital mobility made an independent monetary policy and domestic inflation control effectively impossible. Thus, Tokyo and Bonn both resisted currency internationalization. They took regulatory and other measures to discourage wider international use of their currencies (Eichengreen et al. 2016).

This left only sterling to rival the dollar. Sterling had been used as widely as the dollar in international trade and payments in the 1920s, and more widely than the dollar in the 1930s, as we saw in the previous section. London remained an international financial center. Sterling balances were still widely held by central banks and governments after World War II. Taken at face value, as in Fig. 1, postwar sterling balances were even larger, and by a considerable margin, than dollar balances in foreign hands, Britain's allies having accumulated them in payment for wartime goods and services rendered.

But, in any meaningful economic sense, this image was an illusion. The British economy was one of the most slowly growing in Europe. The United Kingdom was no longer able to finance substantial foreign military commitments, a fact made clear by its 1947 withdrawal from Greece and its acquiescence to colonial independence movements. Its balance of payments was in chronic deficit, as underscored by the failed attempt to restore current-account convertibility in 1947, devaluation in 1949, and then ballooning external deficits in 1951–1952. This was not the kind of stability expected of a leading international currency – to the contrary.

Foreign central banks and governments thus had a clear incentive to liquidate their sterling balances – to convert them into merchandise or dollars – while they still had value. The loyalty and sense of comradery that had been felt so powerfully during the war, among members of the British Commonwealth in particular, was no longer so pervasive following its conclusion (Schenk 2010). As a result, the British government was forced to take measures to limit the conversion of sterling balances. Those balances were held in London: hence they were subject to UK regulation on their withdrawal (in the case of bank deposits) or sale (in the case of bonds). A mitigating factor was that the United Kingdom itself desired market access and, in negotiations with the United States, access to technology and financial assistance. The result was a negotiated settlement in which the sterling balances of different countries were treated differently. The balances of Sterling Area countries could be used for purchases of merchandise and other financial assets within the Sterling Area itself but not elsewhere. The so-called Transferable Account countries, mainly European countries in practice, were permitted to use their sterling reserves for payments between Transferable Accounts and Sterling Area accounts but not for payments to so-called American account countries (members of the Dollar Area).

The effect was to limit opportunities for converting sterling into dollars and using it to purchase merchandise in the Dollar Area. Sterling could be

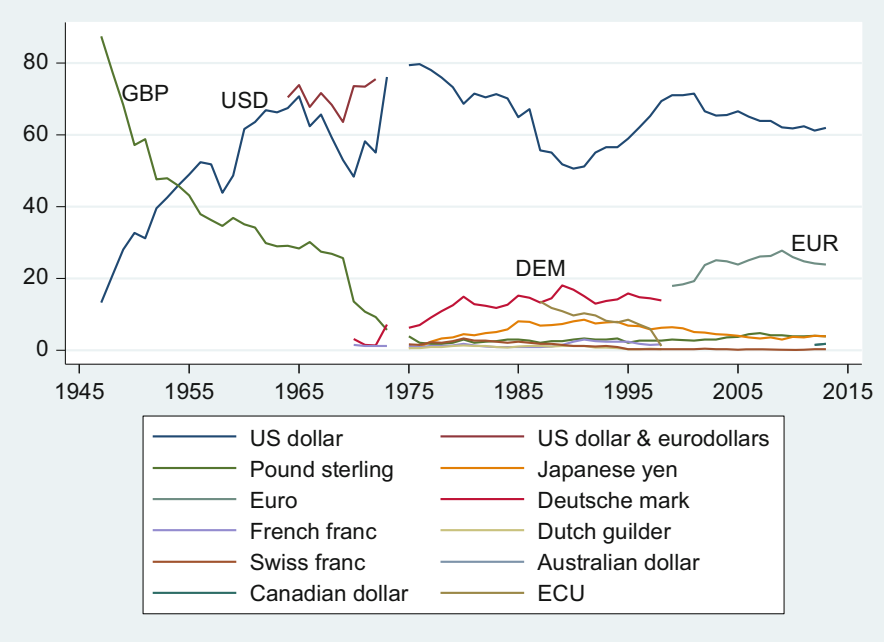


Fig. 1 Currency composition of globally disclosed foreign exchange reserves, 1947–2013, in percent of total. (Sources: Eichengreen, Mehl and Chitu (20,160). Notes: The currency shares are derived from US dollar-denominated amounts for the period 1947–1969 and 1999–2013 as well as from SDR-denominated amounts for the period 1970–1972. The currency shares for the period 1973–1999 are directly provided by the IMF in its annual reports (based on SDR valuation). Starting in 1979 the Fund added the SDR value of ECUs issued against the US dollar to the SDR value of US dollar reserves; after 1987 the ECU was treated as a separate unit. The currency shares reported here exclude unallocated foreign exchange reserves post-1994 (i.e., about 40% of total foreign exchange reserves at the end of the sample)

redistributed among Sterling Area countries, but residents could liquidate their sterling reserves only by using them in settlements with the United Kingdom itself. The British government fostered the practice by maintaining trade and capital controls.

The consequences are evident in Fig. 1, which shows not the sudden liquidation of sterling balances but an ongoing decline in its relative position (i.e., its share of global total foreign exchange reserves) stretching over several decades. In particular, there was a relatively rapid decline in the share of sterling in identified global foreign exchange reserves in the first post-World War II decade, a somewhat slower decline in the second postwar decade, and then accelerating decline again in the third following another sterling devaluation in 1967 and intensifying British balance-of-payment problems. Numerically, sterling was supplanted first by the dollar, for all the reasons described above, and then by the deutschemark, the yen, and a variety of subsidiary currencies, as these and other countries liberalized their financial markets and opened their capital accounts.

Also evident from Fig. 1 is that the dollar was never the sole international reserve currency at any point in the second half of the twentieth century. When currency holdings are valued at current exchange rates, the dollar's share peaked at the end of the 1970s at around 80% of the global total. There was then an undulating decline in that share toward a relatively stable 60%. Thus, the dollar's share of identified foreign exchange reserves at the beginning of the twenty-first century was roughly the same as sterling's share at the beginning of the twentieth. Like sterling a century before, the dollar now accounted for the largest single fraction of official foreign currency holdings, but not the entirety.

The other international reserve currency in this period was not, in fact, a currency: it was the IMF's Special Drawing Rights. SDRs are accounting units credited to IMF members, upon agreement by 85% of the institution's membership, which members are obliged to accept in official transactions with the IMF itself and one another. The idea of a multilateral source of international liquidity to supplement national currencies had been mooted by Keynes at Bretton Woods, but serious negotiations only got underway in the second half of the 1960s. Agreement was reached in 1969, and the first allocation of SDRs occurred on January 1, 1970. Following additional allocations in 1971 and 1972, SDRs accounted for approximately 10% of global non-gold reserve assets.

Additional SDR allocations followed periodically. But the overall trend in the SDR's share was downward, to the point where it accounted for little more than 2% of non-gold reserve assets in 2015. This trend reflected constraints on both the supply and demand sides. On the supply side, agreement of 85% of IMF membership, as required to amend the institution's Articles of Agreement, was a formidable hurdle. To achieve this it was necessary to agree on the distribution of additional drawing rights across countries – whether to allocate them in proportion to countries' quota shares in the institution, in which case the lion's share would accrue to high-income countries, or instead to distribute them to low-income countries, in which case a different, ad hoc formula would have to be devised. Insofar as the purpose of creating additional SDRs was to supplement and eventually supplant the dollar, the fact that the United States alone possessed more than 15% of voting power in the Fund constituted a further obstacle.

On the demand side, there was the fact that the SDR lacked a number of the essential attributes of the dollar and other national currencies used in international transactions. The SDR couldn't be used to settle private transactions, only official transactions among governments. There was limited demand by private investors for international bonds denominated in the basket of currencies comprising the SDR. There was no liquid secondary market in SDR-denominated instruments. A few token SDR bonds were issued in the 1970s, and the World Bank sold SDR-denominated bonds in the Chinese market in 2016, but no more. Banks and other private financial institutions might have issued SDR-linked or denominated securities – private financial institutions having shown themselves as good, sometimes too good, at concocting innovative financial instruments – but there was no demand. With the progressive liberalization of domestic and international financial markets, it was possible for investors to construct a bond portfolio with currency

weights of their choice, that dominated the IMF's unit of six national currencies with fixed weights.

The SDR was created in response to the idea that the Bretton Woods System was unstable. There was a contradiction at the heart of a system predicated on a stable dollar price of gold that also depended on dollars as the incremental source of international liquidity, a point made by Triffin (1960) and his followers. Absent rapid progress on substituting SDRs for dollars, it might become impossible to keep the dollar price of gold stable. The dollar might have to be devalued, undermining confidence in the principal source of international liquidity and throwing the global economy and international financial system into chaos.

In the event, the forecast of dollar devaluation was right, but the prediction of international financial chaos was wrong. The Bretton Woods System collapsed in 1971–1973, when first the dollar was devalued and then other currencies were floated. But there was no sharp reduction in the demand for international reserves, as some experts had predicted, and no sharp shift away from the dollar as a form in which to hold them. From the time of the Roman aureus through the post-World War II dollar, international currencies had been minted from or otherwise linked to precious metal. The architects of the Bretton Woods System carried on in this tradition when they obliged IMF member countries to declare par values for their currencies "in dollars of constant gold content" and singled out the dollar because it alone was convertible into gold at a fixed price by official foreign holders. Many observers therefore concluded that with the end of the dollar's fixed link to gold, the currency's international role would be significantly attenuated if not eliminated.

That this was not the case is again evident from Fig. 1. As had been true for two millennia, the acceptability of a national currency in international transactions still depended on the size, stability, and security of the issuer and the liquidity of its financial markets. But now stability was gauged by more than the stability of its value in terms of gold. It was gauged rather by the stability of its finances, its policies, and, ultimately, its economy.

Looking to the Future

While history provides no road map for the future, it does offer hints for how to think about it – in this case, for how to think about the future of international currencies. Many historical periods have seen a dominant unit, where a significant fraction of cross-border transactions were conducted in a particular national currency, network increasing returns serving as a powerful attractor. But any such dominance has regularly fallen short of absolute. Working in the other direction is the desire of central banks, governments, and other investors not to put all their eggs in one basket but rather to hold a diversified portfolio of foreign currencies and to conduct cross-border transactions using settlement mechanisms utilizing different national units. That said, the desired degree of portfolio diversification can in general be achieved by accumulating and utilizing a limited number of currencies. Nor is there reason to

expect the international currencies in question to be held in equal amounts; one currency like the guilder, then sterling, and most recently the dollar has always been disproportionately important in international markets, reflecting the continuing sway of those network effects.

The long sweep of history suggests that the balance is tipping, if slightly and gradually, away from network effects and toward portfolio diversification. With the development of modern financial markets and instruments, it becomes less advantageous to utilize the same national unit as one's trading partners when engaging in international transactions and easier to exchange one currency for another. With advances in information technology, it becomes easier to compare prices denominated in different currencies. With the development of national financial markets, it becomes more attractive to diversify portfolios – to hold and transact in a variety of different currencies. This historical trend, if it continues, suggests that the dollar will have more rivals, and perhaps more consequential rivals, in the future than the past.

The euro and the Chinese renminbi are most obvious such rivals, since only the Euro Area and China compare to the United States in economic size. But historical experience also suggests questions about whether these economies possess the other attributes required of the issuer of an international currency. In the case of the Euro Area, these center on whether the monetary union can develop the strong state typically associated with a sound and stable currency that is widely accepted and utilized internationally. This means a state apparatus with the capacity to efficiently regulate financial markets – banking union, in other words. It means a state with the fiscal capacity to execute its core functions in noninflationary ways – fiscal union or at least significant fiscal integration at the level of the union. It means a state capable of defending its integrity and dispatching doubts about its future. Historically this has meant a state with a strong military. In the European case, it probably means a union of states capable of collectively securing their common external borders and sufficiently committed to their monetary-union project to eliminate doubts about its permanence. This is not a project that will be completed overnight.

In China's case, in contrast, what is needed is not a stronger state but a more limited state whose authority is subject to checks and balances. Just as political change acknowledging the interests of landowners, merchants, and other investors was needed to secure the status of international currencies from the Byzantine solidus and the Venetian ducat to the Dutch guilder and the British pound, so too political change limiting the arbitrary exercise of power by the Standing Committee of the Communist Party will be needed to secure widespread international acceptance of the renminbi. Limits on the arbitrary exercise of political power and confidence in the arm's-length adjudication of contract disputes, including disputes involving foreign investors, will be needed for the development of deep and liquid financial markets on which there is active foreign participation. How far these political changes will have to go is uncertain. What is certain, once again, is that they will not be completed overnight.

But then, as any issuer of aureus would tell you, Rome was not built in a day.

Cross-References

- ▶ European Monetary Integration
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ The Role of Money in the Economies of Ancient Greece and Rome

References

Ali W (1999) The Arab contribution to Islamic art from the 7th–15th centuries. American University in Cairo Press and Royal Society of Fine Arts, Amman

Bloomfield A (1959) Monetary policy under the international gold standard, 1880–1914. Federal Reserve Bank of New York, New York

Boyce R (2004) Government-City of London relations under the gold standard, 1925–31. In: Michie R, Williamson P (eds) The British government and the City of London in the twentieth century. Cambridge University Press, Cambridge, pp 215–235

Brownworth L (2009) Lost to the west: the forgotten Byzantine Empire that rescued Western civilization. Crown Publishers, New York

Broz L (1997) The international origins of the Federal Reserve System. Cornell University Press, Ithaca

Chinn M, Frankel J (2007) Will the euro eventually surpass the dollar as leading international reserve currency? In: Clarida R (ed) G7 current account imbalances: sustainability and adjustment. University of Chicago Press, Chicago, pp 283–338

Chown J (1994) The history of money from AD 800. Routledge, London

Cipolla C (1967) Money, prices and civilization in the Mediterranean world: fifth to seventeenth century. Gordian Press, New York

Clapham H (1944) The Bank of England: a history, 2 volumes. Cambridge University Press, Cambridge

Cohen B (1998) The geography of money. Cornell University Press, Ithaca

De Cecco M (1974) Money and empire: the international gold standard. Blackwell, Oxford

Dehing P, Hart M'T (1997) Linking the fortunes: currency and banking, 1550–1800. In: Hart M'T, Jonker J, van Zanden JL (eds) A financial history of the Netherlands. Cambridge University Press, Cambridge, pp 37–63

Drelichman M (2003) The curse of Moctezuma: American silver and the Dutch disease. Unpublished manuscript. University of British Columbia (November)

Drelichman M, Voth H-J (2014) Lending to the borrower from hell: debt, taxes and default in the age of Philip II. Princeton University Press, Princeton

Dwyer G Jr, Lothian J (2002) International and common currencies in historical perspective. Unpublished manuscript. Federal Reserve Bank of Atlanta and Fordham University (May)

Eichengreen B (2011) Exorbitant privilege: the rise and fall of the dollar and the future of the international monetary system. Oxford University Press, New York

Eichengreen B, Flandreau M (2009) The rise and fall of the dollar (or when did the dollar replace Sterling as the leading reserve currency?). Eur Rev Econ Hist 13:377–411

Eichengreen B, Flandreau M (2012) The Federal Reserve, the Bank of England and the rise of the dollar as an international currency 1914–39. Open Econ Rev 23:57–87

Eichengreen B, Mehl A, Chițu L (2013) When did the dollar overtake sterling as the leading international currency? Evidence from the bond markets. J Dev Econ 111:225–245

Eichengreen B, Mehl A, Chitu L (2016) Stability or upheaval? The current composition of foreign exchange reserves in the long run. IMF Econ Rev 64:354–380

Eichengreen B, Mehl A, Chitu L (2017) International currencies past, present and future: two views from economic history. Princeton University Press, Princeton
Epstein S (1996) Genoa and the Genoese 958–1528. University of North Carolina Press, Chapel Hill
Fetter F (1965) The development of British monetary orthodoxy 1797–1873. Harvard University Press, Cambridge, MA
Flandreau M, Jobst C (2009) The empirics of international currencies: network externalities, history and persistence. Econ J 119:643–664
Flandreau M, Ugolini S (2011) Where it all began: lending of last resort and the Bank of England in the Overend-Gurney Panic of 1866. Norges Bank Working Paper no. 2011/03 (March)
Glasner D (1989) Free banking and monetary reform. Cambridge University Press, Cambridge
Harris J (2014) Byzantium and the crusades, 2nd edn. Bloomsbury Academic, London
Herrin J (2007) Byzantium: the surprising life of a medieval empire. Princeton University Press, Princeton
Keynes JM (1930) A treatise on money. Macmillan, London
Kindleberger C (1996) World economic primacy, 1500–1990. Oxford University Press, New York
Lindert P (1969) Key currencies and gold, 1900–1913. Princeton Studies in International Finance, vol 24. International Finance Section, Department of Economics, Princeton University, Princeton
Lopez R (1951) The dollar of the middle ages. J Econ Hist 11:209–234
Madden T (2012) Venice: a new history. Viking, New York
Moggridge D (1971) British controls on long-term capital movements, 1924–31. In: McCloskey D (ed) Essays on a mature economy: Britain since 1840. Princeton University Press, Princeton, pp 113–138
Najemy J (2006) A history of Florence 1200–1575. Blackwell, Oxford
North D, Weingast B (1989) Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. J Econ Hist 49:803–832
Quinn S, Roberds W (2005) The big problem of large bills: the Bank of Amsterdam and the origins of central banking. Federal Reserve Bank of Atlanta working paper no. 2005–16 (August)
Rothbard M (2002) A history of money and banking in the United States: the colonial era to World War II. Ludwig Von Mises Institute, Auburn
Schenk C (2010) The decline of Sterling: managing the retreat of an international currency 1945–1992. Cambridge University Press, Cambridge
Spufford P (1988) Money and its use in medieval Europe. Cambridge University Press, Cambridge
Stein S, Stein B (2000) Silver, trade and war: Spain and America in the making of early modern Europe. Johns Hopkins University Press, Baltimore
Triffin R (1960) Gold and the dollar crisis: the future of convertibility. Yale University Press, New Haven
Wood E (1939) English theories of central banking control, 1819–1858. Harvard University Press, Cambridge, MA
Yang L-S (1952) Money and credit in China: a short history. Harvard University Press, Cambridge, MA

Part IV

Money and Metals

American Precious Metals and Their Consequences for Early Modern Europe

14

Nuno Palma

Contents

Abstract

Over the early modern period and beyond, massive amounts of silver and gold were found and mined in the Americas. This chapter reviews the consequences for the European economies. Some second-order receiver countries such as England benefited in both the short and long run. First-order receivers such as Spain and Portugal also benefited in the short run, but their continued exposure to the arrival of massive quantities of precious metals eventually led to loss of competitiveness and an institutional resource curse.

Keywords

American precious metals · Early modern period · Dutch disease · Political institutions · Economic growth · Comparative development

N. Palma (✉)
Department of Economics, School of Social Sciences, University of Manchester, Manchester, UK

Instituto de Ciências Sociais, Universidade de Lisboa, Lisboa, Portugal

CEPR, London, UK

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_58

Introduction

Over the early modern period, massive quantities of silver and gold were produced in the Americas. Most gold and much silver arrived to Europe, even if some of the silver was sent directly to Asia via the Pacific (Manila) route and some was retained in the Americas. Nevertheless, Europeans were the main direct beneficiaries of this windfall, which dwarfed the initial quantities available. It is hence natural to ask what where the effects that these precious metals had for the European early modern economy. In particular, were they a curse or a blessing?

Contemporaries were already puzzled by these questions. For example, scholars of the Salamanca school discussed an early formulation of the Quantity Theory of Money in the sixteenth century. By the seventeenth, it was clear to several observers that the Spanish economy was in a process of secular decline. And by the eighteenth, David Hume held the view that "[S]ince the discovery of the mines in America, industry has increased in all the nations of Europe, except in the possessors of those mines" (Hume 1987/1742, p. 33). He believed that increased monetization benefited industrial development, at least temporarily: "labor and industry gain life; the merchant becomes more enterprising, the manufacturer more diligent and skillful, and even the farmer follows his plough with greater alacrity and attention." Additionally, monetization made it easier for the state to collect tax revenues: "[W]here money is . . . scarce . . . the prince can levy few or no taxes . . . it is evident that such a kingdom has little force even at home; and cannot maintain fleets and armies to the same extent, as if every part of it abounded in gold and silver." Hume is often interpreted as having argued that the effects are only transitory (e.g., Lucas 1996), though in my view this is not clear-cut, as this interpretation is supported by some passages but contradicted by others.

In this chapter, I review recent research which provides some answers. I argue that the precious metals had qualitatively different effects for first- and second-order European receivers and on different horizons. Spain (and from the late seventeenth-century onward, Portugal) were first-order receivers. These countries benefited from the inflows in the short run, but the continued arrival of precious metals ultimately caused a net negative effect which persisted for a long time as it damaged their export sector and ultimately their domestic political institutions. By contrast, some second-order receivers such as England benefited unambiguously from the additional availability of precious metals. Below, I explain the mechanisms behind these differential results.

From Precious Metals to Money

Figure 1 shows the variation over time in the quantities of silver and gold produced in the Americas during the early modern period. According to one estimate, New World production corresponded to 85% of the world's production of silver and 70% of gold during 1493–1800 (Barret 1990, p. 224). As Table 1 shows, over

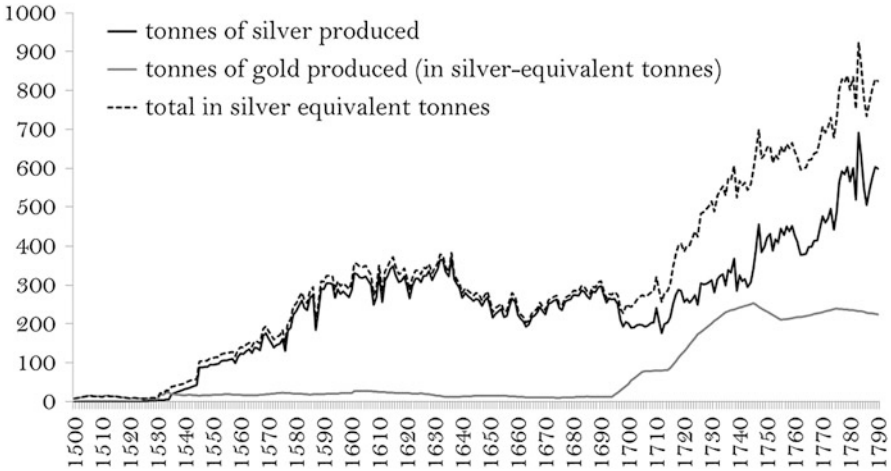


Fig. 1 Silver and gold (in silver-equivalent units) produced in the Americas. (Source: Palma (2019), based on the data in TePaske (2010))

Table 1 Gold and silver stocks and flows to Europe. Sources: For the initial stocks, sources as follows. In the case of the baseline European initial stock, it is calculated as 23% of the Velde-Weber figure, using the European share of global output in PPPs from the Maddison Project Database (MPD; Bolt et al. 2018). The countries included in the numerator are all the European countries currently present in the MPD with data for 1492 or earlier (England, Germany, France, Italy, the Netherlands, Spain, Portugal, Switzerland, Poland, and Greece). When the value for 1492 was missing, it was obtained from linear interpolation from the closest two observations. The denominator is the sum of the GDPs of all the countries in the MPD and was calculated in an analogous manner. Flows from Morineau (2009, p. 570). For bimetallic ratios (to convert gold to silver-equivalent units), I used 1:11 for the sixteenth century, 1:14 for the seventeenth, and 1:15 for the eighteenth. This is in line with the evidence discussed in Barret (1990, p. 238) or Spooner (1972, p. 21). The unit tones in the table refers to metric tons

	Fine silver, tons	Gold, tons	Gold, silver-equivalent tons	Total in silver-equivalent tons
Stocks in 1492				
World stock in 1492 (Velde and Weber 2000, p. 1230)	3,600	297	3,267	6,867
Baseline European stock (combining Velde and Weber with European output shares)	828	68	751	1,579
Alternative European stock (based on the stock of money figure by Glassman and Redish 1985, p. 40)	–	–	–	3,542
Imports to Europe				
Sixteenth century	7,500	150	1,650	9,150
Seventeenth century	26,168	158	2,212	28,380
Eighteenth century	39,157	1,400	21,000	60,157
Total imports	73,825	1,708	24,862	97,687

the sixteenth century alone, imports to Europe overtook the 1492 stock. Production only continued to increase over the next centuries.

Some discussion of the numbers of Table 1 is warranted. My 1492 baseline as well as that of Glassman and Redish are considerably smaller than the numbers that Braudel and Spooner (1967, p. 445) give for Europe in 1500. My estimate relies on the Velde and Weber world stock which may be too small due to assumptions such as 1% annual depreciation. Additionally, they as well as Glassman and Redish rely on rather old secondary sources. Still, their methodology is more solid than that of Braudel and Spooner, which, while endorsed by Parker (1973) and many others, is not well grounded in economics (Glassman and Redish 1985, p. 40). Finally, Morineau (2009, p. 571, 580) gives a figure of 15,000 t for 1500 – almost 10 times higher than the baseline I adopt here – which I do not consider credible, as it is calculated using arbitrary assumptions.

While precious metals (bullion) were not the same as money, silver and gold were key inputs for producing commodity money in the form of metallic coins. This input was available in Europe (from places such as Kutná Hora or Jáchymov in Bohemia) but under relatively inelastic supply. Once minted, coins typically circulated by tale – that is, by count (coins did not begin to have a numeric face value until relatively late). This meant that the purchasing power of a coin was usually superior to the market value of its melted materials. This was especially true of coins primarily made of copper, which mattered for small denominations and whose intrinsic value was typically negligible. As a result, coins of such small denominations were tokens which behaved more like fiat money than commodity money. But, at least in Western Europe and most of the time, even larger denominations circulated by tale, so the distinction between commodity money and fiat was never binary, since there was always an element of credibility to the value of currency. Paper money (usually backed by reserves hence not fiat) and other credit instruments which for functional purposes classify as money appeared gradually, but for most countries did not become an important part of the money supply until rather late (Palma 2018b).

In modern Western societies characterized by high levels of fiscal capacity and a large degree of independence in monetary policy, credit issued by commercial banks and placed in the deposit accounts of consumers and firms can be a substitute for liabilities of central banks in the form of currency. In such societies, fiat money circulates at no discount. But in early modern Europe (as in most premodern societies), there was no fiat money available and no independent monetary authority committed not to destabilize the value of coin existed. In premodern societies, where potential political instability and hence the specter of fiscal dominance were always present, coin and credit were complements, not substitutes. This was for the following reasons (Palma 2018b). First, personal credit instruments had to be based on private knowledge with respect to the creditworthiness of others, and so they often did not circulate even locally. Second, the decision of whether to issue credit to someone was in part based on the anticipation of whether the prospective debtor would have the liquidity to honor the bill at maturity and that in turn depended on the overall availability of coin. Third, notwithstanding the bill of exchange, due to information problems in many

contexts the transaction costs involved when making credit payments were higher than those of using coin (despite the lower transportation costs). And finally, coin and credit were also complements in a supply sense to the extent that reserves in the form of bullion or specie were used to back forms of credit – including paper (but non-fiat) money during the nineteenth century Classical Gold Standard. In such cases, credit was not a suitable substitute for coins.

Hence, without access to new silver and gold, sustained monetary expansion was difficult. Nevertheless, adjustments to the silver value of the monetary unit were sometimes required and made in the form of defensive debasements. Yet, such measures only served as temporary remedies at best. In worse cases where debasements were motivated by fiscal purposes, they could lead to significant depreciations and inflation (Karaman et al. 2019).

The availability of silver and gold in the Americas led to increased minting activity which helped overcome Europe's problems related to monetary scarcity (Day 1978). In the case of the Spanish Empire, silver was coined both in Spain and in the Americas. As time went by, the latter option became more frequent, and most silver ore was molded into bars (ingots) and sent to the nearest local mint where it was transformed into coins before being shipped to Spain (TePaske 2010). Precious metals were privately extracted and owned. The industry was taxed by the Crown but also received a variety of state support, for instance, in the subsidized provision of quicksilver (mercury), an important input in the production process. In the case of colonial Brazil, it was also the case that most remittances were sent as specie: more than three quarters of the gold arrived to Portugal already as coin (Costa et al. 2013, p. 63). The Iberian receiving countries used the inflows to finance the import of goods and military payments abroad. The second-order receiver countries (such as England) then either used these coins to trade with Asia – where Spanish dollars (*pesos*) were a standard means of payment and even circulated locally – or reminted them into their local currencies.

Figure 2 shows the correlation over time between the production of precious metals in the Americas (in silver-equivalent metric tons, i.e., tonnes) and nominal mint output in two European countries, France and Holland. The data available is usually not ideal, because mint output is a gross measure. This implies that some of the peaks are spurious since they correspond to recoinages, i.e., periods when old coins were recalled and substituted. A net measure is not available, however, except for the case of England (Palma 2018a) and, with stronger assumptions, Spain (Brzezinski et al. 2019). In spite of these data difficulties, Fig. 2 shows that an abundance of New World precious metals usually coincided with increased minting activities in Europe. This indicates that the amount of precious metals available to European countries indeed posed a binding constraint on monetary expansion.

The importance of New World precious metals for early modern Europe is also confirmed by the chemical analysis of the coins in circulation. The variation in the isotopic abundance composition of English coins reveals that much of the English silver coinage minted during the sixteenth to seventeenth centuries had a Spanish-American (and in particular, Andean) provenance (Desaulty and Albarede 2013). This evidence, which came to light only recently, is in line with the views of

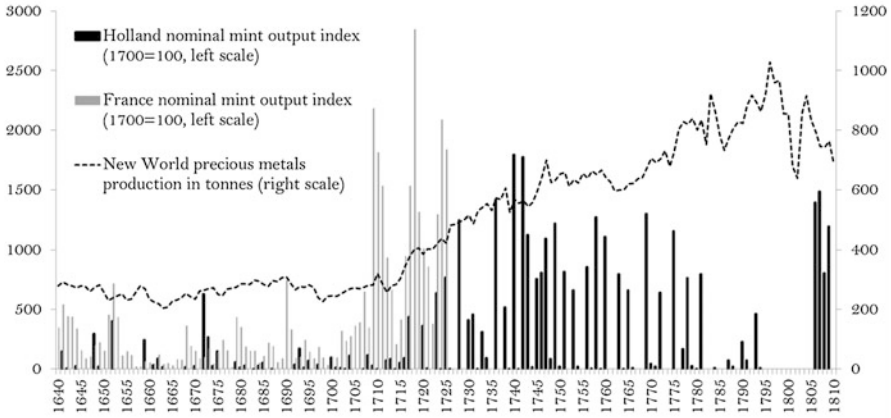


Fig. 2 New world precious production (in silver-equivalent metric tons) and nominal mint output in Holland and France. French data is only available until 1725 and zeros in the graph refer to missing data. (Sources: Palma (2019), relying on data by TePaske (2010), Spooner (1972, pp. 334–341), Stapel (2016), Zuijderduijn et al. (2018))

historians who argue that "there can be no doubt that Spanish bullion did come into the [English] mint, and did so in such quantity that . . . [at times] it formed the core of mint supply" (Challis 1978, p. 195; see also Mayhew 1999, p. 63). Indeed, for the English case, for which data on the coin stock is available, there is a close correlation between availability of precious metals and higher mint output, as one would expect (Palma 2018b, 2019).

Similar evidence is available (also in relation to other European countries) for the eighteenth century, when Mexican silver and gold, and Brazilian gold, became important (Barrandon et al. 1999; Desaulty et al. 2011). Some second-order receivers got a large share, in particular England, which received a disproportionate share of the silver in the decades after the 1630s (Palma 2018b), plus two-thirds of the Portuguese-Brazilian gold extracted over the eighteenth century (Fisher 1971). The fact that England got a large share may have had to do with country-specific characteristics but was also the result of exogenous geopolitical matters, as detailed in Palma (2018b). These also applied to the Netherlands (including the territories of modern Belgium), which were the receivers of large quantity of Spanish military payments, and also performed well during much of the early modern period.

Consequences for Second-Order Receiver Economies

For both first- and second-order receivers, additional money led to higher nominal and real GDP, and after a considerable delay, rises in the price level (Palma 2019). The fact that interest rates became lower (Brzezinski et al. 2019) and prices were

sticky interacted to cause real GDP to rise, and typically the effects were still positive several years later (Palma 2019).

As discussed, there were considerable short-to-medium run effects for first- and second-order receiving countries. But at least for England, there is evidence that money mattered beyond the medium-run as well. As Fig. 3 shows, from the 1630s, the scale of the increase in English coin supply was not matched by an increase in prices – and hence this was necessarily even less the case for less narrow measures of money. The reason does not have to do with changes in velocity, which was, to a first degree of approximation, stable over the early modern period (Palma 2018a). Instead, the cause for this development was economic growth.

The fact that a narrow measure of the money supply (the coin stock) increased as much as it did over time with a much more muted response from prices, as seen in Fig. 3, seems to contradict the Quantity Theory of Money (QTM). But that is not surprising. By assumption, the QTM holds real income and velocity constant in the long run. Now, consider the equation of exchange (MV=PY), which says that money times velocity equals nominal income, that is, the price level times real income. This is just an accounting identity, unlike the QTM. But mechanically, rewriting it as $P = MV/Y$ shows that under real growth, prices must not necessarily rise when there is an increase in money supply, even if V stays constant.

The relevant question is whether the increases in monetization themselves partly caused the levels of economic growth which were observed. I argue that this was the case. The increased availability of precious metals "allowed for a substantial increase in the monetization and liquidity levels of the economy decreasing transaction costs, increasing market thickness, changing the relative incentive

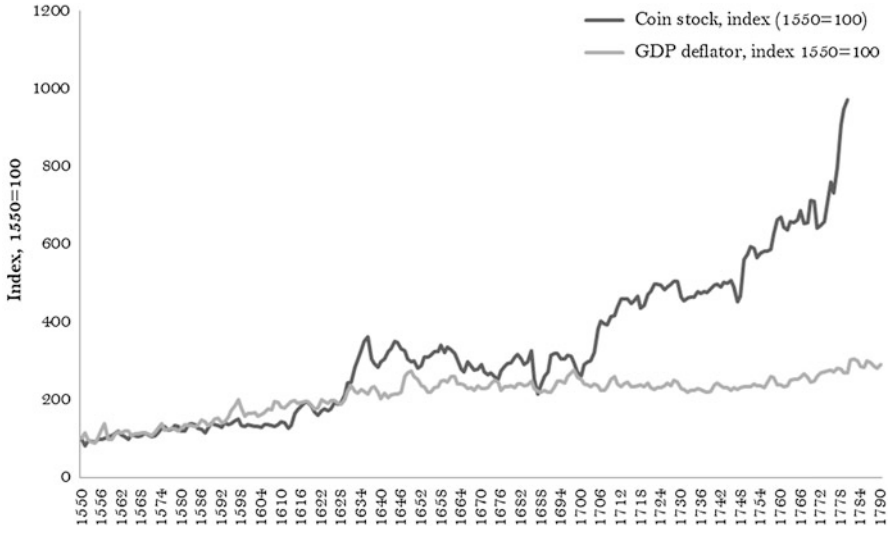


Fig. 3 Coin supply and price level in England, 1550–1790. (Source: coin supply from Palma (2018a, b); the price level is the GDP deflator of Broadberry et al. (2015))

for participating in the market, and allowing agglomeration economies to arise" (Palma 2018b). It additionally made trade with Asia possible on a much greater scale and contributed to high fiscal capacity levels by making tax collection easier. I now discuss each of these mechanisms in turn.

Before proceeding with a detailed study of England's case, it is necessary to introduce the concept of deep monetization. Lucassen (2014, p. 74) defines deep monetization as the existence of denominations which are equal to 1 h or less of waged work and exist in a per capita quantity of at least 5 h of waged work. He shows that the Netherlands were deeply monetized during parts of the early modern period, as well as after 1840. England became deeply monetized from the 1630s, at a time when the so-called Cottington treaty was signed (Palma 2018b) and when structural change in England started (Wallis et al. 2018).

Precious metals impacted England directly through increasing the country's money supply. This effect trickled down to small denominations which could be used for daily payments. Figure 4, which relies on new data on random coin finds, shows that small denominations – defined as one penny or less – first became more than half of the total coins lost from the reign of Charles I. This evidence contradicts earlier claims about widespread lack of small coinage (e.g., Muldrew 2008; Selgin 2008). Of course, coin was not as readily available as today – but the important point is that in England, from the mid seventeenth century coins became much more available than they were before, including those of relatively small denominations. This was sufficient to encourage market participation via a reduction of transaction costs (which supported additional market integration). Comparative work with other countries would be helpful, but so far it does not exist, except for the Netherlands (Lucassen 2014). However, evidence from early modern China suggests that the lack of a stable minting standard and the fact that the government was not sufficiently credible to mint coins (other than small-denomination copper coins) "imposed high transaction costs on market exchanges" (Ma 2013, p. 59; see also Ma and Zhao 2018). Monetary fragmentation had a similarly damaging effect on the countries which were part of the former Spanish empire, once it disintegrated (Irigoin 2009). Overall, the evidence suggests that a relatively efficient level of monetization is a necessary (but not sufficient) precondition for sustained economic growth to take place.

There were other ways in which American precious metals mattered. By making trade with Asia possible, precious metals induced demand for new desirable goods which could be imported, such as silver, porcelain, and tea (Palma and Silva 2016). Without American precious metals, early modern Euro-Asian trade would have been much smaller (Palma and Silva 2016). The tea, silk, and porcelain which arrived to Europe then mattered because they generated an industrious revolution (de Vries 2008) and encouraged industrialization (Berg 2005).

Increased monetization and market participation also made tax collection easier (Palma 2018b; see Besley and Persson 2011 for the importance of state capacity). This helped the government build up fiscal capacity and provide for public goods. Figure 5 shows the timing of the sustained increase of per capita government

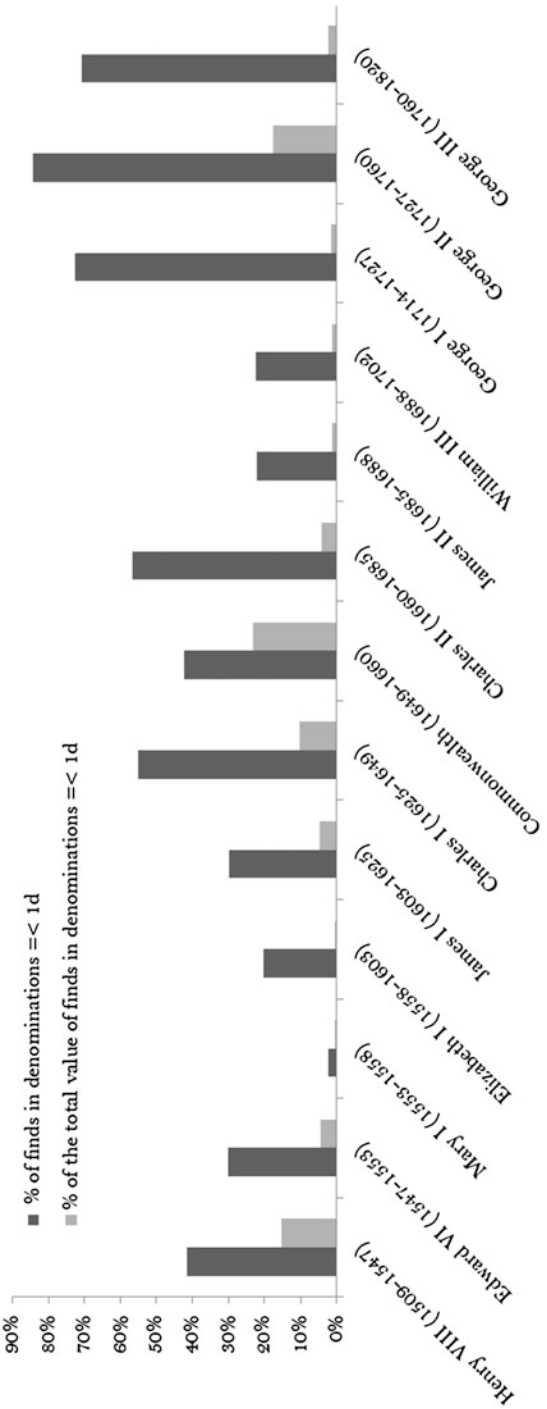


Fig. 4 Distribution of denominations for different periods. (Source: Palma 2018b)

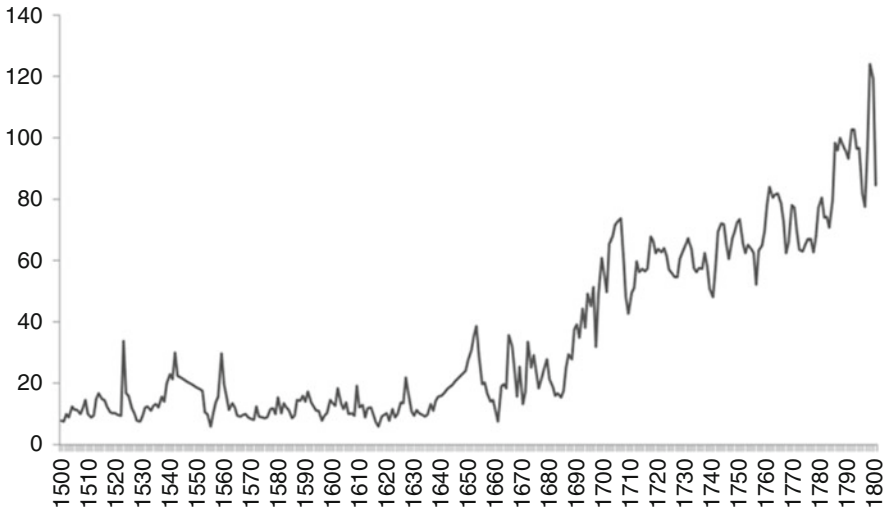


Fig. 5 English government real per capita revenues, 1500–1800. (Sources: Karaman et al. (2017). The unit is per capita tax revenue in grams of silver divided by a daily cost of the Allen (2001) respectability basket; hence the unit corresponds to "days of Allen's basket")

revenues in England (from the 1630s) matches well the timing of the monetization of the English economy (Palma 2018b).

Bonfatti et al. (2017) define monetary capacity as "government's capacity to issue sufficient liquidity for the markets to work properly. . . and for the collection of taxes to be efficient." England and the Netherlands are examples of a symbiotic relationship between high monetary and fiscal capacity, and the state failures of Qing China (or Poland) exemplify that complementarity not being present.

Finally, the additional wealth at impact to the Iberian first-order receivers meant increased demand for the tradable sector products of some second-order receivers (both in direct exchange for goods and indirectly in the context of military payments to Spanish troops in the Netherlands). As Cipolla (1993) writes, "[t]hrough both legal and smuggled imports, effective Spanish demand, sustained by American silver, promoted the economic development of Holland, England, and other European countries."

Overall, it is fair to say that we do not yet know enough about the diffusion and final distribution of the precious metals across Europe (and elsewhere). Nonetheless, as Braudel and Spooner (1967, p. 448) write, "Spanish silver, for political and commercial reasons, had preferred circuits. They greatly benefited northern Europe." Hence, although the precise consequences for the early modern economies (especially in the long-run) remain to be determined with more precision, it is likely they had widespread positive effects for several second-order receivers. As the historian Tim Blanning writes with respect to the eighteenth century, "the rapidly expanding output of Brazilian mines helped alleviate the chronic shortage

of coin and, among other things, allowed the stabilization of European currencies" (Blanning 2007, p. 95).

Consequences for First-Order Receiver Economies

First-order receivers (Spain and Portugal) benefitted from the same type of short-and-medium run effects as England did. As Palma (2019) shows, the impact of additional availability of American precious metals was stronger and faster (i.e., lags were shorter) than it was for second-order receivers. The magnitude of the effect is about double as large and the peak earlier than for the second-order receiving countries. However, these particularly strong effects came with a caveat: while the short and medium-term effects on income were stronger, the effects on the price level added up over time, and by 1600 the price level, measured in silver, had risen considerably more than in other European countries.

In the case of early modern Spain, it is also possible to see what happened when there were unexpected losses of ships from the silver fleets that were lost at sea. These ships were expected to arrive but were lost in maritime disasters (typically due to storms, not pirate attacks). Brzezinski et al. (2019) use these episodes as natural experiments to identify the effects of unanticipated money supply changes (i.e., shocks) on the Spanish economy. We show that 1 percentage point reduction in the money growth rate (or equivalently, a 1% reduction in the money supply) led to a 1.3% drop in real output that persisted for several years. The price level fell, and it did so permanently, but only with a lag. Additionally, tighter credit markets temporarily increased lending rates in Spanish cities, but not in other cities which can be thought of as placebos or control groups (see Brzezinski et al. 2019 for details).

The results of Palma (2019) and Brzezinski (2019) go in the same direction. They show that up to a few years after impact, additional precious metals had a positive effect for Spain – just as they typically did elsewhere in Europe. But we cannot conclude from this that the long-run effects were neutral. As time went by, the repeated arrival of massive amounts of silver from the New World changed the nature of the Iberian economies in two ways. First, Dutch disease set in: inflation caused the appreciation of the real exchange rate, which led to a loss of competitiveness of the export sector, in line with some of the arguments put forward by Hamilton (1934). Second, institutional resource curse set in. I will now consider each of these mechanisms in turn.

While early modern writers often point out that American precious metal mines could not be a sustained engine for long-run Spanish economic growth, and while not all aspects of Hamilton's work have aged well, overall his work represented a great advance both conceptually and empirically in our understanding of these issues. Recent work using alternative baskets for the calculation of inflation rates confirms Hamilton's general pattern, often finding that he even underestimated the inflation in the prices of many goods and services (González Mariscal 2015; Agudo 2019).

Dutch disease manifested itself, for instance, in Spain's woolen textiles industry, which declined steadily over the early modern period (for a classic work on the Mesta, the association of transhumant sheep ranchers, see Klein 1920; for a more recent and topical treatment, see the edited volume of Enciso 2001). While in the middle ages, Spain specialized in the production of finished textiles, during the early modern period, it gradually deurbanized, deindustrialized, and reverted to exporting raw wool (as well as other primary products such as fruit; see Reher 1990). Forsyth and Nicholas (1983) and Drelichman (2005a) argued along these lines, though more work would be welcome, using data on exchange rates, exports, and a sectoral breakdown of industries into tradable and non-tradable kinds, including at the local level.

The region of Spain which first received the metals, Andalusia, experienced the largest economic decline over 1530–1591. Being the richest region of continental Spain in 1530, it was one of the poorest as early as 1591. It then stagnated (Álvarez-Nogal and Prados de la Escosura 2007; p. 353). Spain as a whole declined from 1590 onward, with only a timid recovery during the eighteenth century (Álvarez-Nogal and Prados de la Escosura 2013). Spain and Portugal clearly diverged from Western Europe from the eighteenth century onwards, and would only enter a process of modern economic growth and convergence in the second half of the twentieth century. Portugal's economic activity also boomed during the period of growth in Brazilian gold production, but incomes then steadily declined from the mid-eighteenth century onward, as the percentage of the population working outside agriculture declined from 46.5% in 1750 to 33.1% in 1850 (Palma and Reis 2019).

Dutch disease-type problems could only have temporary consequences for economic growth if long-run monetary neutrality holds (or the closely related notion of "superneutrality," which relates to changes in growth rates rather than levels). Still, this is not necessarily the case if hysteresis-type effects were at play. Multiple equilibria could have existed, for instance, if the destruction of industry led to a new history-dependent pattern of trade networks. On top of this, and possibly even more importantly, the negative effects of the continued arrival of the precious metals for the long-run performance of the Spanish economy were likely not limited to Dutch disease-type problems, even when interacting with path-dependence and agglomeration economies, but were also related to institutional resource curse. Furthermore, the development literature suggests Dutch disease and institutional resource curse could positively interact too (Collier 2010).

The arrival of increasing amounts of precious metals was also associated with a decline in the quality of Spanish institutions over time, as some scholars suggest (e.g., Drelichman 2005b, 2007; Drelichman and Voth 2008). Acemoglu et al. (2005), Hough and Grier (2015), and Tilly (1992) confidently assert that Iberian institutions were already worse than those of England from the beginning of the early modern period. This was not the case (Henriques and Palma 2019). Iberian institutions were not worse than those of England around 1500, but instead, there was a gradual worsening of their quality over the early modern period.

Acemoglu et al. (2005) write that "The more rapid economic growth took place in societies with relatively non-absolutist initial institutions, most notably in

Britain and the Netherlands. In contrast, countries where the monarchy was highly absolutist, such as Spain and Portugal, experienced only limited growth in the subsequent centuries" (Acemoglu et al. 2005, p. 547). The claim that Spain and Portugal grew slower after 1500 can now be tested: While Maddison's numbers prior to 1820 were guesstimates (e.g., Maddison 2001), over the last decade, economic historians have reconstructed historical national accounts for the medieval and early modern periods from actual primary sources (Palma forthcoming). As Table 2 shows, the data rejects the idea that Spain and Portugal grew less, at least until the mid-seventeenth century. It was instead England (and Holland) which did not grow (until 1650 and 1700, respectively). After those dates the two countries did experience growth, and in particular England experienced fast growth by the standards of the time. By contrast, Portugal, which had experienced fast growth until c. 1750, initiated a period of fast decline afterward. But the important thing to retain is that this timing of events does not fit with the "initial" institutions hypotheses of Acemoglu et al. (2005), Hough and Grier (2015), or Tilly (1992).

A comparison between England, Portugal, and Spain shows that the dominance of England was not clear-cut at the start of the early modern period (Fig. 6). Due to the uncertainties associated with the building of truly comparable Purchasing Power Parity benchmarks (Jong and Palma 2018), when interpreting these numbers, it is best to place emphasis on changes in growth rates over time and not on any small differences on income levels. For instance, using the same growth rates over time but instead going backward from the alternative early nineteenth century income benchmarks of Prados de la Escosura (2000) would imply that Spain and Portugal were richer than England until the second half of the seventeenth century (Palma and Santiago-Caballero 2020). With these limitations in mind,

Table 2 Average annual per capita real growth 1530–1800. For England, Broadberry et al. (2015); for Holland, van Zanden and van Leeuwen (2012); for Spain, Álvarez-Nogal and Prados de la Escosura (2013); for Portugal, Palma and Reis (2019); for Italy, Malanima (2011); for Poland, Malinowski and van Zanden (2017); for France, Ridolfi (2016); and for Sweden, Krantz (2017) and Schön and Krantz (2012). Annualized growth rates were calculated using the familiar compound growth formula. As per the available data, for France dates are not until 1800 but 1790; for Poland they are not until 1800 but until 1795

	Annual real p.c. growth, 1530–1650	Annual real p.c. growth, 1530–1700	Annual real p.c. growth, 1530–1750	Annual real p.c. growth, 1530–1800
England	−0.16%	0.20%	0.20%	0.24%
Spain	−0.14%	0.00%	−0.02%	0.04%
Portugal	0.03%	0.12%	0.24%	0.05%
Holland	0.20%	0.00%	0.05%	0.08%
Poland	−0.05%	0.01%	0.05%	−0.01%
France	0.06%	0.08%	0.05%	0.07%
Sweden	0.00%	0.19%	0.00%	−0.04%
Italy	0.00%	0.03%	0.04%	−0.01%

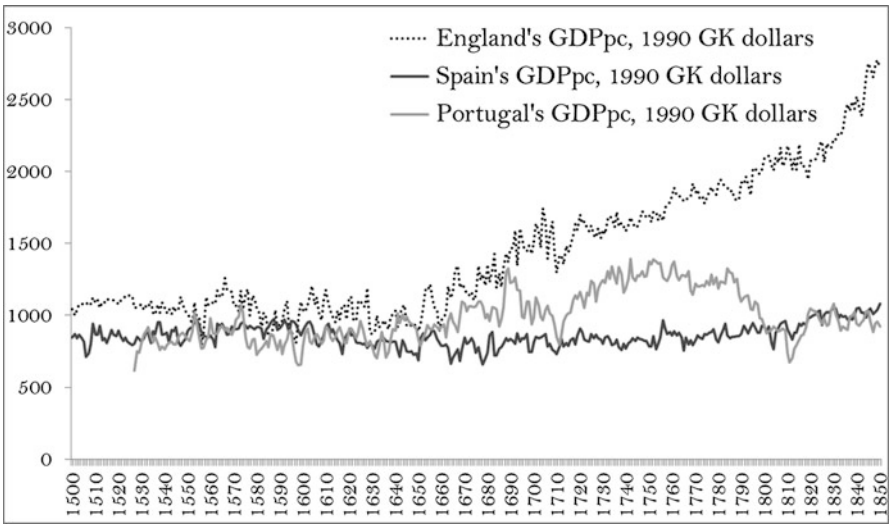


Fig. 6 GDP per capita in constant, 1990 "international" Geary-Khamis dollars for England, Spain, and Portugal, 1500–1850. (Sources: for England, Broadberry et al. (2015); for Spain, Alvarez-Nogal and Prados de la Escosura (2013); for Portugal, Palma and Reis (2019)) In the case of England, levels are extrapolated backwards from the 1870 level for Great Britain, with growth rates corresponding to the borders of England until 1700 and Great Britain afterwards. Note: Changes in growth rates over time, not small differences in income levels, is what needs to be noticed here, because a change in 19th c. benchmarks can lead to a reversion of earlier levels (i.e. each series as a whole shifting up or down) as explained in the text

Fig. 6 suggests that no significant differences existed between either the levels or the growth rates of Iberian and English GDP before the seventeenth-century.

What Fig. 6 does suggest is that the continued arrival of precious metals was temporally correlated with negative economic performance for the Iberian economies – not at impact but with a lag of about four decades. American silver and gold (while discovered earlier) started arriving in large quantities from the 1540s to Spain (TePaske 2010) and from the 1710s to Portugal (Costa et al. 2013). Both countries experienced good economic performance in the first 40 years from the arrival of large quantities of precious metals (mostly silver in the Spanish case and gold in the Portuguese case) but strong economic decline afterward.

Why have these economies worsened so drastically after the arrival of precious metals? In the case of early modern Spain, Drelichman (2005b, 2007) suggests that political institutions worsened as a consequence of the arrival of the silver, a view which finds modern-day parallels in the follow-up to the commodity booms in many poor countries (Collier 2010). But is there evidence that this was a direct result of the import of precious metals? Even given the macroeconomic performance documented in Table 2 and Fig. 6, could it be that Acemoglu et al. (2005) and Hough and Grier (2015) were right that Iberian institutions were already

worse in the Middle Ages, so the early modern period was simply a continuation of the previous institutional trend?

Institutions are notably difficult to measure quantitatively, which makes cross-country comparisons difficult. But in recent work, Henriques and Palma (2019) build measures which allow for an explicit comparison of institutional quality over time. The paper considers several potential measures including the number of crown refusals at legislative assemblies, the number and strength of episodes of coin depreciation, the number and strength of ad hoc taxes introduced, and real interest spreads for public debt. All of these suggest that the political institutions of Spain (at least as measured by Castile) and Portugal were by no means worse than those of England until the mid-seventeenth century. Hence, while North (1990) was correct in assuming that Iberian institutions eventually did became worse than those of England, this was not yet true around 1500 or even 1600.

For example, consider interest rates paid on long-run public borrowing. North and Weingast (1989) argue that the fall in English interest rates observed following the Glorious Revolution of 1688–1689 can be seen as evidence of the higher credibility of the post-revolution regime, since less credible regimes carry a risk premium which internalizes the possibility of default. Their analysis is not comparative, but a comparison of the situation in England, Holland, Portugal, and Spain shows that if this argument is taken at heart, then English political institutions only converged to Iberian credibility levels over the second half of the seventeenth century. England was first unable to issue long-term debt (until 1650), and even afterward, it paid higher real interest rates than those paid in Spain and Portugal until after 1710 (Henriques and Palma 2019). For the importance of 1710, see Macdonald (2013).

Public long-term interest rates are just one illustrative example, as most of the Henriques and Palma (2019) measures in fact show that at the start of the eighteenth century, a divergence in institutional quality was already taking place. For instance, parliament – which sometimes exerted executive constraints – met increasingly frequently over the second half of the seventeenth century, and it became permanent during the eighteenth. By contrast, in Spain and Portugal parliament (the *Cortes*) rarely met over the 1675–1800 period, while in earlier periods, meetings were much more frequent.

What were the fundamental causes of the comparative worsening of quality of Iberian institutions? The first part of the explanation lies with England's success rather than Iberian failure per se. Once we set aside the idea that there was anything special about England's economy before the seventeenth century, it becomes clear that on a political level, there were two stages to England's success. The most well-known and emphasized by a variety of authors including North and Weingast (1989) is the Glorious Revolution of 1688–1689. This was, however, preceded by important fiscal reforms emphasized by O'Brien (1988, 2011). Even Pincus and Robinson (2014), in their defense of the Glorious Revolution which lays emphasis on parliamentary sovereignty and the Whig program of modernization, recognize that "the English fiscal-military state had been growing by leaps and bounds since the 1640s."

A second part of the explanation as to why Iberian institutions worsened comparatively lies in the forces that acted directly on Portugal and Spain. The fact that the reasons of the eventual divergence of Iberian political and economic performance are not medieval but must instead be found during the early modern period makes American precious metals a plausible causal candidate, as argued by scholars such as Drelichman (2005b, 2007) for the Spanish case and Macedo (1982) for the Portuguese case. Precious metals are a better candidate than the influence of colonial trade generally, because the relationship between colonial trade and institutions is more nuanced than it may appear at first glance. This can be illustrated with the case of Portugal. Tilly (1992, p. 62) claims that the fiscal dependence of the Portuguese crown on maritime revenues meant that no representative institutions developed. Tilly (1992, p. 124–5) even compares early modern Portugal to today's oil-exporting countries where, thanks to colonial revenues, the leaders have much autonomy vis-à-vis the general population because they do not need to collect fiscal revenues by taxing a broad tax base. In fact, Portuguese fiscal capacity was comparable to that of other Western European countries, and empire-related revenues were typically not a large share of total state revenues, except during parts of the sixteenth century (Costa et al. 2019). Instead, the colonial empire grew to have a large effect on both fiscal revenues and the economy during the country's less representative period, the eighteenth century (Costa et al. 2015).

Conclusion

There is no contradiction between the view that precious metals had different effects for first- and second-order receivers and also different effects at different time horizons. Precious metals had a positive effect for the English economy, and possibly other second-order receivers such as the Netherlands, in both the short and long run (Palma 2018b, 2019). It also had a positive effect for Spain and Portugal in the short run (Palma 2019; Brzezinski et al. 2019). They were, however, likely to have had net negative long-run consequences for the first-order receivers, the Iberian economies. This is because, for these economies, the indirect negative effects (the interaction of Dutch disease, institutional resource curse, and comparatively deficient growth in sustainable fiscal capacity) eventually overtook the positive consequences of monetization which also took place in some second-order receivers such as England.

Early modern European history teaches us that the relationship between money and economic growth is more nuanced than some monetary neutrality narratives suggest. It is now widely accepted among most macroeconomists that money is not neutral over a business-cycle frequency, though prominent deniers remain. The experience of early modern Europe suggests that the majority are right and additionally that the magnitude of monetary non-neutrality effects (which is more debated) can be large.

By contrast, most macroeconomists believe in the notion of long-run monetary neutrality (and superneutrality, which concerns growth and inflation rather than

changes in levels). But the increase in money supply made possible by the increased availability of precious metals from America did lead to higher economic growth in early modern England, with a visible but muted inflation response (by comparison with the magnitude of the arrivals). The financial history of early modern England suggests that in societies with a deficient provision of liquidity, additional money supply can induce economic growth through a variety of channels which include lower transaction costs. These effects were also present for first-order receivers of the precious metals (Spain and Portugal), but in these cases, they were eventually dominated by negative and indirect consequences in the forms of Dutch disease and institutional resource curse.

Cross-References

► Money Markets and Exchange Rates in Preindustrial Europe

References

Acemoglu D, Johnson S, Robinson J (2005) The rise of Europe: Atlantic trade, institutional change, and economic growth. Am Econ Rev 95(3):546–579

Agudo DG (2019) Prices in Toledo (Spain): sixteenth and seventeenth centuries. Soc Sci Hist 43(2): 269–295

Allen RC (2001) The great divergence in European wages and prices from the Middle Ages to the First World War. Explorations in Economic History 38(4):411–447

Álvarez-Nogal C, Prados De La Escosura L (2007) The decline of Spain (1500–1850): conjectural estimates. European Review of Economic History 11:319–366

Álvarez-Nogal C, Prados De La Escosura L (2013) The rise and fall of Spain (1270–1850). Econ Hist Rev 66(1):1–37

Barrandon J, Morrisson C, Cécile M (1999) Or du Brésil, monnaie et croissance en France au XVIIIe siècle. Preface by Le Roy Ladurie Emmanuel. Cahiers Ernest Babelon, vol 7. CNRS Editions, Paris

Barret W (1990) World bullion flows, 1450–1800. In: The rise of merchant empires. Cambridge University Press, Cambridge

Berg M (2005) Luxury and pleasure in eighteenth-century Britain. Oxford University Press, Oxford

Besley T, Persson T (2011) Pillars of prosperity: the political economics of development clusters. Princeton University Press, Princeton

Blanning TC (2007) The pursuit of glory: Europe, 1648–1815. Penguin, New York

Bolt J, Inklaar R, de Jong H (2018) Rebasing Maddison: relative prices and the shape of long-run global development. GGDC Research Memorandum, vol GD-174. Groningen Growth and Development Centre

Bonfatti R, Karaman KK, Palma N (2017) Monetary capacity. Unpublished manuscript

Braudel F, Spooner F (1967) Prices in Europe from 1450 to 1750. Camb Econ Hist Eur 4:374–486

Broadberry S, Campbell B, Klein A, Overton M, van Leeuwen B (2015) British economic growth, 1270–1870. Cambridge University Press, Cambridge

Brzezinski A, Chen Y, Palma N, Ward F (2019) The vagaries of the sea: Evidence on the real effects of money from maritime disasters in the Spanish Empire. *CEPR Discussion Paper 14089*

Challis CE (1978) The Tudor coinage. Manchester University Press, Manchester

Cipolla CM (1993) Before the Industrial Revolution: European society and economy 1000–1700. Routledge, London
Collier P (2010) The plundered planet: why we must – and how we can – manage nature for global prosperity. Oxford University Press, Oxford
Costa LF, Rocha MM, de Sousa RM (2013) O ouro do Brasil. Imprensa Nacional-Casa da Moeda, Lisboa
Costa LF, Palma N, Reis J (2015) The great escape? The contribution of the empire to Portugal's economic growth, 1500–1800. Eur Rev Econ Hist 19(1):1–22
Costa LF, Henriques AC, Palma N (2019) Portugal's early modern state capacity: a comparative approach. Unpublished manuscript
Day J (1978) The great bullion famine of the fifteenth century. Past Present 79:3–54
de Macedo JB (1982) Problemas de História da Indústria Portuguesa no Século XVIII, 2nd edn. Lisbon, Querco
de Vries J (2008) The industrious revolution: consumer behavior and the household economy, 1650 to the present. Cambridge University Press, Cambridge
Desaulty AM, Albarede F (2013) Copper, lead, and silver isotopes solve a major economic conundrum of Tudor and early Stuart Europe. Geology 41(2):135–138
Desaulty AM, Telouk P, Albalat E, Albarede F (2011) Isotopic Ag–Cu–Pb record of silver circulation through 16th–18th century Spain. Proc Natl Acad Sci 108(22):9002–9007
Drelichman M (2005a) The curse of Moctezuma: American silver and the Dutch disease. Explor Econ Hist 42:349–380
Drelichman M (2005b) All that glitters: Precious metals, rent seeking and the decline of Spain. Eur Rev Econ Hist 9:313–336
Drelichman M (2007) Sons of something: taxes, lawsuits, and local political control in sixteenth-century castile. J Econ Hist 67(3):608
Drelichman M, Voth H-J (2008) Institutions and the resource curse in early modern Spain. In: Helpman E (ed) Institutions and economic performance. Cambridge University Press, Cambridge
Enciso AG (ed) (2001) El negocio de la lana en España (1650–1830). Ediciones Universidad de Navarra S.A, Pamplona
Fisher (1971) The Portugal trade: a study of Anglo-Portuguese commerce 1700–1770. Methuen, London
Forsyth P, Nicholas S (1983) The decline of Spanish industry and the price revolution: a neoclassical analysis. J Eur Econ Hist 12:601–610
Glassman D, Redish A (1985) New estimates of the money stock in France, 1493–1680. J Econ Hist 45(1):31–46
González Mariscal M (2015) Inflación y niveles de vida en Sevilla durante la Revolución de los Precios. Revista de Historia Económica/J Iber Lat Am Econ Hist 33(3):353–386
Hamilton EJ (1934) American treasure and the price revolution in Spain, 1501–1650. Harvard University Press, Cambridge, MA
Henriques AC, Palma N (2019) Comparative European Institutions and the little divergence, 1385–1850. CGEH Working Paper, University of Utrecht
Hough JF, Grier R (2015) The long process of development: building markets and states in pre-industrial England, Spain and their colonies. Cambridge University Press, New York
Hume (1987/1742) Essays, moral, political, and literary. Liberty Fund, Indianapolis. Accessed May 2014 from http://www.econlib.org/library/LFBooks/Hume/hmMPL26.html
Irigoin MA (2009) Gresham on horseback: the monetary roots of Spanish American political fragmentation in the nineteenth century. Econ Hist Rev 62(3):551–575
Jong H, Palma N (2018) Historical national accounting. In: Blum M, Colvin C (eds) An economist's guide to economic history. Palgrave Macmillan, Basingstoke
Jordà O (2005) Estimation and inference of impulse responses by local projections. Am Econ Rev 95(1):161–182

Karaman KK, Pamuk S, Yildirim-Karaman S (2017) Long run patterns in fiscal capacity. Bogazici University, Mimeo
Karaman K, Pamuk S, Yıldırım-Karaman S (2019) Money and monetary stability in Europe, 1300–1914. *Journal of Monetary Economics*
Klein J (1920) The Mesta: a study in Spanish economic history, 1273–1836. Harvard University Press, Cambridge
Krantz O (2017) Swedish GDP 1300–1560: a tentative estimate. Lund papers in economic history, vol 152. Lund University
Lucas RE (1996) Nobel lecture: monetary neutrality. J Polit Econ 104(4):661–682
Lucassen J (2014) Deep monetisation: the case of the Netherlands 1200–1940. Tijdschr Soc Economische Geschiedenis 11(3):73
Ma D (2013) Chinese money and monetary system, 1800–2000, overview. In: Caprio G (ed) Handbook of key global financial markets, institutions, and infrastructure, vol 1 (eds: Calomiris C, Neal L). Elsevier, Oxford, pp 57–64
Ma D, Zhao L (2018) A silver transformation: Chinese monetary integration in times of political disintegration during 1898–1933. LSE Economic History Working Paper No. 283
Macdonald J (2013) The importance of not defaulting: the significance of the election of 1710. In: Questioning credible commitment: new perspectives on the glorious revolution and the rise of financial capitalism. Cambridge University Press, Cambridge
Maddison A (2001) The world economy: a millennial perspective. OECD, Paris
Malanima P (2011) The long decline of a leading economy: GDP in central and northern Italy, 1300–1913. Eur Rev Econ Hist 15:169–219
Malinowski M, van Zanden JL (2017) Income and its distribution in preindustrial Poland. Cliometrica 11(3):375–404
Mayhew NJ (1999) Sterling: the rise and fall of a currency. Lane, Allen
Morineau (2009) Incroyables Gazettes et Fabuleux Métaux: Les Retours des Trésors Américains d'Aprés les Gazettes Hollandaises (XVI–XVII Siècles). Cambridge University Press, London. Originally published 1985
Muldrew C (2008) Wages and the problem of monetary scarcity in early modern England. In: Lucassen J (ed) Wages and currency. Global comparisons from antiquity to the twentieth century wages and currency. International and comparative social history, vol 10. Verlag Peter Lang, Bern
North DC, Weingast BR (1989) Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. J Econ Hist 49(4):803–832
O'Brien PK (1988) The political economy of British taxation, 1660–1815. Econ Hist Rev 41:1–32
O'Brien PK (2011) The nature and historical evolution of an exceptional fiscal state and its possible significance for the precocious commercialization and industrialization of the British economy from Cromwell to Nelson. Econ Hist Rev 64(2):408–446
Palma N, Silva AC (2016) Spending a Windfall: American Precious Metals and Euro-Asian Trade 1531–1810. *GGDC Research Memorandum 165*
Palma N (2018a) Reconstruction of money supply over the long-run: the case of England, 1270–1870. Econ Hist Rev 71(2):373–392
Palma N (2018b) Money and modernization in early modern England. Financ Hist Rev 25(3):231–261
Palma N (2019) The existence and persistence of monetary non-neutrality: evidence from a large-scale historical natural experiment. University of Manchester discussion paper
Palma N (forthcoming) Historical account books as a source for quantitative history. In: Christ G, Roessner P (eds) History and economic life: a student's guide to approaching economic and social history sources. Routledge, London
Palma N, Reis J (2019) From convergence to divergence: Portuguese economic growth, 1527–1850. J Econ Hist 79(2):477–506

Palma N, Santiago-Caballero C (2020) Patterns of Iberian Economic Growth in the Early Modern Period. Forthcoming in: An Economic History of the Iberian Peninsula, 700–2000 (general editor: Pedro Lains)
Parker G (1973) The emergence of modern finance in Europe, 1500–1700. In: Cipolla CM (ed) The Fontana economic history of Europe. Collins, London
Pincus SC, Robinson JA (2014) What really happened during the glorious revolution? In: Galiani S, Sened I (eds) Institutions, property rights, and economic growth: the legacy of Douglass North. Cambridge University Press, New York
Prados de la Escosura L (2000) International comparisons of real product, 1820–1990: an alternative data set. Explor Econ Hist 37(1):1–41
Reher D (1990) Town and country in pre-industrial Spain: Cuenca, 1540–1870. Cambridge University Press, Cambridge
Ridolfi L (2016) The French economy in the longue durée. A study on real wages, working days and economic performance from Louis IX to the Revolution (1250–1789). PhD dissertation. IMT School for Advanced Studies, Lucca
Schön L, Krantz O (2012) The Swedish economy in the early modern period: constructing historical national accounts. Eur Rev Econ Hist 16(4):529–549
Selgin GA (2008) Good money: Birmingham button makers, the royal mint, and the beginnings of modern coinage, 1775–1821. University of Michigan Press, Ann Arbor
Spooner FC (1972) The international economy and monetary movements in France, 1493–1725. Harvard University Press, Cambridge
Stapel R (2016) Coin production in the low countries 14th–19th c. http://hdl.handle.net/10622/D5WXZZ, IISH Dataverse
TePaske J (2010) A new world of gold and silver. Edited by Kendall W. Brown, Brill
Tilly C (1992) Coercion, capital and European states, AD 990–1992. Blackwell, Cambridge, MA
van Zanden JL, van Leeuwen B (2012) Persistent but not consistent: the growth of national income in Holland 1347–1807. Explor Econ Hist 49:119–130
Velde F, Weber W (2000) A model of bimetallism. J Polit Econ 108(6):1210–1234
Wallis P, Colson J, Chilosi D (2018) Structural change and economic growth in the British economy before the Industrial Revolution, 1500–1800. J Econ Hist 78(3):862–903
Zuijderduijn J, Stapel R, Lucassen J (2018) Dataset: coin production in the low countries: fourteenth century to the present. Tijdschr Soc Economische Geschiedenis/Low Countries J Soc Econ Hist 15:2–3

Rise and Demise of the Global Silver Standard

15

Alejandra Irigoin

Contents

Abstract

In the early modern period, the world economy gravitated around the expansion of long-distance commerce. Together with navigation improvements, silver was the prime commodity which moved the sails of such trade. The disparate availability and the particular demand for silver across the globe determined the participation of producers, consumers, and intermediaries in a growing global economy. American endowments of silver are a known feature of this process; however, the fact that the supply of silver was in the form of specie is a less known aspect of the integration of the global economy. This chapter surveys the production and export of silver specie out of Spanish America, its intermediation by Europeans, and the reexport to Asia. It describes how the sheer volume produced and the quality and consistency of the coin provided familiarity with, and reliability to, the Spanish American peso which made it current in most world markets. By the eighteenth century, it has become a currency standard for the

A. Irigoin (✉)
Economic History, London School of Economics and Political Sciences, London, UK
e-mail: M.A.Irigoin@lse.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_13

international economy which grew together with the production and coinage of silver. Implications varied according to the institutional settings to deal with specie and foreign exchange in each intervening economy of that trade. Generalized warfare in late eighteenth-century Europe brought down governance in Spanish America and coinage fragmented along with the political fragmentation of the empire. The emergence of new sovereign republics and the end of minting as known meant the cessation of the silver standard that had contributed to the early modern globalization.

Keywords
Silver specie · International currency · Monetary capacity · Currency trade · Global Smithian growth

Arguably, before the Gold Standard, there was another monetary standard for the international economy, which bore similar currency but little resemblance with the institutions of the gold-backed pound sterling. In the early modern period, a coin of silver made in the New World was the base on which prices and exchange rates were established in far distant places as Leghorn (Livorno), Smyrna, Kingston, Bourbon Island, Surat, Manila, Macao, Cadiz, Havana, Lima, Philadelphia, Buenos Aires, and Bantam. Equally well known to Cotton Mather and to Alexander Hamilton or to the Mughal in Agra and Louis XIV of France, the Spanish American silver coin changed names and some features throughout the sixteenth, seventeenth, and eighteenth centuries but remained the most successful world money before the nineteenth century. Coins manufactured in Spanish America were known in France as piasters, *duros* or "hard" pesos in Spain, or *ryals* and Spanish dollars in the English world, whereas indeed the name is peso. Here both words are used interchangeably. Coins were demanded by Canton Customs for the "measurement" and demurrage charges in Whampoa; it was the unit of account for bills on London of the cash-strapped English East India Company and of the Dutch, French and Danish West Indies. It made the largest share of the Bank of Amsterdam collateral in the 1763 crisis and the numeraire of international movement of capital of the Jesuit Order and most ecclesiastical investments overseas; it was the unit of denomination of the Continental Congress paper money and the template on which Alexander Hamilton anchored the US dollar and the money of account of British Singapore well into the 1830s. In a world where the geography of money revolved more around individual cities than nationwide states (Flandreau et al. 2009), the coin minted separately in three sites far apart in the Spanish New World was the first currency everywhere. As the Company factor John Lockyer indicated for Asia "dollars (pesos) are worth more in specie than when they are melted down" (Lockyer 1711, p. 141).

Together with the discovery of the sea route to India, Adam Smith considered the discovery of the New World was one of the "two greatest and most important events recorded in the history of mankind" (Smith 1776/2007, III, ch 7, p. 829). Indeed, it meant a defining change for the Old (Eurasian) World economies. Thus, the New

World is associated with a number of windfalls to their discoverers, like the "Columbian exchange" (Crosby 1972), the "ghost acreage" (Pomeranz 2000), and improvements to European living standards (Hersh and Voth 2009), all of which allowed Europeans to overcome the Malthusian limits to their path of development. Silver was paramount in this transformation. In Smith's words it was "one of the principal commodities by which the commerce between the two extremities of the world was carried on, and by means of it, in a great measure, that those distant parts of the world are connected with one another" (Smith 1776/2007, I, ch 11, p. 287).

The Spanish American Silver

Unlike spices and exotic foodstuff like cacao, tomato, or maize, gold and silver were known to Eurasian people and rulers. But the acquisition of the American continent allowed an access to precious metals of unimagined proportions. In Europe, the German silver mining boom of the mid-sixteenth century had increased fivefold the continent silver output. At its peak in the 1540s, output in Saxony, Thuringia, Bohemia, Slovakia, and Hungary combined amounted to 52–55 t per annum (Munro 2012). This volume was six times larger than the incipient inflow of American silver to Seville at the time. By the 1560s quantities were comparable; in the next decade 1560s–1570s, European production was probably only half of New World's inflows. Thereafter, silver mining in Central Europe stagnated or declined, and American silver flooded the European and world economy once Potosi in Upper Peru (Bolivia today) and Zacatecas (Mexico today) went into production. Still in the 1790s, Baron Humboldt was bewildered that one single mine in Guanajuato – the "La Valenciana" – could regularly supply 30,000 marks of silver per month, a quantity equal to "half of what is furnished by all the mines of Saxony (combined)" (1801/1811, pp. 171, 174). In Asia, Japan was too a very significant producer, and her maximum output in the early seventeenth century has been estimated in 150–200 t a year – but declined sharply to 60 t by the 1640s and remained very low thereafter. So at a steady annual average of 350 t for more than 250 years, Spanish American silver dwarfed the contemporary production of Central and Northern Europe, Central Asia, and Southeast Asian mines in Siam, Burma, and Cochin China (Vietnam) combined. A total of 86,000 t of silver and 1700 t of gold was the total volume of precious metals mined in the Iberian possessions up to the 1800s. This represented approximately 70–80% of the world stock of silver and 40% of the world's gold stock throughout the eighteenth century. Scholars use Soetbeer (1879) and Barret (1990) estimates, but their figures include "silver equivalent" values – i.e., gold expressed in silver pesos. Data used in this essay originated from the registered metal which was assayed, paid taxes, or was coined; figures here should also be considered an absolute lowest bound (TePaske and Brown 2010, pp. 53, 67, 140).

This increment in the production and shipment overseas did not slacken until 1808 when Napoleon invasion of Spain brought down her rule in Spanish America's mainland, but output and exports lapsed only after the 1820s. It had started by the

mid-sixteenth century when large silver veins found in the mainland outshone the considerable – by standards of the time – gold mining in the Spanish Caribbean soon depleted by 1550. Thereafter, New Granada, today Colombia, became the most consistent producer of gold in the Spanish New World. With varying relative shares, production continued in silver-endowed areas like Mexico and Bolivia today – and boomed in Chile by the end of the Spanish rule.

Several other rich mining districts sprang up over time fuelling and financing the territorial expansion of the Spanish American economy. With a total population of about 12 million inhabitants – roughly the same as Spain's and less than of England's – by 1800 the colonial economy stretched from the new mining regions of Chihuahua and Sonora (near the Mexico-United States border today) in the north to Copiapo and Coquimbo in Chile, in the south. Silver-rich New Spain (Mexico) counted a number of additional very productive mines scattered throughout the territory, i.e., San Luis Potosi, Durango, Guanajuato, Guadalajara, and the central region neighboring Mexico City, the main administrative center in the Spanish kingdom. In the second half of the eighteenth century, other secondary sites flourished in Pachuca, Sombrerete, Bolaños, and Rosario and in marginal regions in Central America like today Guatemala (see map in "Appendix"). A substantial gold rush since the 1690s in Portuguese Brazil added nearly 1100 t of fine gold to the precious metals exported out of the New World – more than half of the total production until 1800. Yet, silver was consistently the chief commodity out of the Americas, for much larger values and for longer periods of time than any other export commodity produced in America with African slave labor like sugar of coffee.

Endowments were far more concentrated in South America – the "Rich Mountain" of Potosi produced 80% of the New World silver outside Mexico and sourced half or more of the world total during the seventeenth century. Minor sites – in proportion – in today Peru, Bolivia, and Chile also contributed to Peruvian output although some regions boomed, and a few declined too after few years, like Carangas, Cailloma, and Castrovirreyna. Still at the turn of the nineteenth century, other sites were incorporated like Norte Chico (Chile), Oruro (Bolivia), and Pasco (Peru). Peru also contained the very large quicksilver mines of Huancavelica, which supplied a vital input for silver refining. Amalgamation – the blending of pulverized ore with quicksilver to be washed away with salted water – replaced indigenous cupellation with blast furnaces in the refining and smelting of silver in Mexico in the 1560s and in Potosi in 1573. Together with reforms that instated the Spanish *mita* on indigenous labor in 1570 – known as *Ordenanzas del Virrey Toledo* – this new technology doubled production within 20 years, to 2.5 million pesos or 62 t of fine silver. Associated with the exploitation of Indian labor, the extraction of silver is a well-known – though ill-understood – feature of the European rule in the New World. Wrongly, institutionalist economists often assume the coercive nature of this labor institution, though the collective property rights of indigenous to land were instrumental during the demographic crisis, but eventually raised the incentives to work for a wage when population replenished (Coatsworth 2006).

A stereotypical characterization of Spanish colonialism has overlooked other remarkable aspects of the silver economy in the early modern Spanish American world. For instance:

1. The nearly complete *private* nature of the industry, which the monarchy taxed in proportion of the output given her domain of the subsoil.
2. The existence of miners of very diverse scale – predominantly small and only few, large, mining companies like the "La Valenciana" which appeared later in eighteenth-century Mexico.
3. The subsidization of the industry by the Royal Treasury which afforded inputs, like gunpowder and quicksilver, at lower than market prices.
4. The private-royal partnership – a feature of Spanish colonialism that extended to most other capacities of the state – that supplied quicksilver from Almaden or imported it from Germany at the expense of the royal purse, whereas the Huancavelica mine was leased to a local refiner's guild.
5. Labor in the Peruvian region was subsidized too as *mitayos* – a form of coerced indigenous labor – were indeed tributaries to the King, who assigned them to miners for regular work while guaranteeing effectively their collective property rights to land (Bakewell 1984; Tandeter 1993; Barragán 2017); *mitayos* were only a part of – unskilled – labor in Peruvian mining, whereas in Mexico however free wage labor was prevalent. Further, in a "developmentalist" state fashion (Grafe and Irigoin 2012, p. 5), taxes on mining were halved in Peru in 1736 and made equivalent to the 10% rate levied on output in Mexico. The sale price of quicksilver was further reduced in 1767 and again in 1777.
6. A greater agency of the state after the 1730s – paradoxically – reduced intermediation costs increasing the returns of miners and contributed to another leap in the silver production in the second half of the eighteenth century. Indeed, silver mining in Spanish America was largely a mix of private and commercial and nonmarket principles, which steadily expanded over the territory and grew in output over the centuries.

Private too was the bulk of the silver exported over the Pacific via Manila and to Europe via Seville – or Cadiz after 1700 – throughout. Consistently more than 80% of the silver transported over the Atlantic in the treasure fleet – convoy of a dozen ships a year to and from Veracruz and Portobello – was private. The proportion persisted when shipping was deregulated after the 1740s; 80 ships from Spain plied a year by the 1770s and around a hundred in the next decade (Hamilton 1929; Morineau 1986). When war in Europe cut off the colonies in the 1790s, US vessels were licensed to call in Spanish American ports expanding the carrying capacity and trade. Throughout though, interlopers like the Dutch in the 1660s, occasional allies like Louis XIV's France, or royal-foreign private partnerships like the *Asientos* with the South Sea Company or Genoese consortia for the introduction of slaves, all of them took a fair share of the Spanish American imports and silver exports. Hence systematically some silver escaped the Spanish tax collector and cannot be accounted for. Well-informed contemporaries at the end of the eighteenth century

estimated that precious metals lost to "contraband" represented a 17–20% of the total volume ever mined (Humboldt 1811/1808).

The Global Scope of Spanish American Silver

Private remittances to Europe meant that most of the silver arrived in Europe as return to trade not as result of fiscal or royal exaction as often assumed. Between 1500 and 1800, global trade grew at an aggregate 1% per annum and the production of Spanish American silver at 1.09% a year. In the eighteenth century, European intercontinental trade grew at 1.26%, while Mexican (recorded) silver production did it at 1.35%, much faster than the European tonnage to Asia – at 1.10% (Maddison 2007; O'Rourke and Williamson 2002, Tables 1, p. 421). It is a rate which replicates the growth rate of European population – inclusive of Russia but higher than that of Asian population growth – 0.4% for India between 1600 and 1800 (Habib 1995, p. 368) and China's at 0.08% (De Vries 2009, p. 730). However, the relation between the production and commerce of silver and the development of the global economy is not well understood yet because of the mild distinction scholars make between silver as commodity (bullion) and specie. Nor is the size or the role of the Spanish American consumption resulting from the purchasing power of silver ever assessed.

This massive inflow of American silver is well represented by the ratio to gold in Europe which stood at 10.75 in 1500 and increased to 15.61 in the 1800s on average (Laughlin 1896, Appendix II Table A). Global monetary historians have emphasized the arbitrage that Europeans enjoyed in trading silver in Asia as a driver of the growth of international commerce in the early modern period. It is unclear however why the silver trade continued and even increased in spite of China's gold-silver ratio converged with the European ones after the 1750s (Flynn and Giráldez 2002) or fluctuated well below and above the European ratio as in India (Habib 1982, Table 9). This insinuates some other intervening factors independently of the two metals' supply.

Large silver inflows during the sixteenth century persuaded economic historians of a "price revolution," which was later explained in Europe as a result of demographic changes, higher productivity in agriculture, and greater urbanization. Silver as reexport to Asia – and to the Baltic and the Levant – had to offset some aggregate effects. Similarly, a fall in the quantities arriving in Spain in the mid−/later seventeenth century has been (wrongly) associated with a global crisis (Flynn 1982; Von Glahn 1996), although direct inflows to other European ports more than compensated for the fall. This was the case of the estimated 250 million livres *tournois* (roughly 1278 t of fine silver), brought to France by St Malo's ships between 1698 and 1724 (Lespagnol 1992/1997). In fact, silver production and exports soared during the eighteenth century: both doubled between 1710 and 1770 and increased by a further 50% between 1780 and 1810. At the same time, China imports reached record levels. However, no significant aggregate effect has been identified in the contemporary development of Europe or Asia. In fact, as Morineau (1996, p. 266) pointed out 20 years ago, "the incorporation of precious metals in the general

circulating medium of European (*and Asian*) economies, their effective role in the development of the economy and of the armed forces, the concurrence with other means of payment, and the transition to the different modern monetary regimes of the 19th century" still remain to be explained.

To account the silver absorption by Europe as a result of the colonial trade of Spain – and Spain's balance of trade with other European nations – is misleading. Silver flows out of Spanish America not only served to pay for imports of European and Asian manufactures in Spain, where mainly French and Genoese houses controlled the trade; they helped also to cancel bills of exchange drawn on Cadiz and later in Barcelona or Corunna by English merchants for their smuggling from Jamaica or Trinidad. Within Europe, bills of exchange dominated the intra-European financial flows as they were a substitute for specie (Quinn 1996); this explains why these instruments were unseen in the commerce that Europeans maintained in China or Spanish America where specie had no substitute (Bernal 1992; Cheong 1978, p. 27), unless business was conducted within financial schemes such of the East India Company. Bills denominated in foreign currencies offered an opportunity for arbitrage between specie and bills in different European markets too if exchange rates were favorable. This did not necessarily have to follow the movements in the stock of silver and gold – as a crude bimetallism would conceive.

Silver Specie

Yet, the most remarkable aspect of the New World silver is that most of what was mined since the 1560s was minted and thus exported overwhelmingly in the form of coins, more precisely in the form of one peso coins. Unlike Brazilian gold, which was partly minted in Lisbon, silver in Spanish America was minted locally under the control of private individuals. Spain, like other European sovereigns, cut their metallic coins from New World precious metals (Challis 1975). In Castile, private and royal mints in Madrid and Segovia coexisted until the eighteenth century. All of them operated under royal specifications for weight and fineness of gold and silver established in 1497 that fixed the ratio for gold-silver exchange and did not change until 1730. War with the United Provinces and other interventions in European conflicts in the seventeenth century prompted large debasements of copper coinage in Spain in 1634 and 1656 that channelled bullion to England. As the main port of arrival of silver, Seville was the largest manufacturer of coins in Spain. Estimates of metropolitan silver coinage calculate that a third of the New World output was minted in Castilian mints before 1640 (Motomura 1997, Table 1; TePaske and Brown 2010, Table 3.20). This represented two thirds of the silver received in Spain in the sixteenth century. Coinage as proportion of imports decreased in the seventeenth century, but silver mintage increased in France and England (Motomura 1997, Table 2, p. 339). The proportion decreased further from a third in the period 1586–1621 to a fifth by 1700 and to barely a 6% of the total silver received in the eighteenth century when around 900 million pesos total was coined in Seville (De Paula 2016, p. 366). The remainder arriving from Spanish America was coined

already – and of high-quality standard. Unlike Britain and her empire after 1821, Spain lacked uniformity in her coinage and suffered from a very volatile monetary environment until well into the nineteenth century (Sardá Dexeus 1948). As a result of the jurisdictional fragmentation of the Hispanic Monarchy, different kingdoms in the peninsula performed with different fractional coins and monies of account, which were not current everywhere or were overvalued in another city/regions.

In Spanish America, state or royal control over coinage was minimal, indirect, and asserted only after 1730. A fraudulent adulteration of silver coinage in Potosi sometime between the 1630s and 1640s triggered major reforms. The free reign of mint officials in Potosi had reduced the content of fine silver in the coin producing a debasement from 15% to 50% between 1620 and 1650. Coinage nearly doubled there, while the registered output stagnated, and even decreased, in the same decades (TePaske and Brown 2010, Table 5.12). The reaction of the crown was swift. The intervention ended with a rare public execution of the responsible officials in Potosi and some large silver merchants of the city (Lane 2015). In both cases, in Spain and Potosi, a re-coinage sought to withdraw the debased specie from circulation: in Spain it achieved some price stabilization after copper inflation; in Spanish America it proved impossible given the chronic shortage of (any) specie in circulation and the limited increase in the pace of coinage. There was no copper coinage as small change and subsidiary moneys developed from pre-Hispanic forms as cacao beans or pieces of cloth and jaggery from sugarcane. Local vested interests in charge of the mints, coupled with the poor fiscal capacity of the colonial state, were constraints to a more efficient management of coinage in the Spanish New World.

Thus, mints in America operated with great deal of autonomy. Precious metals were coined in a few cities: Early in 1535 mints opened in Mexico City and Hispaniola (Santo Domingo), which lasted as long as gold in the island; another one operated in Lima between 1568 and 1588, which was moved to Potosi in 1575. A third mint reopened in Lima in 1684 after the Potosi fraud. Gold coins were cut in Bogota since 1621 and occasionally minted in Lima and Mexico alongside silver, though in much lower proportion. Refining and minting facilities were located at great distance – or independently – of the mining sites. The distance to the Mexico City mint from Zacatecas was 600 kms, from Durango nearly 1000, and 1400 kms from Chihuahua mines. Oruro mines were 300 kms away from Potosi, and Huancavelica more than 500 kms from Lima; Chilean silver and gold had to travel 1326 nautical miles to be coined in Lima. This gave ample opportunity to liquid merchants to purchase the metal from miners and refiners with cash and quicksilver advances. Thus private agents also controlled the coinage business under a distant supervision of Royal Treasury officials or operated the mint house right away as a private venture like in the 1740s in Chile.

Silver was assayed and stamped with the royal mark; this certified the fineness of the bar (bullion) and the payment of taxes. Up to 1728, braceage represented 3 reals out of the 67 struck from a mark of 230 g of fine silver. An additional real for mark was charged as seigniorage. However, chief offices at the mint like the treasurer, assayer, the chief smelter, and the engraver were sold; jobs like marking of the bars, keeping the weights and scales, and minor posts like scribes, guards, and even

porters were purchased also with nearly no requisite of qualification for the post. Like Treasury officials, they were subject to occasional inspection by other local bodies and by special royal envoys in extraordinary circumstances. Office holders received an annual salary plus a compensation proportional to the number of silver marks assayed and minted. Returns to office were very large as to attract wealthy individuals or religious corporations empire-wide, who subsequently leased it to moneyed locals who eventually subcontracted the job to those able to carry on the task.

Royal decrees from 1497 conceived the real as the unit of denomination and a content of 930.5 thousand of fine and a weight of 3.4 g (3.195 g of fine silver). They also established multiples of two, four, and eight reals, the peso, and fractional coins of one-half reals; however throughout the whole period, between 85% and 95% of the coinage in America was struck in the form of 8 real value, hence the piece of eight. These silver lumps (cobs) cut of bars with pliers weighted 27.46 g; at 930.5 thousands standard, they contained 25.61 g of fine silver and performed as specie. Coins bore the initial of the mint of origin and of the assayer, so the notorious Potosi coin was easily discriminated. It traded at discount against the Mexican or the Seville specie in Surat already in 1647 (Foster 1618/1669) and was distinguished in the East Indies Co ledger books. Main markets for silver at the time in Europe like Genoa, Milan, Paris, Flanders, and Konigsberg prohibited its circulation. Even Portuguese Brazil counterstamped the specie lowering its value. The re-coinage that followed changed the design of the peso replacing the Pillars of Hercules for the Jerusalem cross stamped on the coin but kept specifications for weight and fineness. The pillars were the supporters of the coats of arms of King Charles I and the words *Plus Ultra*, and the other engraved with the cross, lions, and castles represented the kingdoms of Castile and Leon, the name of the king stamped together with the motto *Hispaniarum Rex* – for coins struck in the metropolis – *Hispaniarum Rex and et Indiarum Rex* in the American coins. Thereafter the *ryal* started to be known as dollar in English, also "pillar" dollar/peso or *columnaria*.

More significant changes took place in the eighteenth century. Aiming at greater monetary order, there was an increasing intervention of the state in matters of coinage and monetary policy both in Spain and the American possessions. For one the alienation of mint offices from private individuals in Seville and the New World placed coinage of gold and silver under the jurisdiction of the sovereign. In the metropolis a council, *Junta de Moneda,* was established in the 1730s with the purpose of centralizing decisions on monetary issues, but the autonomy of individual mints did not disappear completely. The centralization was limited to the engraving and manufacturing of the dies with which to stamp the coins. In addition, coinage of silver for the metropolis was separated from the one in America. The taille of the silver mark was reduced from 67 to 68 reals in America and to 84 in Spain, reducing the fineness of the latter – the *peso provincial* – to 826.4 thousand which implied devaluation vis-à-vis the Spanish American coin, *peso nacional*. The decoupling did not solve the problems of large outflows of hard currency from Spain paying for imports from all over Europe, so the premium on the Spanish American coin continued.

An overhauling of the minting technology, some institutional reforms, and further changes in the specifications of the coin took production and mintage of silver to record levels. Until the 1730s coins were cut from hammered blanks of a consistent proportion of silver and alloy – hence the "hammered dollar" as known in the USA. Uniformity in the size, weight, and fineness developed over the sixteenth and seventeenth centuries, when the growing volume of coinage granted consistency to the coin. A new technology of minting imported to America after 1730 used horse or mule-driven laminating (rolling) mills and screw presses; thus coinage went through a revolution. The blanks were prepared using roller mills that produced uniformly thick strips of silver from which to cut the blanks with metal punches. Two heavy iron screws pressed the metal to the desired thickness, producing a round or circular coin with a rimmed or milled edge, thereafter known as the "milled dollars." The circular coins were more uniform and the rimmed edge preserved the integrity of the specie. This prejudiced clipping or shaving the piece in an environment plagued with problems of "small change" and allowed greater consistency in the size and appearance of the coin.

Stamping 68 reals per silver mark meant a slight reduction in the weight and fineness of the peso to 916.66 and to 27.064 g total weight, making it for 24.81 g of fine silver. Between 1730 and 1772, Mexico mint house only produced more than 461 million of these pieces – around 288 t of fine silver a year. A further slight devaluation took place in 1772 but the Royal Treasury absorbed the cost of coinage. Charles III debased the peso to 24.43 g of fine silver, i.e., fineness was reduced to 902.7 thousands, and again to 24.25 g in 1786 on the basis of 895.8 thousands of fine. The size and total weight did not change, but the face of the King stamped replaced the pillars on the obverse. They were hence known as carolus and "old heads," "Buddha heads," or *sikong-yin* in China. Over a billion of such coins were cut in Spanish America between 1772 and 1818.

To a higher productivity of mint houses contributed the direct control and funding the purchase of refined silver bars by the state from 1730, making coinage another monopoly like the supply of quicksilver and gunpowder. Royal treasuries earmarked revenues from (lower) mining taxes, and seigniorage built the capital for the mints. A more direct state management improved the returns of miners, who formerly had to exchange their bars for coins at heavily discounted prices from concessionaires who afforded liquidity. The *rescate*, i.e., the smelting, weighting, and assaying of silver bars, was made available to other mining districts beyond the cities with mints. New and larger facilities for the existing mint houses were built, and additional mints opened in Guatemala in 1733 and in gold-rich Popayan (Colombia) in 1758; and the state took over the Santiago Mint in 1772. In 1732 royal appointees replaced concessionaires in Mexico. Royal management was imposed to Lima mints in the 1750s and Potosi in 1771. Mechanization together with institutional reforms propitiated that a greater share of the output passed through royal mints increasing coinage. Results are visible by 1770s: coinage leapt doubling or more the levels produced before; it peaked in the 1790s and 1800s when volumes replicate trends of mining and exports (Fig. 1).

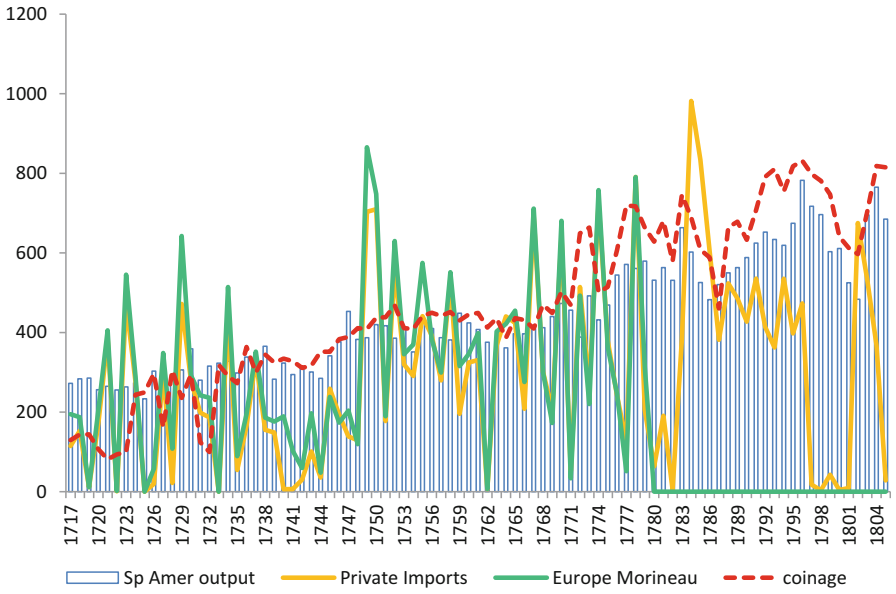


Fig. 1 Spanish American silver production, exports, and coinage, 1717–1805 (in tons). (Sources: Spanish American output (TePAske-Brown, 2010), private imports to Spain (Garcia Baquero Gonzalez 1996; Cuenca Esteban 2008); arrivals in Europe (Morineau 1986), coinage (Cespedes del Castillo 1996)

Mint houses continued in operation until the first decade of the nineteenth century, when the French occupation of Spain decisively impacted in America. The control over refining of silver bullion and coinage broke down together with the Spanish governance; mintage plummeted and exports volume abated. Within 10 years in the 1820s, Potosi and Mexican coinage fell to a third of the previous decade. Whereas additional mints opened in Spain, Cadiz, Valencia, Barcelona, and Galicia, during the war against the French occupation, in America new outlets for coinage sprang up in any region relatively well endowed with silver. Six new houses opened in the 1810s in Mexico only, in Zacatecas, Chihuahua, Durango, Guadalajara, Guanajuato, and Sombrerete, which started manufacturing their own peso coin, and the number increased further in the following decades. The fragmentation of the coinage extended to all other silver-rich regions mirroring the territorial fragmentation of the Spanish empire. Within 5 years of independence, by 1825, Peru had two additional mints in Cuzco and Arequipa; another one opened in the vicinity of Bogota and Popayan in Pasto. Republican governments in Bolivia ultimately maintained control over Potosi but started a steady debasement of the coin after 1826. Within few years the minting multiplied, and cash-starved republics coined silver of any quality and weight or, worse, started issuing unbacked paper money fuelling a beggar-thy-neighbor process in monetary affairs among formerly integrated economies. Thus the features of the Spanish American peso standard disappeared which henceforth rarefied global markets for silver (Irigoin 2009b).

World Silver: Bullion or Specie? Supply or Demand?

By 1500 the monetary stock of Europeans was roughly estimated in 3500 t of gold and 37,500 t of silver (Parker 1974, II, p. 527). New World precious metals doubled or more the stock of silver and increased that of gold by a half. By 1810 this volume represented more than 3400 billion of coins. Obviously neither the geographical distribution nor the progression of coinage was even throughout; half of the 86,000 t of silver were minted in the previous 80 years. It peaked in the later eighteenth century adding about 28 million pieces of the same specie a year. This represented a yearly 700 t of fine silver, 70% of which was produced by Mexico only (TePaske and Brown 2010, p. 113).

Even if there were no consequential effects in the sixteenth century, the acceleration of silvers flows in the world economy in the eighteenth century – and its sudden stop in the 1820s invites some scrutiny. A recent estimate puts the contribution of precious metals to the growth rates in Europe GDP for the period 1530–1790 at 1.3% Palma (2015). Silver "advanced monetization" in both Europe and China; it allowed more efficient functioning of productions and markets (Pomeranz 2000) and eased ongoing Smithian growth, although the relative impact in the productivity growth of each term of the comparison is open to question (De Vries 2015). However, scholars discerning between the "strength" of European demand and the inelasticity of Asian supply leave aside the role of American demand – and supply of specie – in the early modern global economy. It is not to argue for monetary factors in the Great Divergence, but issues explaining the "strength" of European demand and the inelasticity of Asian supply often argued do not fully consider the role of silver and the New World. Trade historians do not consider precious metals because of their monetary role and do not consider relevant their "impact of intercontinental silver flows on aggregate price levels" (O'Rourke and Williamson, 2002 fn 4). Yet, silver had a higher purchasing power in Asia where apparently there was a persistent demand – as it was "a public necessity" (de Vries 2015, p 24) – thus price (and wage) differentials persisted. Given the structure of the Spanish American silver trade, there was another arbitrage to be made by Europeans trading there, where specie abundance traded European and Asian consumer goods – and capital goods like slaves – for silver coins at the lowest international relative price. This "currency trade" expanded the demand potential of the Europeans in Asia, on top of the windfall from the "ghost acreage." As noted by Adam Smith, "the general advantage which Europe, considered as one great country" derived from the discovery of America and of the Passage to the East Indies "consisted first, in the increase of its enjoyments, and secondly, in the augmentation of its industry." The relative importance of Asia demand of silver in relation to Spanish American production and coinage is displayed in the figure below (Fig. 2).

In spite of the largesse of its silver production and exports, Spanish America did not record comparable inflationary effects in prices (Garner 1985, 1995). Contrarily, it was "the gradual enlargement of the market for the produce of silver mines in America, (was) probably the cause . . . which has not only kept up the value of silver in the European market, but has perhaps raised somewhat higher that it was about the

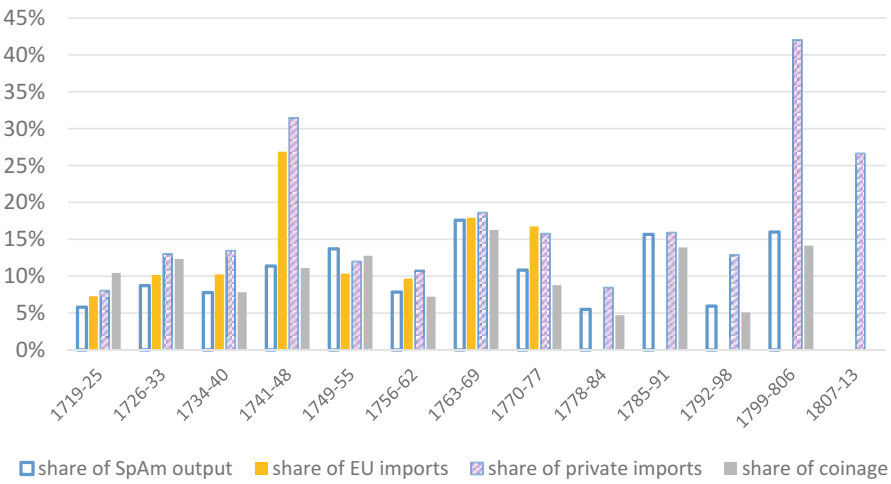


Fig. 2 China's silver imports as share of Spanish American output, coinage, and trade. (Source: China imports from Dermigny (1964), *La Chine et L'Occident*, II p. 735; Output, EU imports, Private imports and coinage idem Fig. 1)

middle of the (18th) century." On top of population recovery, the market had become more extensive in America – as in Asia – and as noted by Adam Smith, the "greater part of Europe has much improved, England, Holland, France and Germany, even Sweden and Denmark and Russia have all advanced considerably both in agriculture and in manufactures" (Smith 1776/2007, Book I, ch XI, III, p. 161, also Book IV ch VII, III, p. 457). Silver coins permitted the extension – and specialization within – of the global market although economic historians tend to disregard the monetary aspects of the Eurasian trade with American silver (Chaudhuri 1975). Emphasizing the colonial trade of each European nation, their role of carrying trade and the demand potential in the New World are often overlooked. After all, "the re-exportation (of East India goods) to other countries brought back more gold and silver to that which carried on the trade than the prime cost of the whole amounted to" (Smith 1776/2007 Book IV, ch III, I, p. 364).

To great extent, this is an outcome of looking at the Great Divergence solely through the archives of the chartered companies. Over these centuries foreign silver accounted (roughly) for 90–93% of the European cargo values to China and 75–80% of the India-bound cargo (Dermigny 1964, II, p. 686). By the mid-seventeenth century – without silver endowments – Mughal India obtained silver mostly from overland routes through Central Asia and the Levant. China sourced it from Japan (Tashiro 1991, p. 77) where gold-silver ratios were 1:11/12 vis-à-vis China's at 1:5 and 1:8 by the late sixteenth and early seventeenth centuries. The availability of Spanish American silver from European intermediaries altered these flows within Asia. It filled in for the reduction in Japanese silver going to China following restrictions to gold and silver exports that historians considered a move to protectionism (*Sakoku*) by Tokugawa from the 1630s. In 1695 Shogun Tsunayoshi directed

a debasement of the standard of the first unified Japanese currency (Keicho) established by 1601–1606 – 80% fine minted silver down to 64% (the Genroku ingot) (Hellyer 2009, pp. 52–68; Innes 1980, II, p. 582). With further debasement into the 1700s and growing silver flows through Manila, Chinese merchants withdrew from the coast of Japan – though Dutch and Korean merchants who intermediated larger intra-Asian trade remained (Tashiro 1991, pp. 78–79).

The arbitrage between (relatively) silver-abundant Europe and silver-scarce Asia engaged the Portuguese and Spaniards in Southeast Asia first, and successively European companies joined in throughout the eighteenth century. Measured by the silver amounts taken to Asia over the Indian Ocean, the VOC was leader until the 1720s; the French prevailed in the 1740s and the English were dominant after the Seven Years War. With the Napoleonic Wars, Dutch and French merchants almost disappeared from this commerce; and the "supremacy" of the English company was soon challenged by the very fast-growing free trade of the USA already in the 1790s when US vessels, sailing around Cape Horn, joined the East Indies trade. After the American Revolution, the silver that England obtained in the West Indies since the Free Ports Act (1767) was diverted partly to the former British colonies. The USA became close second to the British in China and in India in the 1800s when intermediation of Spanish colonial trade as neutrals expanded their access to silver. Thus Western silver trade and commerce overall with Asia prolonged beyond the life span of the monopolistic companies.

Commerce around the Cape of Good Hope was not the single – and possibly not the most important – source of silver. A continuous inflow arrived from the Pacific on the Spanish Galleon – another private-royal partnership of sorts for commerce in the East Indies – over 250 years. The low state capacity of the Spanish Monarchy bore no control upon – nor derived sizable revenues from – that trade either, so amounts are less known. However, evidence from bills of lading attests of a volume for Asian imports and a value for silver arriving in Manila comparable – and often larger – than the yearly "investments" of the East India Company (Morse 1919; Bonalian 2010). Estimated at roughly 50–75 t of silver a year, this flow was directed to China mainly, but thanks to the extensive intra-Asian carry trade, silver permeated to British India, Java, Surat, Laos, and even Japan and Korea. Southeast Asia continued deriving Spanish American silver beyond the Galleon well into the nineteenth century (Kobayashi 2017, 2018) (Fig. 3).

Most of the silver arriving in India, coined or otherwise, was melted and re-coined in the various Mughal and the East India Company mints alike after 1757. It was not the case in China, however; she lacked mints to coin silver in any uniform standard. Taels of the highest purity but very diverse weight and size were manufactured privately. Alongside a very inconsistent coinage of copper, often debased, and private paper notes, the availability and consistency of the Spanish American peso satisfied the demand for a certified means of payment that Qin China lacked. Distinctly, Asian imports of coins – or rather European exports of specie – increased throughout the eighteenth century. In the 1720s the Dutch exported for a maximum value of 63 t, 80% of which were made of pesos and rupees (Van der Wee 2012, p. 95). France's exports between 1725 and the 1780s had a 72% share of silver, of

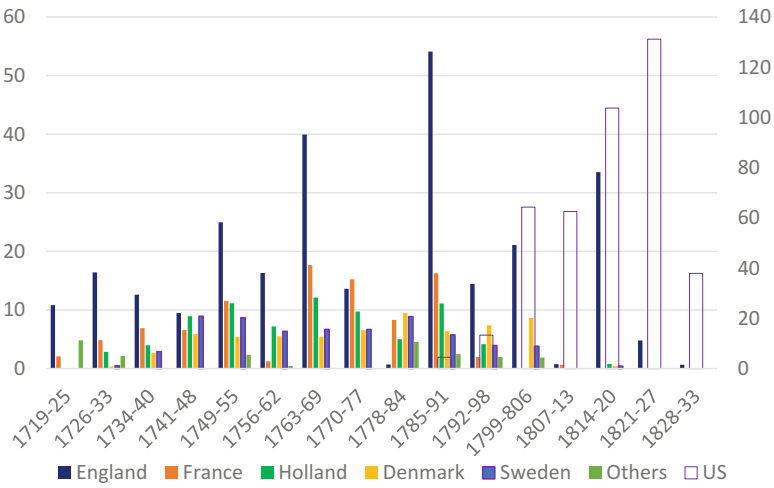


Fig. 3 China silver imports, composition by origin, in tons (US right axis). (Source: own elaboration on Dermigny (1964), *La Chine et L'Occident*, II, p. 735)

which 78% were *piastres* (Dermigny 1960, pp. 122, 124, 138). Up to 1719, silver made 80% of British exports, half of which were Spanish American reals (Chaudhuri 1968, Table 1); it was 75% of the trade with China in 1710–1759 (Morse 1922, p. 228). Between 1788 and 1809, the English company and private traders together exported an average 70 t of silver a year, 92% of which were foreign – overwhelmingly Spanish American – coins (PP 1810 Bullion Report 1810). Similarly, foreign coins made about 65% of the US total exports to China up to 1825 (Irigoin 2009a) and amounted to 1.7 million of specie exported to India between 1802 and 1808 – two thirds of the total British exports in the same period (Seybert 1818, p. 56) (Fig. 4).

Silver had no substitute in the European-Asia trade until well into the nineteenth century. It was the single means to fund the "return investments" of (any) the Company's ship. Pesos were "the token" by which "the pound sterling and the woollen cloths of England were converted into the taels of silver and the silks and teas of China; and (..) whether the dollar delivered to the Company's treasury at Canton had cost to lay down four shillings, or five, or six, *might depend the degree of profit* on the round voyage" (Morse 1922, p. 228 emphasis added). The company systematically invoiced the Spanish dollar at 5 s, which worked as "anchor" for exchange rates in their financial business in the late eighteenth century. After the 1770s the Company started drawing pesos on London – even in Guangzhou occasionally – at the Company's rate at 365- and even 720-day sight in times of plenty, whereas sterling-denominated bills were used in private remittances through the company. Yet, the market rate of exchange of the peso in London (and Asia) varied below and above the Company's rate for bills. In so doing the Company was capturing any possible arbitrage in exchange rates in London and Asia from a less than open trade.

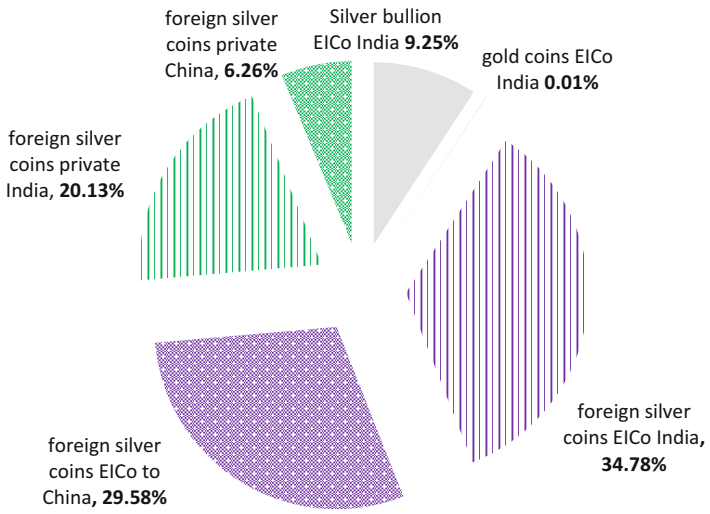


Fig. 4 Composition of Britain's precious metal exports to Asia, 1788–1809. (Source: Parliamentary Papers, HC, The Bullion Report (1810) (PP 1810))

The End of the Global Silver Standard

Seen from Asia, China's divergence is explained by the stress on factor endowments that prejudiced the continuation of Smithian growth by 1800. This is not accurately reflected in trends of silver imports if the US exports are included. In fact, the drop in China's silver commerce was dramatic only after the mid-1820s. Thus, some scholars explain the sharp fall as an exogenous supply-side shock attributed to mining disruption in Spanish America from the independence wars (Lin Man-huong 2006). Yet the lowest ever recorded production in the 1810s was about 400 t – a quantity several times larger than Chinese imports at the same time (60–140 t) (Irigoin 2009a, b). There was plenty of silver in America – even before the mining boom in Nevada – to continue the trade. This suggests a demand side aspect to the development of global silver trade that technological or institutional factors and endowments in Spanish America cannot sufficiently explain. Neither the "precipitous" fall can reflect fully a "languishing" production of exports inside China – in Jiangnan – and the "falling demand for a particular type of money (specie) which lowered the price of silver" (Von Glahn 2012). The fact that silver imports by China resumed in 1857 when there was another reputable standard of a silver coin, the Mexican peso, suggests that even if there were demand side forces behind the shift in China's export performance, this may not be necessarily endogenous. Arguably the disappearance of the silver standard of the Spanish peso in the world economy should have impacted the otherwise growing export economy inside China. This coevolution of the trends

of Spanish American silver minting and exports, with the trend of Chinese imports, and European – and US – reexports, is intriguing.

Only after 1791 was the USA able to establish some sort of (federal) sound money. Along with the establishment of the mint – and other fiscal and financial devices – Hamilton conceived the US dollar as a silver-backed currency, technically a bimetallic system, based on the value of the Spanish American peso. Although with a slightly lower silver content, the US dollar was established at par with the peso, and the new US Mint revalued the price of silver above the English ratio of 15:1 versus 15.6:1 in Britain, hence remaining in a silver standard (Michener and Wright 2006). Added to the extended carry trade in Spanish American commerce – and returns from invisibles – monetary policy channelled specie to the USA, making it the source of silver coins to China and furthering US foreign trade. The 1792 Mint Act fixed the exchange rate with other foreign coins in circulation. The Spanish peso was also the anchor to establish a system for the conversion of monies and debt of the former colonies into a federal money and remained as the single other legal tender in the USA until 1856 (Irigoin 2009a) (Fig. 5).

In spite of the silver inflow, between 1795 and 1806 when President Jefferson suspended minting the federal government coined about $ 1.2 million silver dollars of large denomination – barely 25 cents per capita. The US coin was mostly exported to the Caribbean but was rejected in China. Allegedly, 80% of the specie in circulation in the USA before 1830 was "composed entirely of Spanish coins" (Martin 1968, p. 431), and they made a large proportion of the specie reserves of the Bank of the United States even in 1831 (Gordon Hayes 1933, p. 678). Demand kept them in circulation as they "were *more valuable as money* than as commodity" (Martin 1968, p. 430) as in early eighteenth-century India; hence again the demand for a particular type of specie increased the value of a coin above its intrinsic content.

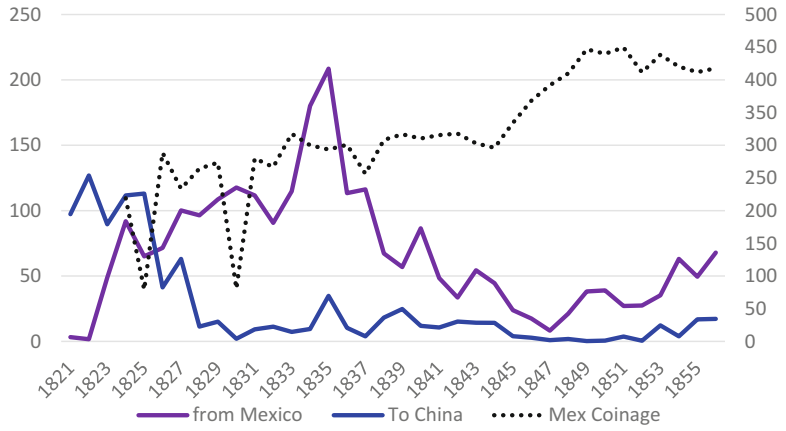


Fig. 5 US silver trade (left). Silver coinage in Mexico (right), in tons. (Source Irigoin 2009a)

The fact that China ceased to import silver in the 1820s, whereas the USA continued importing it from Mexico and other Spanish American republics is revealing. On the one hand, this confirms there was no supply-side shock to China after 1808, nor to the (re) exporter. On the other hand, as each economy had very different institutional settings to deal with silver and specie (and its vagaries), the outcome when the standard disappeared was different. The USA had a mint which assayed and set the exchange rate and parities between gold and silver specie, while a banking system with reserves in silver specie developed credit money. China lacked both. Hence China could not deal with the growing diversity in the quality of the republican silver coins originated after independence from Spain (Irigoin 2009a), with consequences for the real economy. For the very same reason, in turn, India under the control of the Company that minted silver suffered much less from the end of the silver standard.

Monetary Capacity

Indeed, early modern European, American, and Asian economies had very different institutional setups to deal with precious metals and foreign exchange. Even within Europe, different states had also very different monetary capacities despite a notional bimetallic system. Whereas Amsterdam, like Hamburg and Genoa or Venice, had banks and allowed the free deposit and free export of silver, they mostly performed with monies of account and fixed exchange rates that were adjusted over time. For example, the Amsterdam Wisselbank devalued the parity of the current Dutch *gulden* to a content of 10.751 grams of pure silver in 1606–1620, to 10.106 in 1621–1659, 9.717 in 1659–1681, and 9.557 grams of fine from 1681 onward (North 1991, p. 194). They minted different silver coins for different trades: the *rijksdaalder* for the Baltic trade, the *leeuwendaalder* for the Levant, *ducatoon* for the East India, and *patagon* or *kruisdaalder* for domestic use which the banks in Hamburg and Amsterdam tariffed independently of the intrinsic value (Kleeberg 1995, p. 88). On the other hand, sixteenth- and seventeenth-century England had a relatively open domestic market for precious metals in all forms and a large production of silver plate by goldsmiths and silversmiths, until the 1740s when the Bank of England took over the business and goldsmiths faded away (Mayhew 2012; Clapham 1941). But export of English silver was effectively prohibited and coinage nearly disappeared in the eighteenth century, yet plate and specie exports to Asia augmented over the century. Foreign coins of gold and silver – increasingly Spanish dollars – made the Bank of England bullion reserves (Clapham 1958) and the East India Co exports. Interestingly, the Chinese were well acquainted with the English goldsmith mark; thus "old plate (was) the most profitable silver you can carry with you, *when dollars are dear*" (in China) (Lockyer 1711, pp. 136, 140 emphasis added). This should explain the failure of

the Elizabeth's portcullis crown coined for the East India Company in 1601 and that of the Emden trade *thaler* coined by *Preussisch-Asiatische Compagnie* in the 1750s which tried – unsuccessfully – to substitute Spanish American specie in Asia. Still in the mid-nineteenth century, Britain tried to mint a silver coin for the trade in Asia, and the USA eventually did it between 1873 and 1883.

With the 1696's Great Recoinage, the English Mint established the definitive standard for gold mintage in 1717, but the specifications for silver coinage did not change. It also established a mint price for gold and silver which implied a fix ratio for both metals. Because the mint price of silver tended to be systematically below its market price, the system tended to overvalue and attract gold. But also the foreign silver coin, counterintuitively, was priced at a lower rate than foreign silver bullion in spite of transport and coinage costs (Conduitt 1730/1935), "set(ting) a floor or collar price for the sterling price of bullion and a collateral value of full foreign coins" (Hotson and Mills 2015, p. 218). This particular arrangement allowed foreign specie to remain in circulation and to be sent to the East Indies where it had higher purchasing power. It also provided means for the Bank to intervene in the money markets by deals with bullion, specie, and foreign coins and exchange. The Bank derived a margin from differences between the collar or floor price and the mint (and market) price of bullion and foreign specie. It did not however forward the bullion and foreign specie to the mint for coinage but retained it for further business also providing the means to, somehow, sterilize the money market through its foreign exchange interventions. The independence between stocks and flows of silver in England is apparent in the lack of correlation with the exchange rate of the Spanish American silver peso during the eighteenth century shown below (Fig. 6).

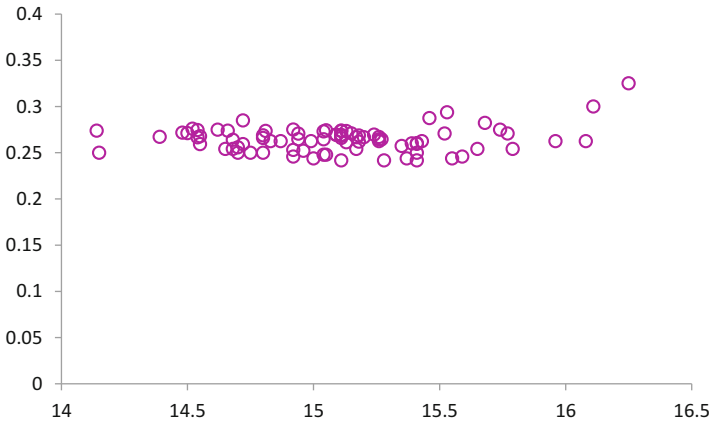


Fig. 6 Gold-silver ratio and Spanish American peso exchange rate, Britain, 1717–1818. (Source: own calculation from gold-silver ratio from Officer (1983) and exchange rate (Parliamentary Papers 1718–1736, 1746–1811 (1812–1813), 1811–1819))

In the early modern "geography of money," England was a full nation-state scale single market, a rare "island" in an "ocean" of individual city markets. England was exceptional as were her institutions to deal with bullion, specie, and foreign exchange. Elsewhere money markets were more local, confined to the commerce and finances of one city. Some countries had mints; others had banks; but none had the ensemble of centralizing institutions as England – as later the USA did. In the 1780s France, for instance, had 17 working *Hotels de Monnaie* of different capacities to coin gold and silver (Dermigny 1955). In Asia, for example, gold coins were more "popular" in East Japan, whereas silver was preferred in the West (Tashiro 1991, p. 77); India comprised different currency zones and several mints: gold coins were current in the South, while the North, including the Mughal Empire, performed with silver monies – and copper served as subsidiary to either. Even until 1835 the Company had separated coinage and currency system in each presidency, but from the later eighteenth century, silver coinage in Calcutta and Bombay was far more important that in Madras, which was better connected to silver-rich Manila. As mentioned, China had provincial and local production of copper coins and private mintage of silver bullion of diverse quality but large size and weight. Thus comparisons on the impact of flows of silver and gold in Eurasian economies – on top of distortions from very dissimilar size and population – may mean little. Within this variety of situations and monies, the Spanish American peso offered a stable, certifiable, and most abundant fractional means of payment to economies orientated to commercial crops, manufacturing, international commerce, services, and capital flows at global scale. It would be inadequate to consider the role of the silver specie in this period with the model of currency substitution in a neat bimetallic standard. The peso was a very suitable complement to large and small denomination means of payment worldwide. In turn the Bank of England reduced the denomination of its notes throughout the century from £ 50 (1696), £20 (1745), £ 10 (1759), and £ 5 (1793) to £ 1 and £ 2 after 1797.

By 1800 it was apparent that what China demanded was not silver bullion per se, but specie, a universally reliable means of payment made of silver: the coin minted in Spanish America since 1772 (Irigoin 2009a, 2013; Von Glahn 2012). Without its own coinage of silver, China had become reliant on the currency standard that the Spanish American coin provided; when its standard ceased, the de facto "dollarization" of the Chinese economy made it too vulnerable to the resulting disorder in the international system of payments. The turmoil is traceable in the almost simultaneous trends of depreciation of copper cash and the appreciation of silver taels in China. But the workings of a pure bimetallic regime in China are less convincing when observing a *rise* in copper inflation alongside with *rising* imports of silver – as what happened in the 1780–1795 period or during the 1810s and 1820s (Irigoin 2013). Prices in copper skyrocketed after the 1830s to the mid-1850s, together with silver deflation; an anomaly which is often explained by political turmoil from domestic rebellions and foreign aggression.

Wars in eighteenth-century Europe – the Seven Years and the Napoleonic Wars in particular – had major global implications. They particularly impacted Spanish America which ended with her independence from Spain in 1825. They also brought about shifts in their colonies commercial insertion in the world and the destruction of the virtues of the coin minted there. In fine, it meant the cessation of the silver peso standard. During the "restriction period," the "paper price" of the Spanish dollar was a 25–30% higher than its estimated "gold price" (Silberling 1919, Table 6). The Mint and Bank of England took the extraordinary step of countermarking pesos – instead of re-coining the silver – launching some £ 5,000,000 pounds extra into circulation between 1797 and 1811. The pesos counterstamped with a mini George III face inside an octagon added to numerous private issues by colliers and textile millers in Glasgow, Lancashire, and Derbyshire (for unknown quantities) who needed lower denomination coins for wages and could no longer pay in kind with rising costs of labor and dearth of domestic silver specie in an inflationary context. In 1797 the Bank tariffed the coins at 4s9d from the 4s6d that was the mint price; two upward revaluations (or a 22% devaluation of the pound in silver terms) to 5 s in 1804 and 5s6d in 1811 contributed to alleviate the burden on foreign exchange from war expenses and subsidies to allies (Kelly 1976). Britain emerged victorious from the war and the inconvertibility of the pound by 1815–1816 (O'Brien and Palma 2018). It took a further 5 years to redeem the "five shillings" coins and to replenish the bank reserves and find a rate of exchange at which return to convertibility. Further, in 1819 silver was demonetized in England and gold became the single standard of value, formally, in 1821.

Conclusion

Monetary historians informed in modern monetary economics find difficult to conceive a currency standard contributing to an international system of payment prior to the Gold Standard, without central banks but currency traders instead. Global economic historians entrenched in disentangling the divergence within Eurasian growth paths look only at supply-side elasticities of these economies; yet the growth of the trade with the most important international commodity, silver, and the resulting demand potential of Spanish American can reveal important unattended aspects allowing a more comprehensive understanding of the development of the early modern global economy.

This chapter has surveyed the production, manufacturing, and commercial aspects of the silver mined in Spanish America. It has also emphasized the private nature of the production and of the trade with silver and the particular character of silver specie as the main product with which the New World inserted in the international economy at the time. Silver endowments provided a windfall to Europeans – colonizers and intermediaries. Spanish America's wealth in precious metals is legendary, and so is the poor understanding of her role in the development

of the early modern global economy. The fact that Spanish America provided Europeans with the most desirable item of trade with wealthier and more advanced Asian economies has eclipsed the fact that the commodity produced there was indeed money, an almost universally accepted means of payment that her silver coin provided. Because of the sheer volume produced, coupled with the high consistency in the shape, size, weight, and fineness, the South American coin became de facto the most current international means of payment before the gold-backed British pound.

Significant developments in Europe and Asia, and individual economies within, are related in one way or another to the exchange of silver coins in wider and remote international markets. A steady increase in the world supply of silver dollars contributed to the growth of the "dollarized" economies – as the surplus in the goods trade balance (with the West) that Early Qin China enjoyed until the 1820s–1830s. Within a metallic standard with free flows of capital the trilemma does not conceive room for monetary policy. The downside was the vulnerability to external shocks of those economies performing with "imported" currency, as described above. Overall the global currency trade ought to have effects on the trade and output in the importing and intermediary economies along the way. Thus, the growth of the silver specie produced in Spanish America contributed to the contemporary Smithian growth in the global economy.

The New World did not escape from these effects. As Smith himself pointed out, (coined) silver was the means by which the extremes of Eurasia became in contact and traded with each other; it was also the driver of the European commerce over the Atlantic (and the Pacific) with the New World which offered – in turn – a potent demand for goods that a very poor labor-to-land ratio economy could not provide. America was food self-sufficient but labor scarce; tradable goods were relatively dear in comparison to other economies best endowed with labor. Thus, imports of African slaves were both a mean to exploit land where precious metals were lacking and to obtain silver coins where they were abundant. Reliant on European intermediation for consumer goods and capital inputs and technology for mining, abundance provided the lowest international relative price of silver to manufactures which prejudiced any indigenous industrial development of note. A generalized Dutch disease effect prevailed over huge transport costs imposed by distance and slow-changing navigation technology. Colonialism – and lack of a merchant fleet – furthered Spanish American foreign trade dependence on the many different commercial agents who linked producers and consumers globally, imposing higher intermediation costs than otherwise.

These were the drivers of another arbitrage Europeans particularly enjoyed, namely, the provision of consumer and capital goods to the New World (Irigoin forthcoming). This, coupled with the arbitrage originated in supplying silver money to China and other South and East Asian economies, allowed another

sort of "extra-profits" that Europeans obtained in their reexport trade within and beyond. This double de facto monopoly lowered the costs of increasing elasticity of demand of Europeans and eased their needs for liquidity on the margin. Crucial to it was the very particular institutions that one and another economy had to deal with money and foreign exchange – domestically and in the broader international economy. Comparatively, these institutions better reflect the political economy, and success, of European mercantilism. Instead, Asian large empires practically had not a monetary sovereignty to account for or a large enough banking system to effectively create endogenous money in the form of credit. They "outsourced" its "central banking" functions to Europeans who provided a reliable high-quality coin – of a convenient denomination. That specie originated in the largely private coinage in Spanish America and was extracted overseas and circulated worldwide by ways of trade. Thus it supplied a certifiable means of payment to the widest world economy that lowered transactions costs of the ongoing global Smithian growth. The counterfactual would have been a reduced demand for Spanish American silver and a much lower trade for Europeans to intermediate.

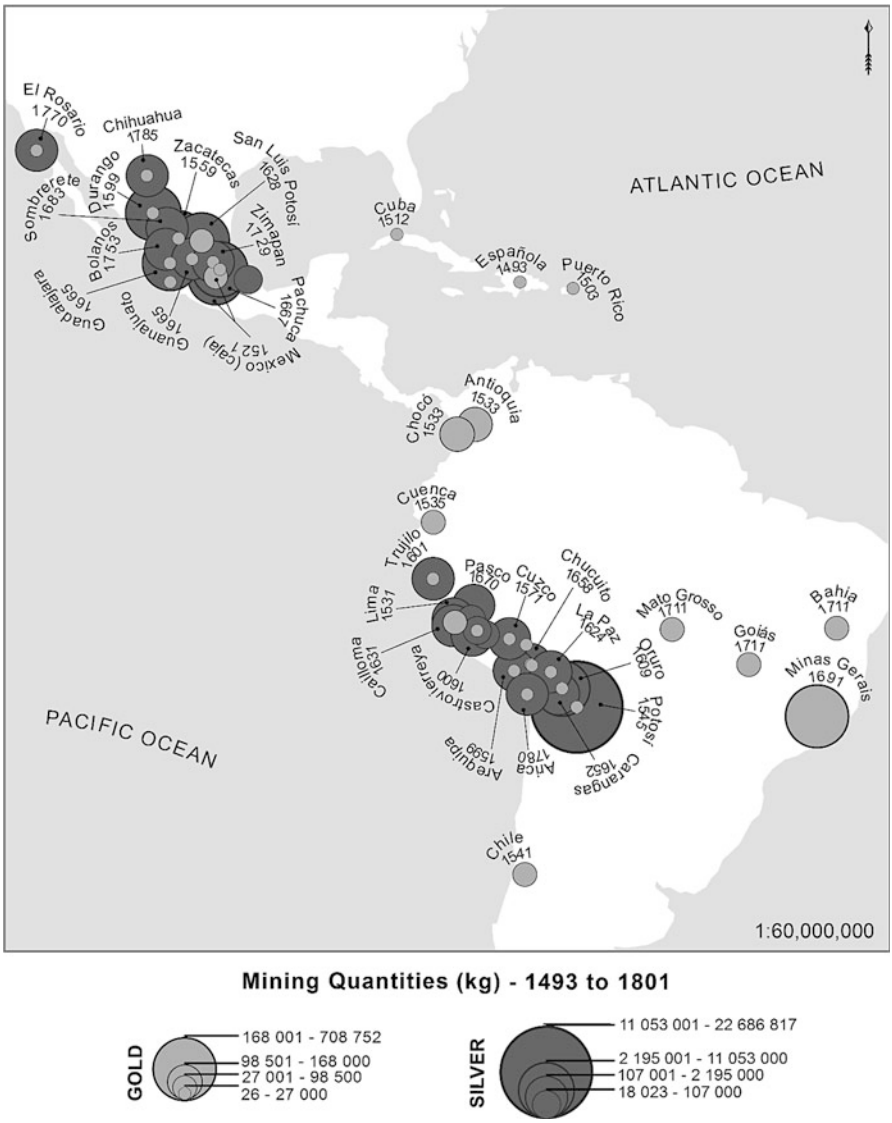
Within Europe, different urban marketplaces organized differently around barely comparable institutions to deal with precious metals, bills, and foreign exchange while maintaining an ever large import trade with Asia exchanging silver specie for textiles and other consumer goods for further reexports. A particular monetary capacity based on an extraordinary set of monetary institutions and policies to deal with the flows of – in and out – specie, which bear little relation with political institutions (Karaman et al. 2018), made eighteenth-century Britain outstanding in that game. Generalized warfare in Europe and overseas by the end of that century put the structure of the trade to test: on the one hand, the consistency of the silver standard ceased with the colonial arrangement for mining and minting silver in Spanish America with definitive implications for producers and final consumers. On the other hand, Britain managed to wage the war and won it on the seas; together with the Navy, the mint, the treasury, and the bank aligned with the privy council – the Parliament aside – managed to overcome the threat to her currency; thus after Waterloo the Gold Standard emerged alongside with England as the new leading industrial exporter of the world economy.

Cross-References

► Monetary System of the "Ancient Régime" (Third to Eighteenth Centuries)
► Money Markets and Exchange Rates in Preindustrial Europe

Acknowledgments The research presented in this chapter has been kindly funded by the Leverhulme Research Fellowship RF-2017-570\3.

Appendix



Source: N. Palma (2015) Harbingers of Modernity: Monetary Injections and European Economic Growth, 1492–1790, Unpublished Ph.D. thesis, London School of Economics, Fig. 2, p. 63. I thank Nuno Palma for allowing the reproduction of the updated version of his map

References

Bakewell P (1984) Miners of the Red Mountain. University of New Mexico Press, Albuquerque
Barragán R (2017) Working silver for the world: mining labour and popular economy in Colonial Potosi. Hisp Am Hist Rev 97(2):193–222
Barret W (1990) World bullion flows, 1450–1990. In: Tracy J (ed) The rise of merchant empires: long-distance trade in the early modern world, 1350–1750. Cambridge University Press, Cambridge, pp 224–254
Bernal A (1992) La financiacion de la Carrera de Indias (1492–1824) Dinero y credito en el comercio colonial espanol con America. Fundacion El Monte, Sevilla
Bonalian M (2010) El Pacifico Hispanoamericano Political Comercial y trafico asiatico entre las Filipinas, Mexico y Peru 1680–1784. Historia. PhD, El Colegio de Mexico, Mexico
Cespedes del Castillo G (1996) La Produccion. In: Las Cecas Indianas en 1536–1825. Fabrica Manual de Moneda y Timbre, Madrid, pp 247–271
Chaudhuri K (1968) Treasure and trade balance: the East India Company's export trade, 1660–1720. Econ Hist Rev 21(3):480–502
Chaudhuri K (1975) The economic and monetary problem of European trade with Asia during the seventeenth and eighteenth centuries. J Eur Econ Hist 4(2):323
Cheong W (1978) Mandarins and merchants: Jardine & Mathesson & Co a China Agency of the early nineteenth century Curzon 1978. Curzon, London
Clapham J (1941) The private business of the Bank of England, 1744–1800. Econ Hist Rev 11(1):77–89
Clapham J (1958) The Bank of England, a history. Cambridge University Press, Cambridge
Coatsworth J (2006) Political economy and economic organization. In: Bulmer Thomas V, Coatsworth J, Cortes Conde R (eds) The Cambridge economic history of Latin America, vol 1. Cambridge University Press, Cambridge, pp 237–274
Conduitt J (1730/1935) Observations upon the present state of our gold and silver coins, 1730, from an original manuscript formerly in the possession of the late Dr Jonathan Swift. In: Shaw W (ed) Select tracts and documents illustrative of English monetary history, 1626–1730. T. Becket, London, pp 210–244
Crosby A (1972) The Columbian exchange: biological and cultural consequences of 1492. Greenwood Publishing, Westport
Cuenca Esteban J (2008) Statistics of Spain's colonial trade, 1747–1820: new estimates and comparisons with Great Britain. Rev Hist Econ xxvi(3):323–354
De Paula Perez, F. (2016). La Casa de Moneda de Sevilla, su historia. Sevilla, Universidad de Sevilla
De Vries J (2009) The limits of globalization in the early modern world. Econ Hist Rev 63(3):710
De Vries J (2015) Understanding Eurasian trade in the era of the trading companies. In: Berg M et al (eds) *Goods from the East, 1600–1800: trading Eurasia*. Palgrave Macmillan, Basingstoke, pp 7–39
Dermigny L (1955) La France à la fin de l'ancien régime: une carte monétaire. Ann Hist Sci Soc 10(4):480–493
Dermigny L (1960) Cargaisons indinennes: Solier et Cie, 1781–1793. SEVPEN, Paris
Dermigny L (1964) La Chine et l'Occident; le commerce à Canton au XVIIIe siècle, 1719–1833. SEVPEN, Paris
Flandreau M et al (2009) Monetary geography before the industrial revolution. Camb J Reg Econ Soc 2:149–171
Flynn DO (1982) Fiscal crisis and the decline of Spain (Castile). J Econ Hist 42:139–148
Flynn D, Giráldez A (2002) Cycles of silver: global economic unity through the mid-eighteenth century. J World Hist 13(2):391–427

Garcia Baquero Gonzalez A (1996) Las remesas de metales preciosos americanos en el siglo XVIII: Una aritmetica controvertida. Hispania 56:203
Garner R (1985) Price trends in eighteenth-century Mexico. Hisp Am Hist Rev 65(2):279–325
Garner R (1995) En torno al debate sobre la inflacion en Mexico durante el siglo xviii. In: Riquer J, Grosso J, Yuste C (eds) Circuitos mercantiles y mercados en Latino America, siglos xviii y xix. Inst JLM Mora UNAM, Mexico, pp 161–178
Gordon Hayes H (1933) Bimetallism before and after 1834. Am Econ Rev 23(4):677–679
Grafe R, Irigoin A (2012) A stakeholder empire: the political economy of Spanish imperial rule in America. Econ Hist Rev 65(2):609–651
Habib I (1982) Monetary system and prices. In: Raychaudhuri T, Habib I (eds) The Cambridge economic history of India. Cambridge University Press, Cambridge
Habib I (1995) "The Eighteenth Century In Indian Economic History." Proceedings of the Indian History Congress 56: 358–78. http://www.jstor.org/stable/44158638.
Hamilton EJ (1929) Imports of gold and silver into Spain, 1503–1660. Q J Econ 43(3):436–472
Hellyer R (2009) Defining engagement. Japan and Global Contexts, 1640–1868. Harvard University Press, Cambridge, MA
Hersh J, Voth H (2009) Sweet diversity: colonial goods and the rise of European living standards after 1492. Discussion paper CEPR. Centre for Economic Policy Research, London
Hotson A, Mills T (2015) London's market for bullion and specie in the 18th century: the roles of the London Mint and the Bank of England in the stabilization of prices. In: Allen M, Coffman D (eds) Money, prices and wages; Essays in honour of Prof. Nicholas Mayhew. Palgrave, Basingstoke, pp 217–227
House of Commons Parliamentary Papers (1718–1736, 1746–1811 (1812–13), 1811–19) An account of the market prices of standard gold in bars, foreign (ie Portugal) gold in coin, standard silver in bars and Spanish dollars or pillar pieces of eighth with the courses of exchange with Hamburg, Lisbon and Paris, extracted from Castaign's and Lloyd's lists first quotation every month. Parliamentary Papers. London House of Commons
House of Commons Parliamentary Papers (1810) Report together with minutes with evidence, and accounts from the Select Committee on the High Price of Bullion House of Commons, London
Humboldt A (1811/1808) Political essay on the kingdom of New Spain. I. Riley, New York
Innes R (1980) The door ajar: Japan's foreign trade in the seventeenth century. History. PhD, University of Michigan, Ann Arbor
Irigoin M (2009a) The end of the silver era: the consequences of the breakdown of the Spanish silver peso standard in China and the US, 1780s–1850s. J World Hist 20(2):207–243
Irigoin M (2009b) Gresham on horseback. The monetary roots of Spanish America political fragmentation in the nineteenth century. Econ Hist Rev 62(3):551–575
Irigoin A (2013) A trojan horse in Daoguang China? Explaining the flows of silver in and out of China. LSE economic history working papers 173/13. London School of Economics, Department of Economic History, London
Irigoin A (forthcoming) The new world and the global silver economy. In: Riello G, Roy T (eds) Global economic history. Bloomsbury, London, pp 271–286
Karaman K et al (2018) Money and monetary stability in Europe, 1300–1914. CEPR, London
Kelly E (1976) Spanish dollar and silver tokens, an account of the issues of the Bank of England, 1797–1816. Spink & Son Ltd., London
Kleeberg J (1995) America's silver dollars. American Numismatic Society, New York
Kobayashi A (2017) Price fluctuations and growth patterns in Singapore's trade, 1831–19. Aust Econ Hist Rev 57(1):108–129
Kobayashi A (2018) International bimetallism and silver absorption in Singapore, 1840–73. Econ Hist Rev (forthcoming) 1–23
Lane K (2015) Corrupción y Dominación Colonial: El Gran Fraude a la Casa de la Moneda de Potosí en 1649. Bol Inst Hist Argent Am Dr Emilio Ravignani 3ra Serie(43):94–130
Laughlin J (1896) The history of bimetallism in the United States. Appleton, New York

Lespagnol A (1992/1997) Messieurs de Saint-Malo: une elite negociante au temps de Louis XIV. Presses Universitaires de Rennes, Rennes
Lin M. h (2006) China upside down : currency, society, and ideologies, 1808–1856. Cambridge, Ma, Harvard University Asia Center
Lockyer C (1711) An account of the trade in India. Samuel Crouch, London
Maddison A (2007) Contours of the world economy, 1–2030 AD: essays in macro-economic history. Oxford University Press, Oxford
Martin D (1968) Bimetallism in the United States before 1850. J Polit Econ 76(3):428–442
Mayhew N (2012) Silver in England 1600–1800: coinage outputs and bullion exports from the records of the London Tower Mint and the London Company of Goldsmiths. In: Munro J (ed) Money in pre-industrial world: bullion, debasements and coin substitutes. Pickering and Chatto, London, pp 97–108
Michener R, Wright R (2006) Development of the US monetary union. Finan Hist Rev 13(1):19–41
Morineau M (1986) Incroyables gazettes et fabuleux métaux: Les retours des trésors américains d'après les gazettes hollandaises, 16e-18e siècles. Les Editions de la MSH, Paris
Morineau M (1996) Rejoindre to A. Garcia Baquero "Las Remesas de metales preciosos Americanos en el siglo XVIII: una aritmetica controvertida". Hispania 56:263–266
Morse H (1919) British trade in China. J R Soc Arts 68(3497–8):17–28, 34–44
Morse H (1922) The provision of funds for the East India Company's trade at Canton during the eighteenth century. J R Asiat Soc 54(2):227–255
Motomura A (1997) New data on minting, seigniorage, and the money supply in Spain (Castile), 1597–1643. Explor Econ Hist 34:331–367
Munro J (2012) https://www.economics.utoronto.ca/munro5/ResearchData.html. Silver mining 2017
North M (1991) Bullion transfer from Western Europe to the Baltic and Asia, 1550–1750: a comparison. Money, coins and commerce: essays in the monetary history of Asia and Europe. E. Van Cauwenberghe. Leuven, Leuven University Press: 185–196.
O'Brien P, Palma N (2018) Danger to the old lady of Threadneedle street? the bank restriction act and the regime shift to paper money, 1797–1821. Eur Rev Econ History. Economics Discussion Papers- Univ of Manchester 1808(June)
O'Rourke K, Williamson J (2002) After Columbus: explaining Europe's overseas trade boom, 1500–1800. J Econ Hist 62(2):417–456
Officer L (1983) Dollar-Sterling mint parity and exchange rates, 1791–1834. J Econ Hist 43(3): 579–616
Palma N (2015) Harbingers of modernity: monetary injections and European economic growth, 1492–1790. Economic History. London, London School of Economics PhD.
Parker G (1974) The emergence of modern finance in Europe, 1500–1730. In: Cipolla C (ed) The Fontana economic history of Europe, vol ii, Cambridge, Cambridge University Press. pp 527–594
Pomeranz K (2000) The great divergence, China, Europe and the making of the modern world economy. Princeton University Press, Princeton
Quinn S (1996) Gold, silver and the glorious revolution. Arbitrage between bills of exchange and bullion. Econ Hist Rev 49:473–490
Sardá Dexeus J (1948) La política monetaria y las fluctuaciones de la economía española en el siglo XIX. Instituto de Economia Sancho de Moncada, Madrid
Seybert A (1818) Statistical annals embracing views of the population, commerce, navigation of the United States of America. T. Dobson & Son, Philadelphia
Silberling N (1919) British financial experience 1790–1830. Rev Econ Stat 1(10):282–297. https://www.fraserstlouisfedorg/files/docs/meltzer/silbri19.pdf. Accessed 26 June 2017
Smith A (1776/2007) An inquiry into the nature and causes of the wealth of nations S. Soares. MEtalibri, Amsterdam/Sao Paulo
Soetbeer A (1879) Edelmetall-Produktion Und WerthverhäLtniss Zwischen Gold and Silber Seit Der Entdeckung Amerika's Bis Zur Gegenwart. Perthes, Gotha

Tandeter E (1993) Coercion and market: silver mining in colonial Potosí, 1692–1826. University of New Mexico Press, Albuquerque

Tashiro K (1991) Exports of gold and silver during the early Tokugawa Era, 1600–1750. In: Van Cauwenberghe E (ed) Money, coins and commerce: essays in the monetary history of Asia and Europe. Leuven University Press, Leuven, pp 75–94

TePaske J, Brown K (2010) A new world of gold and silver. Brill, Leiden

Van der Wee H (2012) The Amsterdam Wiseelbank's innovations in the monetary sphere: the role of 'Bank money'. In: Munro J (ed) Money in pre-industrial world: bullion, debasements and coin substitutes. Pickering and Chatto, London, pp 87–95

Von Glahn R (1996) Fountain of fortune: money and monetary policy in China, 1000–1700. University of California Press, Berkeley

Von Glahn R (2012) Cycles of silver in Chinese monetary history. In: So BKL (ed) The economy of lower Yangzi Delta in late imperial China. Routledge, London, pp 17–71

Part V

Monetary Experiments

Experiments with Paper Money

16

François R. Velde

Contents

Abstract

Experiments in paper money provide the crucial link in the transition from commodity money systems to modern fiduciary currency. China's experience ended shortly before Europe began its own experiments. Using paper instead of precious metal was a way to economize on resources but also allowed governments to finance their deficits. The experiments surveyed show how issuance of paper by a central bank, coexisting with coins made of precious metal, came to be the dominant model in the nineteenth century.

Keywords

Paper money · Fiduciary currency · Inflation

F. R. Velde (✉)
Federal Reserve Bank of Chicago, Chicago, IL, USA
e-mail: fvelde@frbchi.org

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_17

Introduction

In the early sixteenth century, an Italian jurist named Girolamo Butigella contributed to a centuries-old debate about what could constitute valid tender in repayment of debts. At the time jurists only thought of money as taking the form of coined gold or silver, although reports of China's paper currency (brought by Marco Polo) were taken increasingly seriously. Butigella departed sharply from the consensus. For him, monetary objects derived their value not from their intrinsic content, gold or silver, but from public approval: "even if a coin were made of lead, indeed even of wood or leather, as long as it is publicly approved, it would be possible to repay it for another coin."

This revolutionary idea was a step too far for the eminent French jurist Charles Dumoulin, who ridiculed the notion a few years later as "irrational and ridiculous: why, by the same token, it would be possible to make money out of printed paper, and that is just as ludicrous and ridiculous as a children's game, and not only contradicts the origin and definition of money, but also experience and common sense."

Three centuries of experience later, paper money was anything but a children's game. By 1800 most European countries had had direct experience with paper money of various forms, which this chapter surveys.

But first a matter of definition. Paper money is at the intersection of two concepts of money, as a monetary object and as a security. Generally speaking, money often (though not necessarily) takes the form as physical media of exchange, which can be broadly classified into those that have intrinsic value (commodity money) and those that don't. The distinction is not clear-cut: coins made of precious metal (say, silver) usually contain some base alloy (say, copper) and as the proportion varies the coin, if it circulates, may owe less and less of its value to its metallic content. In the extreme case of pure copper, coins usually circulated for far more than their intrinsic worth and were in effect tokens. Paper money is in some respect just another form of token coinage, and indeed one of the earliest examples of "paper money" were the token coins made of recycled books in the besieged city of Leyden in 1574.

Paper money is also related to securities or more broadly liabilities. Indeed the formal similarity of the physical notes to checks (including their horizontal rectangular format) belies their common origin in medieval payment orders. Paper money is the physical evidence of a promise that used to be quite explicitly stated on the paper itself (and still is, charmingly but pointlessly, on Bank of England notes).

This chapter will use as definition of paper money a promise or liability physically represented by a piece of paper that circulates as medium of exchange. Of course this definition does not delineate sharp boundaries between what was and wasn't paper money. On the medium of exchange aspect, it can be difficult to establish clearly how "generally accepted" a paper money was. One useful indicator is the size of the smallest denomination, compared to average income or wages. In eighteenth-century Britain, the £5 note represented a month or two of wages, too large for ordinary retail transactions but less than a pure wholesale payment instrument or a tool reserved for the merchant class. On the security aspect, there are many

potential characteristics, and while paper money is now of a fairly standard form, there has been considerable variation. Some of the key characteristics are the terms, if any, under which the paper money was convertible into an intrinsically valuable object (typically gold or silver coin or bullion) and the terms under which the paper money could serve to discharge debts (i.e., was legal tender) or had to be accepted in transactions ("cours forcé").

This chapter will proceed to survey various experiments country by country. China is first and somewhat special, as there is little direct link between its experiment, which ended in the fifteenth century, and the others which began in the seventeenth century in Europe and led eventually to the late nineteenth-century gold standard with paper issued by central banks (Van Dillen 1964).

China

The use of paper money began under the Northern Song dynasty (960–1126), arising at first in the Sichuan province where bronze currency, used elsewhere, was unavailable and replaced with iron (Von Glahn 2005). The inconvenience of this medium led to the emergence of privately issued paper money, soon restricted to a government-appointed monopoly, and later taken over by the local government itself in 1024. These notes, called *jiaozi*, were in use for decades in Sichuan but sank to 10% of their value by 1107 due to overissue and were eventually replaced by an inconvertible currency called *qianyin*, which was demonetized in the early 1110s. After the Jurchen people conquered the north of China (establishing the Jin dynasty) the Song retreated south. Lack of currency prompted once again private issues of paper money called *huizi*, in 1135. Fiscal pressure led the Southern Song government to issue its own *huizi* from 1160. For several decades, the value of the currency remained stable, sustained by occasional redemption into silver, but from 1190 inflation became chronic, due to both civil wars within the Song domains and constant warfare with the Jin dynasty to the north. By 1207 the principle of convertibility had been abandoned; by 1240 the money supply had increased by a factor of 50 relative to 1190. Parallel developments took place in northern China under the Jin dynasty: lacking copper mines and also fighting against the Mongols to the North, the Jin resorted to paper money (called *jiaochao*) from the 1150s, at first intermittently convertible into silver and later inconvertible and heavily depreciated. Ultimately both dynasties succumbed to the Mongol invasion, and the Mongol ruler Kublai Khan founded the next dynasty, the Yuan, in 1271.

In Xanadu did Kublai Khan a paper currency decree: the Yuan adopted paper money from the Jin and organized it into the sole currency to the exclusion of other forms of coinage, introducing it into their new territories as their empire expanded. They issued the currency in a large range of denominations and established redemption offices throughout the Empire. Although denominated in bronze coins, the notes were in fact convertible into silver bullion. Large notes were subject to a fee upon redemption, presumably justified by their convenience in long-distance trade. Marco Polo happened to visit the Khan just about at that time, and he was duly impressed by

the brand-new and well-designed system. But it did not take long for things to go wrong due to excessive issue: within a decade convertibility was abandoned. In 1287 the currency was devalued by 80% and a new one introduced, which did not fare much better. A third currency introduced in 1307 collapsed within a few years and was withdrawn. From 1311 the two previous paper currencies circulated concurrently at stable exchange rates to silver until popular revolts in the late 1340s led to further massive issues and deep depreciation.

With such a record, one might have expected the Chinese to abandon paper money. On the contrary, the newly established Ming dynasty once again faced difficulties in supplying enough bronze coins and turned to paper money from 1375, as a supplement to coins. For a while it produced alternately metallic coins and paper money, but reckless spending eventually led to more printing, depreciation, and final abandonment in the mid-1430s. Henceforth the Chinese authorities stopped providing currency, either paper or metallic, and silver bullion became the unit of account and medium of exchange, a state of affairs that lasted until the late nineteenth century.

Sweden

Sweden's experiment started in the mid-seventeenth century (Wettenberg 2009). The early success of Amsterdam's *Wisselbank* (founded in 1609) had been spurring various projects to establish a bank in Sweden since the 1620s. The first to be realized was due to a merchant from Riga (then a Swedish city) of Dutch origin, Hans Witmacker, known as Johan Palmstruch after his ennoblement. In 1656 Palmstruch received a royal charter to set up a loan and exchange bank, the *Stockholms Banco*, in order to fulfill two goals: stimulate trade by making loans and provide a stable medium of payment by exchanging book-entry credits for coins. Although Amsterdam only presented a model for the latter function, Hamburg's *Wechselbank* (founded in 1619) also had a lending activity.

Palmstruch's enterprise was run by him and largely controlled by the executive, i.e., the king. It first started as a deposit bank but soon began lending as well as paying interest on deposits to fund itself. In 1661 Palmstruch found a new way to fund lending, namely, issuing bank notes in standard denominations, payable on demand to the bearer. The idea came to him from the receipts that the mining company in Stora Kopparberg issued to miners in exchange for copper ore and which were seen to circulate hand to hand. The notes had an enormous element of convenience because Sweden, since the 1620s, had sought to support its copper industry by making copper into a medium of exchange. Given that the metal was roughly 100 times cheaper than silver, Sweden's full-bodied coinage was extremely cumbersome, with some "coins" weighing 20 kg. From the start the notes were issued in a wide range of denominations, and the lowest was equivalent to around 35 g of silver, the equivalent of a large silver coin and worth 1 or 2 weeks' wages for a laborer. Palmstruch's bank, however, failed quickly because of overissue: it suspended payments in 1663 and was later liquidated. In its place the Estates

(Parliament) created their own bank in 1668, which is now Sweden's central bank, the *Sveriges Riksbank*.

Sweden underwent two more episodes of paper currency, both much more extensive than Palmstruch's experiment. The Bank of the Parliament was directly involved in the first one, despite its charter's prohibition on note issue. The inconveniences of the continued copper standard, continually altered to finance the wars of King Charles XII, meant that the bank's customers took the habit of using payment orders (effectively certified checks) as hand-to-hand currency. Eventually the Bank issued "transfer notes" which were made legal tender for taxes in 1726, further enhancing their popularity. Note issue was restrained at first, but after the death of Charles XII, Parliament regained control during the "Age of Liberty" (1719–1772) and used its bank to satisfy its constituencies' demands for cheap loans and to finance its aggressive foreign policy. By 1745 the volume of notes had reached such proportions that the Bank was forced to suspend convertibility. Prices rose and foreign exchange fell, but many observers of the time attributed this to balance of payments issues. Ultimately, after a change in the balance of power in Parliament, the government embarked on a policy of monetary contraction to reverse the currency's depreciation, from 1767 to 1769. The resulting economic damage contributed to the restoration of autocratic rule under Gustav III in 1772, who restored convertibility of the currency and established a silver standard in 1777.

The second episode was driven by two wars with Russia (1788–1790 and 1808–1809). In the first, Parliament refused to let the Bank finance the war but was sidestepped by the establishment of a National Debt Office which soon began to issue its own paper, which circulated alongside and soon depreciated relative to the Bank's own notes. A plan to make the Debt Office notes legal tender on the same footing as the Bank's notes led a disgruntled rentier to shoot Gustav III during a masked ball in 1792. A few years later, a plan was drawn to stabilize the value of the Debt Office notes and turn over their management to the Bank, and Sweden enjoyed a few years of stability but became entangled in the Napoleonic wars with disastrous consequences. The Bank was compelled to assist in war finance, the money supply doubled, and the Bank's notes became inconvertible. After military defeat Parliament regained permanent control and forced the king to abdicate in 1809, but it took many years to return to a metallic standard. In 1834, all notes became convertible in terms of silver at a fixed rate, starting a period of monetary stability that would continue until the twentieth century.

England

England (or more precisely, London) developed an early practice with representative media of exchange in the seventeenth century (Clapham 1945). During the Commonwealth large numbers of merchants remedied the lack of small change by issuing copper tokens redeemable on demand. After the Restoration in 1660, London goldsmiths received deposits of cash and helped their clients manage them. This entailed both the use of early checks (i.e., orders by the customer to the goldsmith to

pay a third party) and the issue of promissory notes (promise by the goldsmith to pay a named party or the bearer on demand) which circulated and were accepted by other goldsmiths. Thus by the time of the Bank of England's foundation, merchants and wealthy clients in London were used to a paper-based medium of exchange.

The Bank of England was founded in 1694 mainly to float a government loan. In a process of securitization that would be repeated with the East India Company and the South Sea Company, investors were offered equity in a commercial venture instead of a claim on government. The money raised was lent to the government, and this debt to the Bank formed its first asset; the other was the right (later privilege) to conduct banking operations in London. This, in effect, was a way to enhance the value of the loan to the investors: collectively they received the interest on the Bank's loan to government but also enjoyed the potential profits of the Bank's franchise. As a bank it carried out the same business as the goldsmiths (taking and managing deposits, making payments, issuing notes payable on demand) and private bankers (making loans, discounting bills, dealing in foreign exchange). This implied that, in contrast with the existing public banks of Amsterdam and Hamburg, the Bank of England would operate with a fractional reserve: only part of its assets were in the form of coins, liquid but yielding nothing.

The Bank's beginnings were inauspicious. England was at the time engaged in an expensive war that stretched the financial resources of the government which soon turned to the Bank for help. Plans to create other banks circulated, and at least one, the Land Bank, was implemented and threatened the profitability of its business plan. Within a few months, the poor state of the silver coinage determined Parliament to recoin the whole money supply, but the process, hastily conceived and chaotically executed, left the country with too little currency; a flow of notes brought in for redemption forced the Bank to suspend convertibility on May 1696, less than 2 years after its founding. The Bank's bonds sank to a 20% discount in early 1697. In addition the Exchequer started issuing its own bills, as small as £5 and bearing 4.5% interest, a direct competition to the Bank's own notes. But the war was nearly over: peace was signed in September 1697. The same year the Bank was able to negotiate an extension of its charter and a monopoly on incorporated banking in exchange for swapping old government debt for Bank stock. Later, in 1707, the Bank was able to take over the circulation of the Exchequer bills by setting up a fund to ensure their convertibility on demand, restrict their denomination to £25 and above, and prevent further issues without its consent. It also obtained a further extension of its charter and a prohibition on note issue by partnerships of more than six people, all in exchange for a loan to the government. By that point the Bank had firmly entrenched its monopoly and made itself indispensable to government finances.

Note issue was not cited explicitly in the Bank's charter, which was in fact rather vague on what kind of business it could carry out, save for a prohibition on buying and selling goods. It quickly adopted the methods of the goldsmiths, allowing depositors to transfer funds or make payments with "drawing notes" (the equivalent of checks) but also issuing "running cash notes" payable on demand to the bearer. These notes bore interest in the first years only. Fairly quickly they were issued in standard denominations up to £1000. The smallest note (aside from a few £5 notes in

the early years) was £20 until 1751, £10 until 1793, and £5 until 1797 (in 1777 an Act had prohibited the issue of notes under £5). These were large denominations at a time when a laborer's monthly wages were between £1 and 2.

The Bank came close to complete disaster in 1720, at the same time as, and because of, Law's experiment in France. Nine years earlier, in 1711, the South Sea Company had been created to securitize existing government debt: in this instance the inducement for bondholders to convert was the expected profit from trading with the Spanish colonies in South America, including along the Pacific Ocean (then known as the South Sea). The South Sea Company, close to the Tories, was a potential rival to the Whigs of the Bank of England. This rivalry broke out openly in late 1719, when the South Sea Company, following the model of John Law's debt conversion in France, offered to convert the whole British national debt (including the Bank of England's annuity) into its own stock. The government would reap a lower debt burden; the investors would enjoy whatever profits the Company could expect, not only from the trade to the South Sea (which by then held little promise) but also from the management of all government debt and the eventual replacement of the Bank of England as privileged partner of the government. The threat was grave enough to lead the Bank in a bidding war with the South Sea Company; fortunately for the Bank, the South Sea Company won in February 1720, although the Bank's annuity was left out of the conversion. The ensuing speculative fever induced government creditors in great numbers to accept overvalued stock for the debt they held. When the stock price fell, it was too late for the creditors to back out, and the Bank of England, solicited for help, demurred. Parliament refused to rescind the conversion but eventually remitted the price that the South Sea Company had promised to pay, lessening the hardship for its shareholders. South Sea shares just became government debt under another name, and the last major threat to the Bank's dominance disappeared.

For the rest of the eighteenth century, the Bank's history was uneventful, compared to its first years and to most of the other experiments described in this chapter. The occasional crises it encountered were at first political (such as the run in 1745 due to the Jacobite invasion) and later financial (such as the Europe-wide crisis of 1763 and the postwar slump of 1783).

As with its peers throughout Europe, the Bank's big test came with the French and Napoleonic Wars. When France went off the metallic standard (see below), gold flowed into England; when the *assignat* experiment ended in 1795–1796, flows reversed, and in February 1797 the Bank was obliged to suspend convertibility for 6 months. Six months turned into 24 years, as the government found it convenient to use the Bank for short-term funding as long as the wars continued, which they did, with only a short interruption from 1801 to 1803, until Napoleon's defeat at Waterloo in 1815. The Bank started issuing small denominations (£1) almost immediately. From 1797 to 1815, note issue tripled. Also adding to the note supply were the country banks, located outside the perimeter of the Bank's note-issue privilege. Prices eventually rose, as did the premium on gold, reaching nearly 40% in 1813. Great Britain's debt was also vastly expanding at the same time, and the government debt held by the Bank was never more than a few percent of the total;

but, if the Bank did not massively finance deficits, it helped the government with short-term debt. Nevertheless the Bank's responsibility in the depreciation was debated in Parliament before the Bullion Committee: the Bank's representatives thought that the adverse trade balance was to blame, while others such as Thornton and Ricardo traced the depreciation to the issue of inconvertible notes (which, moreover, became effectively legal tender at par in 1811).

The end of the wars in 1815 brought a reduction in public spending but also a repeal of taxes, and it took longer than expected for the Bank to be in a position to resume payments, in part also because the government wished to reform the monetary system and formally put in place a gold standard before resuming convertibility. Resumption occurred in May 1821, small denomination notes were withdrawn, and the United Kingdom remained on the gold standard with paper convertible on demand until 1914.

Outside of England but within the British dominions, it should be noted that paper money came of use early in Scotland, where the Bank of Scotland issued notes from 1695 and notes as small as £1 from 1704, followed from 1727 by the Royal Bank of Scotland. In the British colonies of North America note issue by colonial governments began in 1695 in Massachusetts, although French Canada had begun issuing emergency money on playing cards in 1685 and continued to use various forms of paper money until the British conquest in 1760. All 13 colonies followed suit, with considerable variation in the management of the currencies. Typically they were not convertible on demand but were legal tender (sometimes at some future date) in payment of taxes. During the American Revolution, the Continental Congress also resorted to paper money to finance the war against Britain. This is covered in more detail in ▶ Chap. 17, "Money and Prices in Colonial America".

France

The first and brief experiment with paper money started in 1701. The French government was at the time repeatedly taxing the money stock by demonetizing existing coins if they were not surrendered to the mint and overstruck with a new design. The coins with a new design had higher face value, but part of that gain was paid to the mint (in other words, the mint returned fewer overstruck coins than it received). This method had been used in 1689 and in 1693, during wartime. During the 1701 operation, the mint delivered receipts to coin holders instead of returning overstruck coins. The *billets de monnaie* or mint bills were preprinted and the actual sum handwritten. The bills were withdrawn by 1703, but another restriking of the coinage began in 1704, and new bills were issued, this time in large quantities and rarely in exchange for actual coins but as paper money. They were issued in standard sizes ranging from 200 L (around £12) to 1000 L, in principle payable within a month, but reimbursement was repeatedly postponed and an interest of 7.5% allowed. In 1706 they were made legal tender for up to a quarter of all payments, and the quantity issued had reached 180 million L (ten times the bank of England's note circulation at the time). The discount on the bills reached 60% in 1710; they were

eventually withdrawn by conversion into other forms of debt and by accepting them in payment of the seigniorage tax on yet another recoinage in 1709.

The failure of the *billets de monnaie* informed the beginning of the second experiment (Murphy 1997). When the Scots projector John Law arrived in Paris and proposed a plan to create a bank, the authorities were ultimately reluctant to let him implement his initial project, which called for a state-run bank handling all tax-related financial flows. A bank would be useful but had to be distinct from the state whose credit was ruined and could not be a substitute for the harsh measures that were required to straighten the government's accounts. The government allowed Law to create a private bank (May 1716), whose initial capital, like the South Sea Company, consisted in heavily discounted government bills, while a combination of spending cuts, tax increases, and selective default on war contractors brought the budget roughly into balance. Law's bank resembled the Bank of England in form, but it was not grounded in the local financial community nor did it enjoy the same broad political support. Law's power derived largely from the support of the duke of Orléans, regent of the kingdom, and from those seduced by his theories or enriched by his operations. The *Banque générale* that he founded prospered in spite of the financiers' opposition. The notes it issued were denominated in silver coins of specific weight and fineness and thus partially protected from the monetary manipulations that had plagued the kingdom since 1689. The government also acted to enhance their attractiveness. They were made legal tender for taxes and redeemable on demand by all tax collectors, which made them useful for remittances within the kingdom. And when a recoinage took place in 1718, the notes were made redeemable into new coinage at a preferential rate. The *Banque* also proved adept at providing deposit services and dealing in foreign exchange and was credited with lowering short-term interest rates in Paris. Law seemed to achieve in a couple years what it the Bank of England had in a couple decades.

Law, however, soon deviated from the existing English model. In December 1718 the bank was nationalized and became the *Banque Royale*. It issued new notes denominated in units of account, the livre, rather than in specie, which soon became legal tender for private debts. Most importantly Law, who remained director of the bank, ceased to be accountable to shareholders. Concurrently Law had created in 1717 a trading company, again on the English model, except that its main trading asset, the colony of Louisiana, was more substantial than a right to trade with a hostile country, the English South Sea Company's only asset. In addition Law's company quickly expanded by a series of mergers and acquisitions to become known as the *Compagnie des Indes* and encompass the tobacco monopoly, the slave trade, the fur trade with Canada, and all trade with the Far East. It then acquired the right to collect nearly all taxes in France as well as the lease on the royal mints. It finally offered in August 1719 to refinance the national debt, raising the funds to do so by issuing stock – in effect converting government bonds into the stock of a company that monopolized foreign trade and collected all taxes. For the conversion operation to succeed, however, the share price had to stay high, and Law propped it up by offering to buy shares with bank notes. In doing so, he lost control over the money supply even as he was demonetizing gold and silver coins and issuing notes

as low as 10 livres. Foreign exchange rates began to fall and prices to rise. In May 1720 Law tried to stop the depreciation by cutting the face value of the notes in half, but this only succeeded in provoking a bank run. Convertibility, even into small silver coins, was suspended, and the notes began to depreciate. Within 6 months Law's attempts at rescuing his construction had failed, the notes were demonetized, the Compagnie was in effect bankrupt, and Law had fled France. Eventually the notes were converted into long-term government debt, at varying rates.

Law's experiment, the first full-scale implementation of a fiduciary currency in Europe, stands out in many respects, by the scale it reached as well as by the swiftness of its demise. Contrary to many other episodes, however, it was not brought down by fiscal pressures, but rather by its creator's mistakes. It did not stand alone but was closely linked to a scheme whereby public debt was to become equity in a state-like entity, thereby freeing currency from the need to absorb fiscal shocks. Law also stands out because, as an economic thinker of some acumen, he conceived of paper currency as a superior medium of exchange, and therefore his experiment was intentional (and, from our point of view, forward-looking).

The third French experiment began in 1776. France had a new king, Louis XVI, a reform-minded economist as minister of finance, Turgot, a balanced budget and a temporary respite from warfare. Isaac Panchaud, a cosmopolitan figure (born in London of a Swiss father and Dutch mother) obtained permission to create a bank which he called *Caisse d'Escompte*. Initially intended to finance overseas trade, it quickly became a bankers' bank. It could only issue demandable liabilities and engage in bullion trade, lending, and discounting, like the Bank of England. But its core capital was not made of government debt: it was financed with private capital, and, aside from its charter (which conferred neither privilege nor monopoly), it had initially little or no relation with the government. This could not last very long, however: French involvement in the War of American Independence had created large deficits which the government could only fill with increasingly expensive loans. Repeatedly it requested loans from the bank which was its only resource by 1788. Lacking the political authority to raise taxes sufficiently, the government was forced to call a meeting of the Estates General, the event that marked the beginning of the French Revolution. The new constituent assembly found a sufficient financial resource in the nationalization of church lands and decided to issue its own currency backed by the lands, the *assignat*, in 1790. This freed the Caisse from its entanglement with the government and allowed it to resume payment on its notes.

The constituent assembly thought that a free and democratic government could be trusted to manage a paper currency and had no need of a central bank. The assignat were issued in large denominations of 200 L (£8) or more, like the Caisse d'Escompte's notes, and were briefly initially interest-bearing. Their purpose was to repay the government debt: holders of the new notes could use them to buy church lands as they were auctioned off. The first 2 years of the new currency were fairly successful. But the Revolution took a turn for the worse. The confiscation of church lands and the ensuing establishment of the Catholic clergy created strong frictions. Disagreements grew between the king, who had never really accepted the constitutional limits placed on his power, and more radical groups. When France went to war

with Austria in April 1792, the budget situation quickly went out of hand. The conflict became Europe-wide the following year, and money creation became the only resource left for the revolutionary government. Notes were issued in smaller denominations, down to 5 L (£0.2) in 1791 and 0.5 L in 1792. The regime of the Terror successfully forced people to use the currency and limited the extent of depreciation of the currency. Once the Terror was ended in August 1794, the currency's value began to slide rapidly, and by March 1796 it was less than 1% of its original value. A short-lived replacement, the *mandat*, collapsed within a few months.

The Caisse d'Escompte had not survived the Revolution: it had been shut down by the revolutionary government in August 1793. But in 1796 it was essentially reconstituted, under the name of *Caisse des comptes courants*. It formed the nucleus around which a group of major bankers formed the Banque de France in February 1800, with the support of Napoleon Bonaparte's new regime. But in contrast with the other belligerents of the Napoleonic Wars, France's bank did not finance government spending, and its notes remained convertible throughout, indeed all the way to 1914 except for a few brief episodes (1848, 1870 to 1873).

Denmark

Denmark (whose king also ruled Norway until 1814) also has a long history with paper money (Andersen 2016). In 1695 a merchant named Jørgen Thormøhlen obtained a royal charter to issue bank notes in Norway, but he went quickly bankrupt. When Denmark participated in the Great Northern War against Sweden, the king issued notes, ranging from 1/6 to 100 riksdaler, that were legal tender but only for half of the payment (1713–1716). The notes went to some discount relative to silver coin but were eventually all retired at par by 1728. Not long after a private bank was chartered in Copenhagen, the *Kurantbanken*. Its notes were legal tender for taxes only. The bank operated as a classic bank, taking deposits, discounting bills, and making loans, including initially modest loans to government. Its notes were convertible on demand, but the bank was not subject to reserve requirements, and it seems to have kept low reserves, exposing it to runs. Convertibility was suspended temporarily from 1745 to 1747 because of silver outflows from Denmark and again in 1757 when the Seven Years War broke out. This suspension, however, would last for several generations. The government made the notes legal tender for all payments, issued denominations as low as 1 riksdaler (£0.2), and borrowed extensively from the bank. Peacetime efforts to restore convertibility never succeeded: the government nationalized the bank in 1773 and learned to live with, and manage, a large stock of inconvertible paper money.

The system that emerged in the Danish dominions was unusual and complex. Although the stock of Courant notes tripled during the last third of the eighteenth century, the extent of their depreciation remained limited. In 1791 the Danish-Norwegian specie bank was established to manage a dual monetary system: it issued its own notes convertible in silver on demand but also redeemed the Courant notes at

market price. The Danish king was also ruler of Schleswig and Holstein (the duchies), but these kept a separate monetary regime; in particular the Danish Courant notes were never legal tender. In 1776 a bank was created in Altona, in Holstein near the great commercial center of Hamburg. This bank held deposits and made transfers but did not issue notes; a successor state-owned bank created in 1788 continued to support foreign trade and also issued notes, but its strict rules on note issue and convertibility were always abided.

By the late 1790s, Denmark, although neutral in the wars that were ravaging Europe, could not escape their impact. In 1799 silver outflows led to the effective demise of convertibility and renewed issues of paper money through a *Depositokasse*. In 1801 Admiral Nelson destroyed the Danish fleet, and in 1807 Copenhagen was bombed and largely destroyed by the British fleet. The Danish government was forced to resort to more paper money issues, increasing the nominal stock by a factor of nearly 6 to 1813, by which time the currency had lost over 90% of its value. In 1813 a new bank, the Riksbank, was created and its notes replaced the Courant notes at a 6:1 ratio. The new institution's main asset was a lien on all real estate in Denmark. It nevertheless took nearly 30 years to resume convertibility of the notes: this occurred in 1845 in Denmark (Norway, which had been transferred to Sweden in 1814, resumed convertibility in 1842).

Austria

Austria's first public bank, the *Wiener Stadtbanco*, was created on the model of the Venetian *Banco del Giro* to turn exiting government debt into a ledger-based payment instrument (Jobst and Kernbauer 2016). Initially put under the authority of the municipality of Vienna, with dedicated tax revenues as assets, it progressively lost its autonomy and, under Empress Maria Theresia, became largely controlled by the government. In 1762 it was ordered to issue notes (*Bancozettel*) from 5 florins (about £1.7) to 100 florins, which were not legal tender between private parties but were accepted for up to half of tax payments. The notes were convertible on demand into silver or could be used to buy government bonds. The issue, made in the last stages of the Seven Years War, was designed as temporary, but the notes were well received, and the issue was renewed in 1771, this time legal tender for half of all payments. Use of the notes spread further into the Austrian domains in the 1780s.

As elsewhere the French and Napoleonic wars created large fiscal burdens which were partly financed by note issue. In 1796 the notes were made legal tender in every payment, and convertibility was restricted, and then suspended, in 1797. The increasing depreciation of the paper currency led to the melting of the silver coinage and the issue of notes down to 1 florin (and up to 1000 florins). As Austria was engaged in nearly constant warfare against France (and constantly losing), paper issues exploded. The first attempt at stabilization and reform occurred after the Treaty of Schönbrunn, which deprived Austria of 17% of its territory. In 1811 the volume of notes had surpassed 1 billion florins (20 times the volume in 1796), and, after some debate, the government had to default on its debt (reducing interest payments by half) and on the currency, converting it into new notes at a 5:1 ratio.

The new currency, eventually called the Vienna standard, devalued in subsequent years as war, and paper issues, resumed in 1813, tripling the volume of the new currency in 3 years.

In Austria as elsewhere in Europe, the peace of 1815 was a turning point. In 1816 the "Privileged Austrian National Bank" was founded as a privately owned joint-stock company with a 25-year charter. The plan was to have this institution, modeled on the Bank of England, circulate its redeemable notes in parallel with the old currency which would be progressively retired. The first attempt at conversion offered holders of Vienna standard paper florins the option to convert into a mixture of silver coinage and low-yielding government bonds, but the option was too attractive and withdrawn within a few weeks. In 1819 a second, more successful plan left to the Bank the conversion policy; the Bank exchanged notes for silver at a pace of its choosing and returned the notes to the government for silver and 4% bonds. The value of Vienna standard florins stabilized quickly at the Bank's parity of 2.5:1, and the old currency was progressively retired, while the Bank's notes remained redeemable.

By the 1840s Austria might have seemed in as good a position as Denmark or Sweden, but a full return to a metallic standard was repeatedly postponed. In 1848 the wave of revolutions that swept over Europe brought revolts in the Empire's Italian dominions and secession in Hungary: a run on the Bank's soon forced a suspension of convertibility, and the government borrowed from the Bank and issued its own paper. After the emergency the quantities of paper money were stabilized, and resumption of convertibility was planned for 1859 but was postponed by the war against Piedmont and France that year. Resumption was projected for 1866, but another war, this time against Prussia and Italy, broke out that very year. The compromise between Austria and Hungary the following year created two parliaments but left a single central bank for the two halves of the monarchy, which meant that the process for deciding monetary policy was cumbersome. By 1873 the Bank came close to resuming convertibility, but the collapse of bimetallism and the fall in silver prices made it undesirable to resume on the old silver standard. Instead the Bank started targeting a stable gold value for its notes. Ultimately, the Austro-Hungarian Empire never quite returned to convertibility at a fixed parity. Instead, from 1892 the central bank actively managed the exchange rate to maintain parity with other gold-standard European currencies.

Russia

Until the seventeenth century, Russia's monetary system was limited to small silver coins called kopeks, alongside imported foreign coinage (Bloch 1899). In 1655, during the First Northern War, Tsar Alexis minted silver rubles (until then the ruble was a unit of account of 100 kopeks) and then similarly sized copper rubles intended to circulate for the same value. From 1658 overissue resulted in rapid depreciation of the copper rubles, which in 1663 traded at 50 to 1 for silver rubles. This episode of war finance with overvalued currency was repeated many times in Russian history until the 1920s.

Peter the Great in the early eighteenth century designed a European-style monetary system with gold, silver, and copper coinage for different denominations, but Russia did not yet extract gold or silver, and acquiring enough through foreign trade was a slow process. This, as well as continuous warfare, led to repeated debasements of the silver currency as well as large issues of overvalued copper under Peter I and his immediate successors. With the accession of Catherine the Great in 1762, the use of debasement and copper issues ended, but the circulating currency was still dominated by large quantities of copper whose inconvenience, as in Sweden, predisposed the public to accept a paper alternative (Heller 1983). Thus, when Catherine instituted banks in St Petersburg and Moscow in 1768 to issue paper *assignats* redeemable in silver or copper, the response was initially favorable. For the next 20 years, the assignats, issued in denominations from 25 (£5) to 100 rubles, suffered little or no depreciation and served as a medium of exchange.

In 1786 Catherine the Great merged the two banks into one Imperial Assignat Bank and created a bank to lend against real estate to the nobility and urban classes. At the same time, she raised the circulation from around 40 million rubles to a maximum of 100 million and issued smaller denominations (5 and 10 rubles). Over the next few years, Russia was at war with Sweden, Poland, and Turkey, and the self-imposed limit was soon breached. The assignats began to depreciate. After 1794 Russia fought in Poland, Persia, and the Caucasus, and the assignat circulation reached 210 million in 1799. A brief attempt at making it convertible at a 1.3:1 ratio in 1798 failed. Russia's involvement in the European wars against France, as well as incessant fighting on its southern borders with Persia and Turkey and in the Caucasus meant constant deficits; lack of central control over the bureaucracies and inefficient tax collection compounded the problems.

The peace of Tilsit in 1807 with France opened a brief window for reform. The value of the assignat kept falling, to half and by 1810 to a quarter of its face value. A bold plan, put forward by the Tsar's adviser Mikhail Speransky, reform posited that the silver ruble was the true monetary standard, renounced the use of money creation, and froze the circulation of assignats at its current value. Of course Speransky understood that this was only possible as part of a comprehensive reform of finances that would balance budgets on a permanent basis. To stabilize the value of the assignats, he planned to redeem a fraction, raising the necessary funds through sales of state lands. He also planned to found a silver-based note-issuing bank on the model of the Banque de France, partly owned by the state and partly by the private sector and engaging in short-term discounting of commercial paper rather than the long-term land-backed mortgages that the existing state bank offered. Although some steps were taken according to these plans, continuing deficits in preparation for war and the abrupt dismissal of Speransky in March 1812 prevented further progress. In April 1812 the assignat was made legal tender and unit of account for all payments. Russia would continue to use fiat currency for decades.

After Napoleon's disastrous invasion of Russia and the end of the European wars in 1815 finally brought a respite for Russian finances. From 1817 the quantity of assignats was kept constant or decreasing and their value stabilized. The same year the Bank of Commerce was founded, which eventually formed the nucleus for the

central bank. Finally, under Minister Kankrin a reform enacted from 1839 to 1843 replaced the assignats, at a 3.5:1 ratio (the current market value) with credit notes issued by the Bank of Commerce and convertible on demand in silver, backed by a reserve of at least 1/6 of their total amount.

Deficits continued however, and when the Crimean War broke out in 1853, there was no choice but to suspend convertibility in 1857. In 1860 the state bank was founded: although under the government's complete control, it was endowed with a metallic reserve to back the credit ruble, and in 1861 a crawling peg was instituted to bring the credit ruble progressively back to par, but the Polish insurrection of 1863 put an end to this experiment. The war against Turkey in 1877 brought further increases in the circulation of credit rubles, which reached its peak in 1879.

By now the rest of the world had moved to the gold standard, and silver's value on the world market was falling. As in Austria, the credit ruble reached par and even gained relatively to the silver ruble. The government accumulated gold reserves through capital inflows as well as exploitation of mines in Siberia and in 1897 made the credit notes convertible into gold, putting an end to a century of fiduciary currency.

Conclusion

Why was paper money invented? The answer is straightforward: relative to metallic money, which preceded it for nearly two millennia, paper money is more convenient to use and cheaper to produce. However, this answer immediately raises another question: why did it take so long to invent? This survey of experiments with paper money suggests that, while it is a good idea, paper money is difficult to execute.

Paper money's advantage, its low cost of production relative to its potential value in exchange, is also its weakness, leaving it exposed to the twin threats of counterfeiting and overissue. Counterfeiting is not specific to paper, and it has plagued metallic coinage since its inception; but it is more dangerous for paper money or token money in general, that is, money whose cost of production is substantially below its market value. To the extent that the intrinsic content of a gold or silver coin is ascertainable, the problem of counterfeiting is limited; coins that are made of gold but not struck by the official authority entail little or no loss to their owners, nor do they diminish the value of authentic coins. Not so for paper money.

The problem of overissue is specific to paper money. There cannot be too much gold or silver coinage, because if there were users would melt the excess coins and put the metal to other uses. This safety mechanism is unavailable for paper money: the paper is worthless in alternative uses, and some alternative mechanism is needed to ensure its value.

Paper money appeared so late in the history of money because these challenges entailed a number of technological and institutional prerequisites. Paper was necessary, and (for large-scale, standardized, and hard to counterfeit issues) a printing

technology. China had both long before the West and a centralized state that could enforce use and deter counterfeiting.

Paper and printing gave the West books and printed images in time for the Renaissance and the Reformation, but weak and fragmented states meant paper money came later, through a variety of paths. One is the concept of emergency money: a temporary, overvalued currency issued by an authority to serve as a medium of exchange under extraordinary circumstances, with an implicit or explicit promise to retire it at face value after the emergency has passed. The most common documented examples, stretching from the late fifteenth to the early nineteenth centuries, are obsidional currencies, issued during a siege by the local commander to pay his troops, and made of roughly cut metal plate of tin, copper, or sometimes silver.

Still the first experiments (Sweden, England, and France) arose through private entities (banks), albeit state-chartered. This is not surprising. The forerunner of paper money as a circulating claim is the bill of exchange, and to this day the format recalls that origin (as does the antiquated "promise to pay the bearer on demand" on the Bank of England's note). The attempts at economizing on precious metals (or, in Sweden, cumbersome copper) took the form of promises, claims issued by banking institutions. Aside from the well-known London goldsmiths, mention should be made of the Neapolitan banks, a group of nonprofit depository institutions that arose in the late sixteenth century and early on issued circulating liabilities called *fede di credito*. Although they did not take the form of a standardized bank note until the nineteenth century, they point to the source of the innovation in early modern banking practices.

But even if paper money can be said to have arisen from the private sector, government was not far and soon became closely involved. By the mid-eighteenth century, Sweden and Denmark had implemented state-run, economy-wide paper money, soon followed by Austria and Russia. As it turned out, launching a currency was easy: managing it properly was another matter. Long before the Bullion controversy in Britain, interesting debates were taking place in Sweden and Denmark over the determinants of the value of money and the proper management of a fiduciary currency. From the eighteenth-century experiments, two competing models of paper money emerged, one issued by private banks in large denominations and geared toward wholesale payments and the other issued by the state and effectively replacing coins in retail payments.

The European wars from 1792 to 1815 represented a watershed. Both models were tested, and few countries that had paper money were able to maintain convertibility. But the degree of depreciation varied considerably. It was mild in England and was reversed after convertibility resumed at par. In Austria and Russia, the loss of value, while limited in comparison to twentieth-century inflations, surpassed 90% (as it had a few years earlier during the American Revolution). Of the two models, that of the Bank of England (and the Banque de France) clearly triumphed. In the classical gold standard adopted in Britain in 1816, the unit of value was tied to a constant amount of gold, represented by large-value coins. Smaller denominations were issued in silver or bronze, as tokens rather than full-valued coinage.

Notes circulated alongside coins, convertible on demand. As with medieval emergency monies, deviations were understood to occur: a fiscal emergency could lead to temporary suspension of paper's convertibility, and the suspension could be quite long; but resumption at par was expected to follow, as it had in Britain in 1821, as it did in the United States in 1879 after the issue of "greenbacks" during the Civil War. Accordingly, in August 1914 nearly all belligerents suspended convertibility of their paper money, unwittingly opening a quite different chapter of monetary history.

Cross-References

- ▶ International Monetary Regimes: The Gold Standard
- ▶ Money and Prices in Colonial America
- ▶ The Historical Evolution of Central Banking

References

Andersen SE (2016) The origins and nature of Scandinavian Central Banking. Palgrave Macmillan

Bloch J d (1899) Les Finances de la Russie au XIXe siècle. Paul Dupont, Paris

Clapham J (1945) The Bank of England: a history. Cambridge University Press, Cambridge

Heller K (1983) Die Geld- und Kreditpolitik des russischen Reiches in der Zeit der Assignaten (1768–1839/43). Franz Steiner Verlag, Wiesbaden

Jobst C, Kernbauer H (2016) The quest for stable money: Central Banking in Austria, 1816–2016. Campus, Frankfurt/New York

Murphy AE (1997) John Law, economic theorist and policy-maker. Clarendon Press, Oxford

Van Dillen JG (ed) (1964) History of the principal public banks. Frank Cass, London

Von Glahn R (2005) The origins of paper money in China. In: Goetzmann WN, Geert Rouwenhorst K (eds) The origins of value: the financial innovations that created modern capital markets. Oxford University Press, Oxford

Wettenberg G (2009) Money and power: from Stockholms banco 1656 to Sveriges Riksbank today. Sveriges Riksbank, Stockholm

Money and Prices in Colonial America

17

Farley Grubb

Contents

Abstract

The British North America colonies used different media of exchange to transact domestic trades. The creation of efficient domestic barter structures helped maintain chronic outside-money (specie coin) scarcity. To replace this missing outside money, the colonies developed alternative inside transacting media and units of account. I assess whether these alternative media were monies using a new approach to determining what is money and what is not money. These alternative media included tobacco money, wampum money, grain or country pay as money, and, eventually, legislature-issued paper monies. I document how colonial legislatures structured their paper monies, show how the value and performance of these monies can be predicted, and report the estimated relationships between prices and paper monies. Colonial economies were under-monetized, bank-structure constrained, and undeveloped in their ability to produce import substitutes. These conditions maintained chronic outside money scarcity and

F. Grubb (✉)
Alfred Lerner College of Business and Economics,
University of Delaware and NBER, Newark, DE, USA
e-mail: grubbf@udel.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_18

induced colonial legislatures to eventually create inside paper monies. These paper monies, however, were cumbersome media of exchange compared with subsequent post-colonial banknote monies.

Keywords
Bills of credit · Paper money · Tobacco money · Wampum · Zero-coupon bonds

Introduction

Only the British North American colonies to 1775 are considered here. The focus is on how these colonists executed domestic transactions within their colonies, namely on what kinds of media of exchange and monies they used in local trade, and how these monies affected local prices. These colonies began as primitive settler economies primarily oriented to agricultural production, as well as organic land and sea product extraction, both for domestic use and export. Ancillary craft production for domestic use arose to support these primary product sectors. While the colonies developed and grew over time, both in population and in internal trade, the goods produced, the technology used, and their export orientation changed little (McCusker and Menard 1985; Nettels 1934).

For the colonists, specie (gold and silver coins) was the universal money – often called *real* money. It was outside money. It was the money the rest of the world used to consummate transactions that crossed polity borders, and it was the money often used by Europeans to consummate domestic transactions within their respective polities. The colonists did not produce specie, as gold and silver were not yet mined there; besides the British Crown did not allow them to mint coins. If they were going to use specie money to execute domestic transactions, they would have to import that money in exchange for exported goods. The colonists' exports to Britain, primarily tobacco, generated specie (pound sterling) credits that could be accessed through bills of exchange drawn on accounts in England, but it did not provide a significant inflow of specie coins to the colonies that could be used to execute domestic trades. The colonists' exports to Spanish America, however, did yield payments in gold and silver coins which, in turn, became the principal specie monies available in the colonies to execute domestic transactions (Grubb 2004; McCusker 1978; Nettels 1934).

Throughout the period, colonists complained that specie money was chronically scarce. Not enough was present to execute local trades or pay local taxes. Colonists complained that as quickly as specie was imported, it was re-exported, mostly to England, to buy imported goods. These complaints were ubiquitous and insistent (Davis 1964; Grubb 2012; Nettels 1934: 13, 91, 202–208, 281).

Without real money, local exchange was transacted through barter, but not the crude barter that requires a double coincidence of wants. Colonists did not go door to door with piglets under their arms searching for neighbors who had and were willing to trade candles for said piglets. The colonists developed relatively efficient barter structures organized around shop-notes, store book-credit accounts, tobacco credits, promissory notes, individual bonds, and so on, with local goods priced in a common

unit of account which allowed easy relative pricing and account clearing across multiple trades and traders. Early on, the common unit of account was the primary commodity produced in the colony, such as tobacco in Maryland and Virginia. Later, colonial legislatures created units of account that were imaginary colony-specific pounds set equal to (at face value) some fraction of a pound sterling (Grubb 2012; Nettels 1934: 208–228).

Efficient barter, by trading goods in a credit-debt account-based exchange structure, allowed for the time mismatch, as well as some limited geographic and individual mismatch, of demands – thus overcoming the costly limitation of crude barter exchange. Set-offs and net settlements were used to clear accounts between traders and even among a limited set of traders, without using a store-of-value instrument, or money "thing," whose value was independent of the personal reputation of the debtor party to the transaction.

Efficient barter placed a price floor below which local prices could not fall. Thus, as specie left the colony local prices could not fall enough to draw specie back in through changing the terms of trade so that specie could be used again as a local money. The opportunity cost of using specie money to execute domestic transactions was not zero. The colonists could not easily produce or generate substitutes for desired imports – imports that could only be purchased with specie money or specie credits. As such, rational maximizing colonists used specie to buy imported goods, thus sending the specie out of the colony and making it unavailable to execute further domestic transactions. They turned to efficient barter structures to execute their domestic transactions. Chronic specie scarcity was the result (Grubb 2012, 2016b).

Regarding the execution of domestic transactions within colonial economies, the media of exchange (M) equaled the money (M) in the domestic economy plus the efficient barter structures (EB) that obviate the need for a money "thing" to execute a transaction, Eq. (1). Monies are media of exchange, but not all media of exchange are monies. The monies (M) used to execute domestic trade in colonial economies were imported foreign specie coins (Ms) and, in the latter half of the colonial period, paper monies (Mp) issued by colonial legislatures – see Eq. (2). Whether local commodities monies (Mc), such as tobacco, wampum, grain, deerskins, and beaver pelts, belong in M or are part of EB is addressed below. Efficient barter structures had largely displaced specie monies before the colonies turned to issuing paper monies, Eq. (3). Thus, before paper monies were created, assuming that Mc is part of EB, $M \approx$ EB, namely barter structures (including ubiquitously traded local commodities as commodity monies) accounted for almost the totality of the media of exchange for domestic transactions in colonial economies.

$$M = \mathrm{M} + \mathrm{EB} \tag{1}$$

$$\mathrm{M} = \mathrm{Ms} + \mathrm{Mp} + \mathrm{Mc} \tag{2}$$

$$\Delta\mathrm{EB} = -\Delta\mathrm{Ms} \quad \text{up to some threshold } \psi\mathrm{Ms} \quad \text{where } 1 \geq \psi \rightarrow 0. \tag{3}$$

What Is Money? What Is Not Money?

Economists answer these questions poorly. They do not use the tools of economics to independently determine what is M and what is not M. Economists simple assume X is M and Z is not M in their models and applications, and that trades in the economy are sufficiently monetized in X so that X captures the relational patterns of behavior in the theory corresponding to M. For the nineteenth- and early twentieth-century U.S. economy, M seems obvious and easy to define and measure in a way that captures most of the *M* used in domestic transactions, namely it was coins, banknotes, and bank accounts (Friedman and Schwartz 1963; Rockoff 1971). Other than just assuming M, economists often define M in a way that ultimately everything is money – which is not very helpful, or they fall back on the eighteenth-century methodology of using categorization by unweighted characteristics to evaluate monies, such as portability, divisibility, storability, and so on, which still does not determine what is M and what is not M.

Being able to determine what is M and what is not M is important for colonial economies because they developed various inside media-of-exchange methods to execute domestic transactions that took the place of missing outside monies (specie coins). Which of these media were truly monies and which were not? To determine this we need a definition of money that allows us to distinguish between transactions using money and transactions using barter. We also need a definition that is able to distinguish between commodity monies and ubiquitously traded barter goods. Barter is when real goods or real assets directly, or ultimately, pay for real goods or real assets without using an intervening and easily transferable store-of-value instrument that is itself not a real good or real asset. Real assets include claims to specific future real goods or real value obligations. As such, we need a definition of money that separates out, or is independent of, the non-money real goods or real asset value of the object we are identifying with money.

I use the economist's tools of opportunity cost and counterfactual analysis to define money. The money "thing" or "object" is composed of real goods or real asset value and, what will be called here, "moneyness" value. A positive moneyness value is due to the public's willingness to pay a premium above the object's real goods or real asset value, because the object serves as a more convenient transacting medium than the next best alternative. Thus, money is defined in opportunity cost terms. To determine what portion of the object is moneyness value, the value of the object when not used as money must be estimated. This requires a counterfactual exercise. If the object was not used as money, what would its value be? Comparing that counterfactual value with its observed value in the marketplace captures the degree of moneyness in the object. This is similar to a hedonic pricing exercise where a good is broken down into its characteristics and the separate characteristics are priced out. Figure 1 illustrates this conceptualization.

Both opportunity cost and counterfactual analysis are required to measure what is money or, more accurately, what portion of an object's value is moneyness. If it is all moneyness value, then the object is a pure fiat money, namely if the object was not money then it would have no real goods or real asset value. If none of it is moneyness

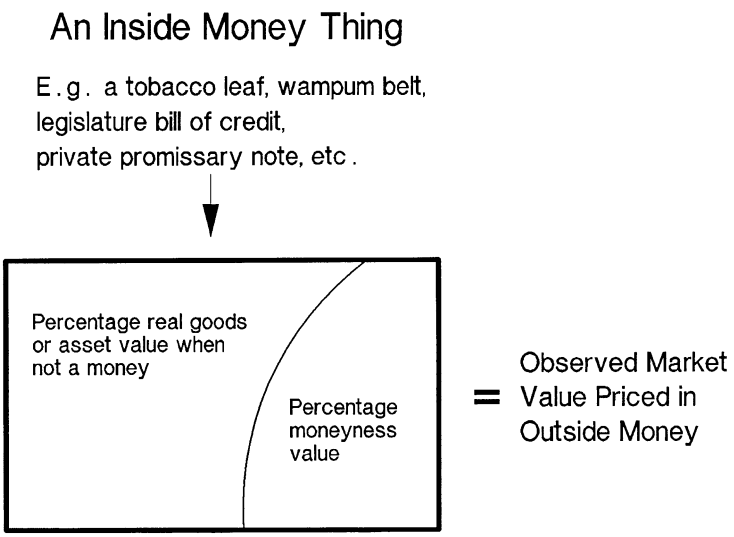


Fig. 1 Counterfactual decomposition approach to measuring money as "Moneyness" value. Notes: In the evaluation of legislature-emitted bills of credit in Figs. 4, 5, and 6, and Eqs. (4) and (5), *MEV* observed market value priced in outside money, *APV* percentage of real goods or asset value when not a money

value, then the object is a pure barter good or barter asset, namely if it was not money then its market value would remain the same. The definition of money used here is close to what scholars identify as pure fiat money. Money here is defined as the fiat money portion of an object – when measured in opportunity cost terms.

Imbedded in this definition of money is faith that subsequent traders will continue to see the object as a more convenient transacting medium than the next best alternative. Legal tender laws, and the designation by the government in what media taxes can be paid, contribute to that faith. Who is the guarantor, and their reputation for following through on their guarantee, of the real goods or real asset portion of the money object also affects that faith. Relatively risky real goods and relatively risky real assets are unlikely to possess moneyness value. Guarantors who are private individuals, such as for promissory notes or bills of exchange, may have occasioned less faith in the subsequent moneyness value of the money object than guarantors who had governmental power and law backing them.

This decomposition, or hedonic, approach to defining money is based on relative measurement. As such, a baseline value is needed against which any prospective inside money can be measured. I assume, but do not systematically estimate, that specie coins had moneyness value. The act of minting coins imbued gold and silver with a value beyond their non-money real goods value. Coins had a convenience in executing transactions above the next best alternative medium of exchange, especially across polity borders. As such, the international outside money, gold and silver specie coins, is used here as that baseline for comparative evaluation.

Commodity "Monies": Tobacco, Wampum, Beaver Pelts, Deerskins, Grain, Country Pay

In the seventeenth and early eighteenth centuries, colonists used local commodity monies to execute domestic transactions, such as tobacco in Maryland and Virginia, wampum in New York, beaver pelts in Canada, deerskins in Carolina, and grain or a mixture of local agricultural goods (called country pay) in the other colonies (Nettels 1934: 208–228). Were these commodities really monies, or were they just ubiquitously traded barter goods? While future research will have to address this question more systematically, the view here is that they were ubiquitous barter goods and so are subsumed into EB in Eq. (1).

Ideally, following the approach in Fig. 1, the counterfactual outside-money value of these commodities when these commodities were not suspected of being monies would need to be estimated. Those counterfactual estimates, when compared with the observed market values of these commodities when they were suspected of being monies, measures whether any moneyness values were present in these commodities. Doing such requires finding a location and time when the commodities in question were not considered to be monies, and then doing a controlled mapping of their relative values into the economy in question where the commodities were suspected of being monies. Doing such is a task for future research. Nevertheless, I will discuss two important commodity monies in colonial America, tobacco and wampum, and argue that they served more as ubiquitous barter goods than monies.

In the seventeenth- and early eighteenth-centuries, Virginia and Maryland used tobacco as a measure of value and, in a particular way, a medium of exchange. People counted in tobacco units, namely in so many physical pounds of tobacco, as though tobacco was a homogeneous good. Court-case restitutions and fines, along with government tax payments and fees, were priced in pounds of tobacco. Farmers and merchants kept accounts in pounds of tobacco.

Tobacco, however, was not a homogeneous good. Besides differing qualities or grades of tobacco from trash to middling to good, tobacco also came in different varieties. There was "sweet-scent" grown mostly in Virginia that was preferred by the English consuming public. There was "Oronoco" grown principally in Maryland but also in substantial quantities in Virginia that was preferred by the French consuming public, and there was "bright-leaf" grown largely in Maryland and preferred in the Dutch market. The Oronoco variety also had distinctly recognizable sub-varieties – some a closer substitute for sweet-scent than others (Grubb 2008; Middleton 1984: 109–111; Price 1995, 2: 510, 1973: 651, 666–668). The different varieties with their different European market niches each with their different quality grades yielded considerable value variation among tobaccos. Yet, local unit-of-account pricing was just in so many pounds of tobacco.

Marylanders did not carry tobacco leaves around in their pockets as cash nor did they typically make payments directly with tobacco leaves for individual items they purchased in person. Tobacco was a fragile commodity that could easily deteriorate when excessively handled or improperly stored. What were typically traded were claims on future tobacco production or the credits arising from the sale of tobacco in

England. Planters would trade barreled tobacco for goods from vessels that arrived in the Chesapeake from Britain, or barter tobacco for local and/or imported goods from domestic storekeepers often on credit. Typically these storekeepers were connected to tobacco shipping merchant-houses in Britain. At stores in the Chesapeake, goods would be priced in so many pounds of tobacco of the next harvest. After the next harvest the planter would deliver the tobacco to clear the book-credit debt. This book-credit style of tobacco money was barter using a commodity *numeraire* on a narrow trade-channel connecting tobacco growers to export/import merchants (Gould 1915: 38–77; Middleton 1984: 112–123; Price 1980: 16–17, 1995, 1: 195–196, 2: 509, 4: 11–14, 7: 65).

Finally, Maryland never made tobacco a legal tender, and British merchants referred to tobacco as a "barter" medium of exchange (Grubb 2008). Tobacco itself had a hard time circulating as an efficient medium of exchange. It was more likely a ubiquitous barter good and commodity *numeraire*, than a money as defined here.

In the seventeenth- and early eighteenth-centuries, New York and New England used wampum, in a particular way, as a medium of exchange. Wampum consisted of tubular beads made from clam and conch shells that were strung together or woven into belts. Wampum belts were a language that could be read. Each belt was unique, more a language than a store of value or means of payment (Shell 2013). A wampum belt was not a storehouse of so many beads. As such, wampum belts do not have a comparative metric. How people counted in wampum is unclear. While counting wampum in terms of handfuls of beads or string lengths was sometimes done, it was done loosely and imperfectly. The concept of 1 wampum versus 5 wampum, versus 10 wampum, versus 20 wampum lacked meaning. As such, wampum could not be effectively used as a unit of account.

Colonial farmers and merchants did not keep accounts in wampum. Even Europeans who dealt primarily in wampum, still predominantly priced their goods in, and kept their accounts in, guilders and florins. Europeans traded wampum to Native Americans in exchange for beaver pelts. Outside that connection, wampum had little value. The centrality of the fur trade gave wampum barter value among Europeans. The outcome of any ancillary exchange using wampum was the accumulation of that wampum by traders, who would eventually barter it for beaver pelts with Native Americans. As such, wampum was more likely a ubiquitous barter good than a money as defined here.

Paper Monies

Colonial governments could not create money per se. That was the prerogative of the British Crown (Davis 1964, 1: 271, 4: 206, 219). Colonial paper money creation was constrained by the British Parliament. In 1741, Parliament extended the 1720 Bubble Act to the colonies. Joint-stock corporations, except those chartered by the Crown, were illegal, thus making banking operations in the colonies prohibitively costly. Banks emitting paper banknote monies backed by fractional specie reserves would not appear before the American Revolution.

Colonial legislatures could, however, create transaction-able or exchangeable debt in the form of paper *bills of credit*, often called paper money by the colonists (Newman 2008). The British North American colonies were the first western economies to experiment with sizable emissions of paper monies. These bills were emitted by their respective legislatures who had paper money printed and placed in their respective colony's treasury. The legislatures directly spent this money on soldiers' pay, government salaries, military supplies, construction projects, etc. In some cases, they loaned it out on interest to their respective citizens who secured these loans with mortgages on the respective citizens' lands. These colony-specific legislature-issued paper monies formed an important part of the circulating medium of exchange in many colonies. These colonial paper-money experiments were neither uniform nor coordinated across the colonies. They were instituted piecemeal – at different times with different motivations and goals – though often they were first initiated to deal with wartime emergencies. Their institutional structures and relative performances varied. While not consciously intended to be so, this was monetary experimentation on a grand scale (Brock 1975; Grubb 2016b; Newman 2008).

Paper money emissions began in 1690 in Massachusetts; 1703 in South Carolina; 1709 in Connecticut, New Hampshire, New Jersey, and New York; 1710 in Rhode Island; 1712 in North Carolina; 1723 in Pennsylvania and Delaware; 1733 in Maryland; and 1755 in Virginia and Georgia. Once initiated, colonial legislatures, with irregular frequency, continued to enact new emissions. With one minor exception, these colonies maintained some bills in circulation through 1774 (Brock 1975; Grubb 2016b; Newman 2008).

Figure 2 shows the quantity of bills in circulation for the six colonies in the middle of the British North American colonial set. The quantity for each colony is measured at their face value and then converted to Queen Anne's Proclamation rate of 1.33–1 pound sterling when not emitted at that rate (see section "Prices and Money" below). This conversion allows for a direct comparison of amounts by value between these colonies and over time. Figure 2 shows the importance of the Seven Years War on affecting the amounts of paper money in circulation (Grubb 2016c; Wicker 1985). Figure 3 provides an example of a bill of credit. For more examples, see Eric P. Newman (2008).

Legislatures put considerable time and effort into crafting their paper money laws, with comparably lengthy legal space devoted to how each emission was structured. Some emissions were made a legal tender; some were not. Some emissions paid an interest; most did not. Because paper money was considered a borrowing by the government that had to be paid back in the future, paying interest seemed both just and necessary to maintain the face value of the money. However, because the borrowing was only from the colony's citizens and these same citizens would have to pay the higher future taxes required to pay the interest back to themselves, paying interest was considered both unnecessary and burdensome. In addition, colonists worried that paper money that paid interest would be held as investments and not be used as intended, namely as a circulating medium of exchange (Grubb 2016b: 156–157, 183–189; Newman 2008: 18).

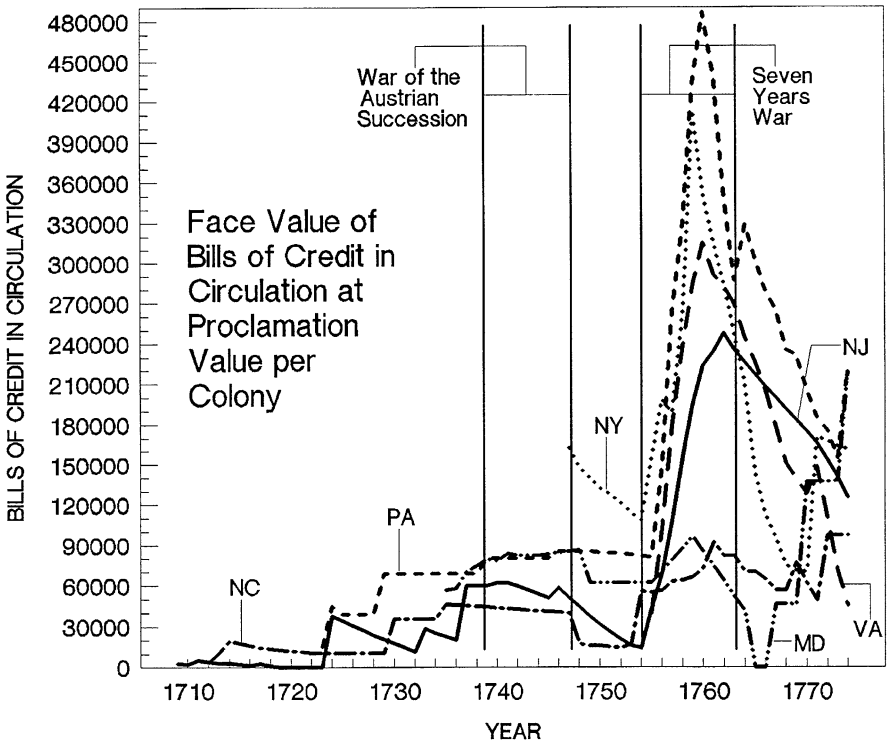


Fig. 2 Quantities of paper money in circulation in six colonies. (Sources: Brock 1975: 82–83, 346–347; Cutsail and Grubb 2017; Grubb 2005b: 35, 2015a: 15–16, 2017: 106). Notes: *MD* Maryland, *NC* North Carolina, *NJ* New Jersey, *NY* New York, *PA* Pennsylvania, *VA* Virginia. While New York began emitting paper money in 1709, data on the yearly amounts in circulation are only currently available from 1748 on. Future research needs to fill in the 1709–1748 amounts for New York. See text for when and by what magnitude amounts had to be converted to proclamation values from non-proclamation values

Emissions had differing redemption (maturity) intervals specified in their authorizing legislation – some with complex and variable maturity structures. Some emissions had portions set aside as land-backed loans for the colony's citizens, and some did not. Some emissions had future redemptions linked to an accumulating sinking fund comprised of specie-denominated assets. Most emissions, however, linked future redemptions to explicit taxes or land-backed loans that could be paid in those emissions or their specie equivalents. All emissions were structured to be like bearer-bonds requiring an explicit redemption exercise, namely establishing a maturity payoff structure, to extinguish the principal expressed on its face, with most being structured like zero-coupon bonds (Celia and Grubb 2016; Cutsail and Grubb 2017; Grubb 2008, 2012, 2015a, 2016a, b, c, d, 2017).

Fig. 3 Virginia paper money (bills of credit), issued 12 October 1758. (Source: Newman 2008: 439, with permission of the Eric P. Newman Numismatic Education Society). Note: Bills of credit were also called treasury notes in Virginia. For more examples among all the colonies, see Newman (2008)

Material printed on the face of the bills informed holders about each emission. Besides the monetary denomination, two additional pieces of information appeared on all bills, namely the date of the law authorizing the emission and the colony emitting the bill. Thus, citizens could always distinguish emissions by the colony-date printed on the face of the bill. This information, when matched to the authorizing legislation, provided citizens with the complete details about how each emission was structured.

The material printed on the face of the bills also stated consistently whether interest would be paid. If a bill paid interest, that interest rate was stated on the face of the bill. By inference, when no interest rate was stated on the bill's face, citizens could be certain that the bill paid no interest. By contrast, redemption dates for specific emissions were only irregularly printed on the face of the bills. In some colonies, redemption dates were often printed on the face of the bills, e.g., see Connecticut, New Hampshire, Rhode Island, North Carolina, Georgia and, after 1740, South Carolina. In other colonies, no redemption dates were printed on the bills even though redemption dates were so legislated, e.g., see New York, New Jersey, Pennsylvania, Maryland, and Virginia (Grubb 2016b; Newman 2008).

With the exception of Maryland, most colonies initiated their paper money systems and made their largest paper money emissions to cover emergency wartime expenditures, with Queen Anne's War, King George's War and, especially, the Seven Year's War placing large claims on colonial budgets. Future taxes were

concurrently legislated to redeem these wartime emissions. In the absence of war expenditures as a vehicle to spend paper money into general circulation, colonial legislatures typically turned to land-bank emissions as their peacetime vehicle to put paper money into general circulation (Grubb 2015a, 2016b, c).

Land-bank emissions entailed printing bills and then allowing citizens to borrow the bills on interest with citizens pledging their lands as collateral for the loans. The acts authorizing land-bank emissions established the interest rate to be paid by the borrower (typically a few percentage points below the market rate), the principal repayment period, how interest and principal payments would be used by the government, how loaned bills would be retired from circulation, a loan size limit on individual borrowings, collateral requirements (typically double the value in land assessment per loaned amount), procedures for assessing land values and for foreclosing on and selling the lands of delinquent borrowers, and a system of county loan offices and loan-office commissioners with their salaries. Interest income from these land-bank loans was a major source of colonial government revenue (Grubb 2015a, 2016b, c).

The value and performance of colonial paper monies are explained by the redemption (maturity) structure as executed by each colony. Paper monies had little extra value as money per se above their expected real asset present value. In effect, paper monies were primarily real barter assets. Their observed market values match what rational bond pricing yields. For New Jersey, Virginia, and North Carolina this outcome is shown in Figs. 4, 5, and 6. Even David Hume referred to these bills as a "commodity" and not as a money (Rotwein 1970: 69).

The observed market value of paper money is what colonists paid in paper money to purchase bills of exchange paying pounds sterling in England. This price, adjusted for the transaction and time cost of getting bills of exchange to England, liquidated, and the specie money returned to the colonies – approximately 7%, measures the specie, or outside money, value of these paper monies in the colonies. *MEV* expresses these values as a percentage of their respective paper money's face value, see Eq. (4) (Cutsail and Grubb 2017; Grubb 2016a, b, d; McCusker 1978).

$$MEV = \text{colonial bills of credit at face value equal to one pound sterling}/\left(\text{colonial bills of credit needed to purchase a one pound sterling bill of exchange}/0.93\right) \tag{4}$$

As zero-coupon bonds, colonial paper monies required time-discounting to ascertain their real asset present value (*APV*). If the public is assumed to know the face value of the quantity of bills in circulation (Mp_j) and the amounts redeemed and retired from circulation each year (RED_t), and if the public responds only to the *expected* redemption of the *average* bill currently outstanding, then Eq. (5) shows how the expected *APV* of the average bill in circulation is calculated in year j. Equation (5) is adapted from the basic continuous discounting present value formula ($PV = FVe^{-rt}$), where PV is present value, FV is face value, r is the discount rate, and t is the time to maturity. The discount rate (r) is taken from statements by colonists about what the typical interest rate was for low-risk assets – what in the

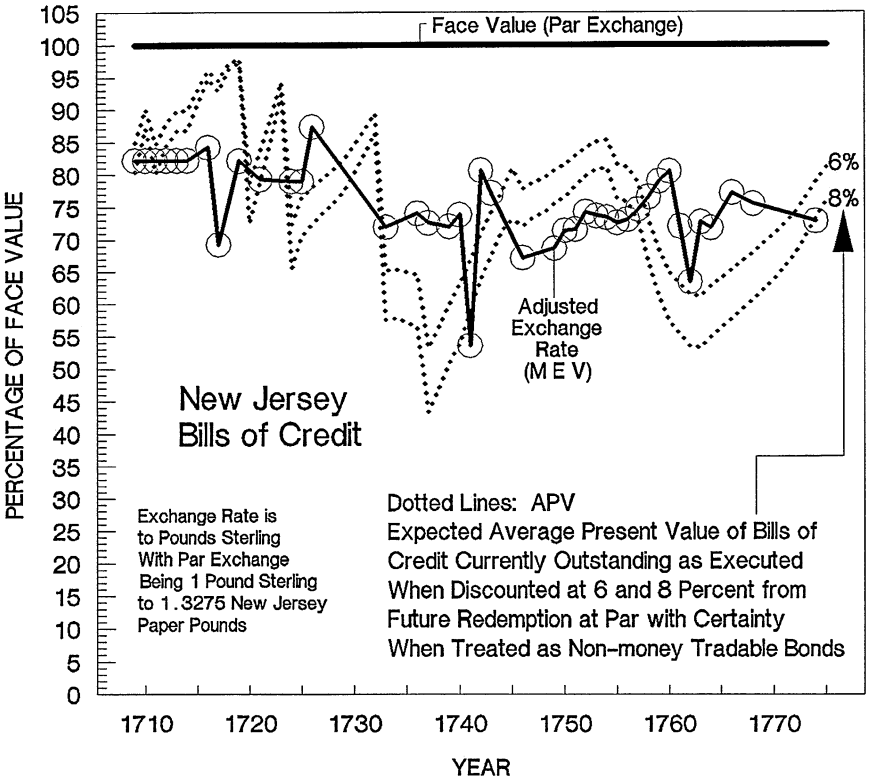


Fig. 4 Colonial New Jersey paper pounds (bills of credit) (Source: Grubb 2016d). Circles represent yearly data points for *MEV* with linear interpolative lines connecting them

modern literature would be comparable to the riskless rate. By far, 6% was the most common rate mentioned, but rates up to 8% were sometimes mentioned (Grubb 2016a, b, d). RED_N is the amount in the last year N that satisfies Eq. (6).

$$APV_{\mathrm{j}} = \sum_{\mathrm{t=j}}^{N} \left(RED_{\mathrm{t}} / Mp_{\mathrm{j}} \right) e^{-r\mathrm{t}} \tag{5}$$

$$\sum_{\mathrm{t=j}}^{N} \left(RED_{\mathrm{t}} / Mp_{\mathrm{j}} \right) = 1. \tag{6}$$

In words, Eq. (5) assumes that the amount of paper money outstanding in a given year is redeemed at face value by all bills actually redeemed in the immediately following years until the year when that original amount is fully redeemed. These yearly redemption amounts are divided by the initial amount outstanding from the chosen year to assign a yearly weight to its contribution in the redemption process.

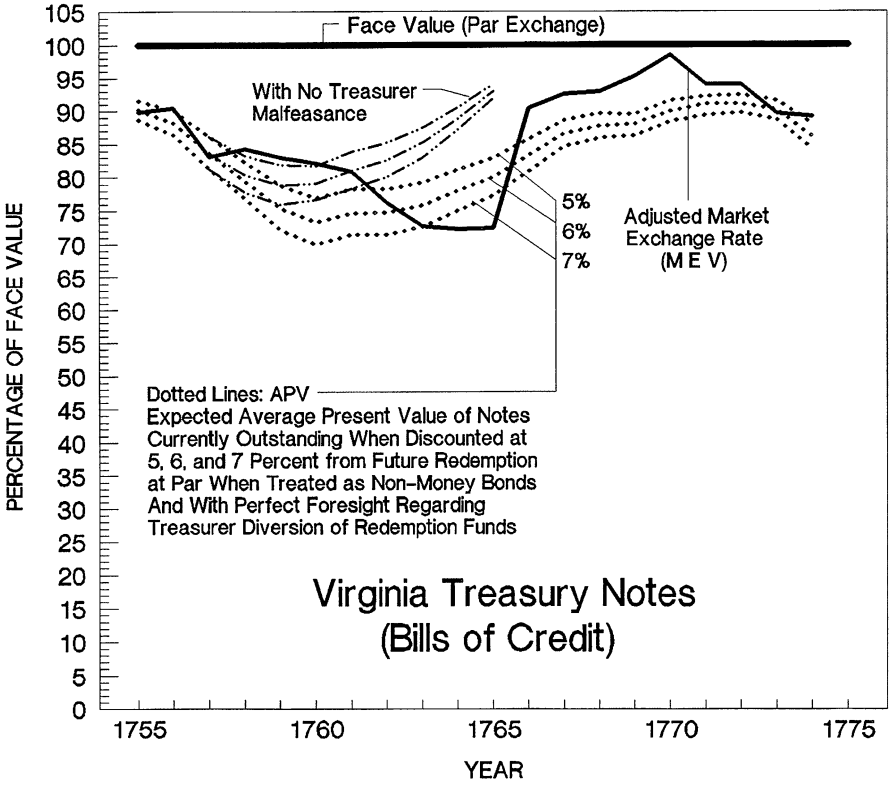


Fig. 5 Colonial Virginia paper pounds (bills of credit also called treasury notes). (Source: Grubb 2016a)

The time discounts between the initial year and the redemption year are multiplied by the contribution-weights for their respective years. The time-discount-weight values for each year are summed to get the expected present value of a representative bill outstanding for that chosen year. *APV*s are counterfactual values for paper money. They are what rational bond pricing would produce under the assumption that colonists saw their paper money not as money per se but only as zero-coupon bonds, and that colonists correctly forecasted the emission and redemption path of those bonds.

MEV could be within a percentage point of that drawn due to measurement error in the transacting cost of liquidating sterling bills of exchange drawn on London and turning them into specie money usable in the colonies. While 6% is the best-guess discount rate, uncertainty over that rate means that 5%, 7%, or 8% could also be used. Considering the range of possible measurement error in *MEV* and uncertainty over which discount rate to use for *APV*, the hypothesis that *MEV* is primarily and predominantly comprised of *APV* cannot be rejected given the data in Figs. 4, 5, and 6. Over their entire paper money era, using a 6% discount rate, $APV \approx MEV$

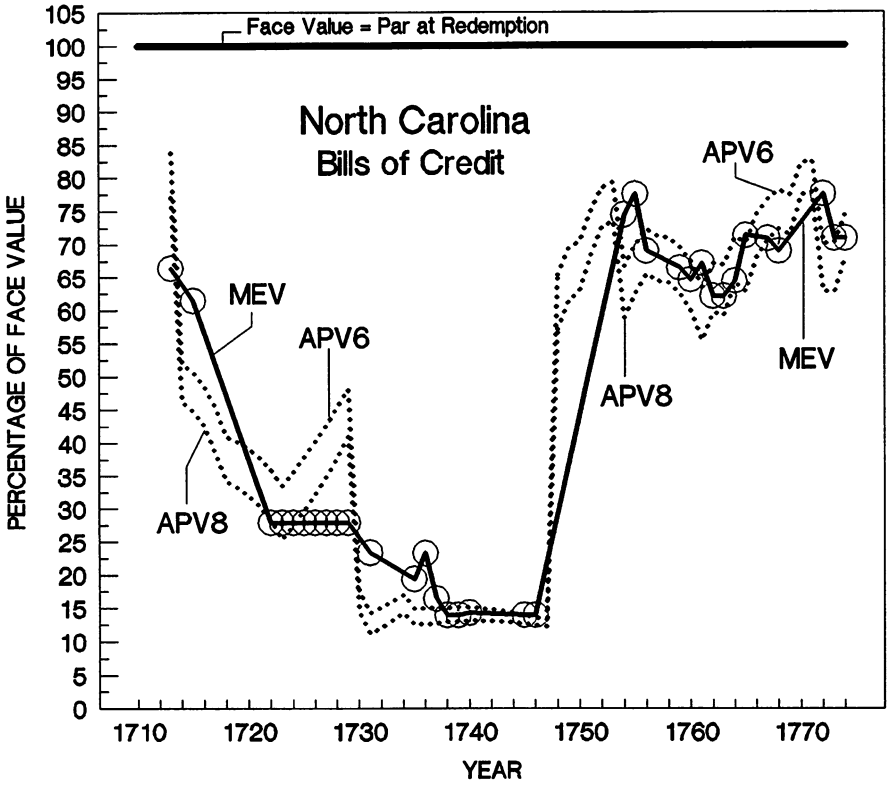


Fig. 6 Colonial North Carolina paper pounds (bills of credit) (Source: Cutsail and Grubb (2017)). Notes: APV6 is *APV* where the discount rate *r* is 6%. APV8 is *APV* where the discount rate *r* is 8%. Circles represent yearly data points for *MEV* with linear interpolative lines connecting them

in New Jersey and North Carolina, and *APV* equals 97% of *MEV* in Virginia. If the analysis is confined to periods when bills were emitted through land-bank loans, and to periods when no issues of credibility in tax redemption structures were suspected, then 92–94% of *MEV* was *APV*, leaving 6–8% of *MEV* as moneyness value. This was enough moneyness value to make the bills the preferred medium of exchange for executing local transactions (Grubb 2016a, d).

While rational bond pricing under perfect foresight appears to be the best single predictor of the actual market value of colonial paper monies, applications to more colonies are required to confirm this conclusion. As a general finding, this result could be considered a manifestation of Ricardian equivalence (Barro 1974; Abel 1987). While colonial paper monies traded below face value (below par), this does not mean that paper monies had depreciated or that inflation had eroded their value. The differences between the bills' face values and their *MEV* were overwhelmingly due to time-discounting and not depreciation. The prior literature has habitually confused time-discounting with depreciation (Brock 1975; Ernst 1973; Newman 2008: 11–12).

Bills of credit displaced efficient barter structures for executing domestic transactions. Judging by the amount of bills colonial legislatures maintained as a reserve in their treasuries to replace worn and torn bills that could no longer continue in circulation, and estimates of the portion of taxes paid in bills versus alternative media, the bills experienced extensive hand to hand circulation (Grubb 2012, 2015b, 2016b). It only takes a tiny premium above the real asset barter value to signal that bills were the preferred media for local exchange. In opportunity cost terms, an infinitesimally small profit makes an actor shift methods.

While bills of credit may have been marginally superior to efficient barter, they were still a cumbersome media of exchange. These bills did not circulate at face value due to time discounting. Merchants adjusted their prices to reflect the time-discounted value of these paper monies. The average time-discounts varied from year to year. Using bills of credit as a medium of exchange required some calculation and guesswork regarding the money's current value.

The typical legislated redemption structure used a $1/N$ redemption rate over an N-year redemption window. Slight variations on the $1/N$ rate were sometimes used, and N often varied, and redemption windows often overlapped, from emission to subsequent emission. Lengthy redemption windows were used to give fiscal credibility to the redemption exercise by smoothing the tax obligation over time. However, legislatures did not establish a method to determine which bills could be redeemed in which year within the N-year redemption window. Thus, while the average expected value can be forecast, see Figs. 4, 5, and 6, the realized value of a particular bill in any given year could vary from face value (one of the $1/N$ bills redeemed that year) to a full time-discounted reduction from year N off the face value. Again, this made for a cumbersome medium of exchange, as well as a poor store of value.

Colonial Maryland was an exception. Maryland's major paper money emissions had unique redemption dates, with all bills redeemed in the legislated year N and none before. Redemption was paid out of a sinking fund filled over time with purchases of stock in the Bank of England using yearly tax assessments in specie on tobacco exports. Citizens could forecast that the fund would be more than sufficient to execute the promised redemption. Unlike the paper monies of New Jersey, Virginia, and North Carolina (Figs. 4, 5, and 6), Maryland's paper money traded in the marketplace at a value well above its real asset present value, i.e., by 19%. While still circulating below face value, it nevertheless had a goodly bit of moneyness value in excess of its rational bond pricing (Celia and Grubb 2016; Grubb 2005b, 2008, 2016b: 168, 180). This difference is consistent with all Maryland bills being relatively more predictable in value in each year, namely having no variance in their individual realized yearly value versus their expected average yearly value.

Post-Revolution, Americans were able to legally incorporate banks. Banknotes always paid their face value in specie on demand at the issuing bank (barring liquidity and reputation crises). This made banknotes competitively superior inside paper monies to bills of credit. The U.S. Constitution also constitutionally banned legislature-issued bills of credit after 1789 (Grubb 2006).

Oversight by the British government and by the proprietors of some colonies constrained colonial paper money creation. Paper money acts could be disallowed. The debates with a colony's proprietor were often not about paper money per se, but about political rights and prerogatives. Typically, proprietors wanted to be exempt from having to accept their colony's bills in payment of the quit rents owed them, and proprietors also wanted a say in how bills could be spent. Colonial assemblies typically wanted the opposite. Wrangling over these issues often delayed and shaped paper money acts (Brock 1975: 354–362; Grubb 2008).

The British government directed its colonial governors to resist paper money acts, especially when not required for war funding. Suspending clauses were required in paper money acts so the acts could be reviewed by the Board of Trade and potentially disallowed. The Board became more aggressive in reviewing and then disallowing paper money acts, especially disallowing land-bank paper money acts, after King George's War (Brock 1975; Celia and Grubb 2016; Ernst 1973; Grubb 2015a, 2016b, c, d; Labaree 1967: 212–238).

Finally, the British Parliament regulated colonial paper money through the Currency Acts of 1751 and 1764. Complaints about New England's paper money led Parliament to pass the Currency Act of 1751. This Act applied only to New England. It outlawed making bills of credit a legal tender in private transactions. It also restricted the emission-to-final-redemption interval to a maximum of 2 years in peacetime and 5 years during wartime. Complaints about Virginia's bills in the early 1760s led Parliament to pass the Currency Act of 1764. This Act applied to all colonies and outlawed making bills of credit a legal tender for any transaction. It did not restrict the emission-to-final-redemption interval as was done to New England by the Currency Act of 1751. After colonial protests, Parliament in 1773 amended the Currency Act of 1764 to allow bills to be made a de jure legal tender for public debts of the issuing government (Brock 1975; Ernst 1973).

The prior literature fails to explain why these Acts took the form they did. The Acts did not prohibit paper money as is often erroneously stated in the literature. They only restricted its emission-to-final-redemption span and its legal tender status. These restrictions only make sense if bills of credit were structured as zero-coupon bonds that required time-discounting to assess their current market value and if courts did not take this into account in tort law cases.

Colonial legislatures, when allowed, made their bills of credit a legal tender within their jurisdiction. They did this to establish and guarantee the payment of face value in specie equivalence at redemption (maturity), and so anchor the time-discounting value-path of the bills. Legal tender laws were not used, i.e., no legal cases can be found, to force citizens to accept payment in bills at their specie-equivalent face value in private transactions, or to exchange bills, pre-redemption, at their specie-equivalent face value. Local prices and exchange rates were adjusted to reflect the bills present value compared with its specie-equivalent face value.

Legal tender laws created problems in assigning restitution in breach-of-contract tort-law cases. To make the plaintiff (creditor) whole, colonial courts had to assign a value in the legal tender. If the court used the paper money's face value rather than its current market or asset present value, the creditor lost out. While there is little

evidence that American courts used the bill's face value to assess restitution value, the fear of such drove British merchant creditors to complain about colonial paper monies (Brock 1975; Labaree 1967: 229–231).

The two Currency Acts were intended to lessen the possibility of loss to British creditors if, by chance, colonial courts assigned the bill's face value, rather than the bill's current market value, when making plaintiffs (the British merchant creditors) whole in breach-of-contract cases. This explains the restriction on the emission-to-final-redemption span in the 1751 Currency Act. A shorter redemption span reduced the gap between the time-discounted asset present value and the face value of the bills. This restriction, however, could not completely solve the tort law problem as some gap in value between face value and current market value would still exist. The complete solution was to ban using bills as a legal tender. This solved the tort law problem, as restitution in breach-of-contract cases were no longer required to be in bills of credit, thus removing British merchant fears of being potentially cheated in colonial-court debt-restitution cases (Brock 1975; Grubb 2016b).

Prices and Money

Prices for domestically traded goods in the seventeenth century are poorly documented. The absence of printed newspapers, which did not appear until the eighteenth century, meant no published price currents exist. In addition, early prices were expressed in the primary commodity unit of the colony, such as pounds of tobacco in Virginia and Maryland or bushels of grain elsewhere. Compiling such prices for enough domestically traded goods from the probate inventories, court records, and merchant account books in sufficient quantities to construct reasonable price indices awaits future scholarship.

Yearly prices for two export goods have been compiled for the seventeenth century, namely codfish in New England from 1634 on, and tobacco in Virginia from 1647 on (Carter et al. 2006: 675–690). Little discernable trend in codfish prices appears between 1634 and 1701. Tobacco prices fell by over half from 1647 to 1665, but then stayed relatively constant to 1697.

These export prices were expressed in colony-specific units of account, called that colony's pound. These units of account were regarded as imaginary monies because there was no physical money "thing" representing them. These colonial pounds were set at a fraction of pound sterling, with that fraction not being the same across colonies, and sometimes being changed within a given colony at different points in time (Nettels 1934: 240–249). Prices in the eighteenth century were mostly expressed in these colony-pound units of account. When each colony emitted its own paper money, the face value of that paper money at redemption was expressed in their respective colonial pound unit of account.

Queen Anne's 1704 Proclamation set the rating for these colonial pounds at a maximum of 1.33 colonial pounds to 1 pound sterling. Colonial pounds set at that rate were called proclamation money. Colonies did not always abided by this proclamation in their laws setting the face value at redemption for newly emitted

paper monies. This was a point of contention between Crown-appointed governors and colonial legislatures. For example, North Carolina maintained its North Carolina pound at 1.5–1 pound sterling until 1748 when it switch to the proclamation rating. Virginia maintain its Virginia pound at 1.25–1 pound sterling throughout. New Jersey maintained its New Jersey pound at 1.55–1 pound sterling until 1723 when it switched to the proclamation rate. New York appears to have maintained its New York pound at 1.55–1 pound sterling throughout, whereas Maryland and Pennsylvania appear to have held to the proclamation rating for their respective colony's pound throughout (Cutsail and Grubb 2017; Grubb 2008, 2015a, 2016b, 2017; McCusker 1978; Nettels 1934: 242–243).

Newspaper price currents, as well as prices compiled from probate inventories and merchant account books, become prevalent enough at various points in the eighteenth century to construct yearly price indices for some colonies. Figure 7 displays these price indices from 1748 to 1775 for six colonies. Statistically, these price indices are trend stationary with short half-lives to shocks. The trends, while positive, are less

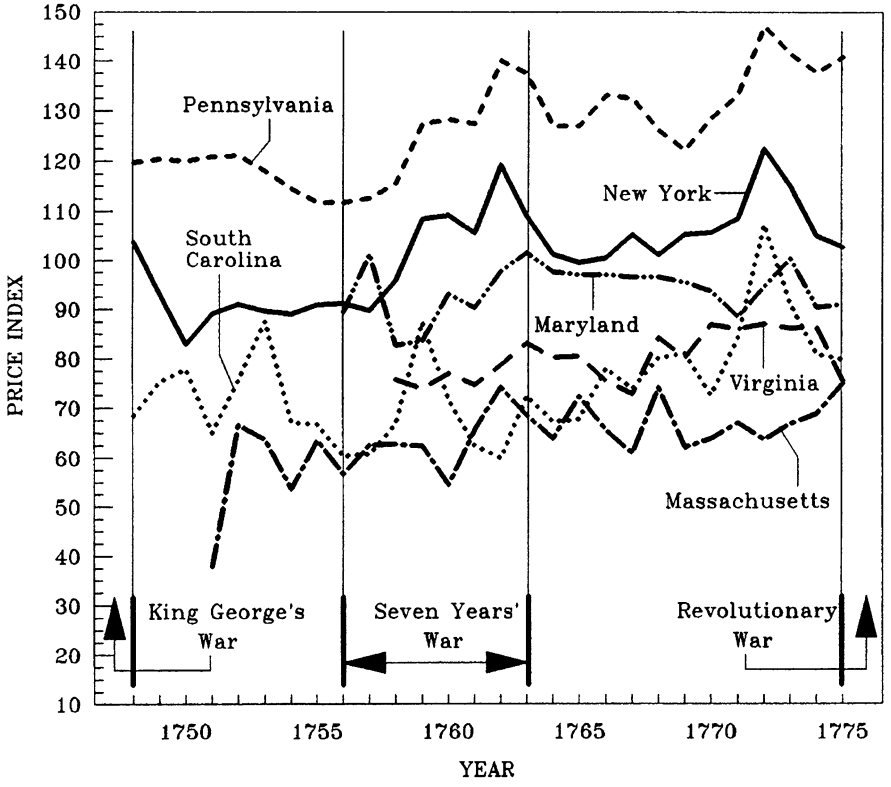


Fig. 7 Price indices by colony, 1748–1775. (Sources: Adams 1986: 643; Bezanson et al. 1936: 388; Carter et al. 2006: 682–687; Cole 1938: 120–122, 135, 155–156; Grubb 2003: 1782–1784; Rothenberg 1979: 983–984)

than, or not statistically different from, the trend in prices in England expressed in pounds sterling (Grubb 2003: 1782–1784, 2005a: 1346). It should be noted that these price indices are based on relatively small sets of goods heavily weighted toward imported and exported goods. More work to compile prices for domestically traded goods in more locations in the colonies is needed.

With the exception of New England (Massachusetts) between 1730 and 1750, no statistically significant relationship of any effective magnitude can be found between changes in the face-value quantities of a colony's bills of credit in circulation and changes in prices in their respective colonies. Nor can any statistically significant relationship of any effective magnitude be found between changes in the face-value quantities of bills of credit in circulation and changes in prices for contiguous groups of colonies treated as a single monetary unit (Cutsail and Grubb 2017; Grubb 2019; Officer 2005; West 1978).

The change in the amount of bills of credit in circulation were often large, especially during the Seven Year's War (Fig. 2). For example, the growth rate in bills of credit in circulation for New York between 1754 and 1759 averaged 47% per year, for New Jersey between 1732 and 1762 66% per year, for Pennsylvania between 1723 and 1760 83% per year, for Maryland between 1764 and 1774 39% per year, and for Virginia between 1755 and 1760 116% per year. Yet prices hardly moved across these periods (Grubb 2019; Wicker 1985).

Even the initial creation of paper money by a colony had little effect on prices. Price series for a few years before, versus a few years after, bills of credit were first emitted exist for Pennsylvania, Maryland, and Virginia. Pennsylvania first emitted bills in 1723, Maryland in 1733, and Virginia in 1755. Prices before paper monies were first emitted were expressed in their respective local units of account, an imaginary money before physical bills expressed in that unit of account were issued. Figures 8 and 9 show that prices in these colonies were not substantially different before versus after paper monies were first emitted – being slightly higher in Pennsylvania, slightly lower in Virginia, and about the same in Maryland after setting aside the influence of the War of the Austrian Succession. The amount of bills per white capita in these emissions for Pennsylvania was 1.2 Pennsylvania pounds in 1724 and 2.3 in 1731. For Virginia, it was 2 Virginia pounds in 1760. These amounts represent only 4–7.5% of the yearly income of laborers in Philadelphia (Carter et al. 2006, 5: 652; Grubb 2008, 2011: 260–261, 2016b). Enough efficient barter trades took place to support comparable pricing even when bills of credit as a medium of exchange for that trade had not yet been created. After bills of credit were created, the use of efficient barter still dominated domestic transactions.

Statistically, the strongest predictor of prices this year in any colony are last year's prices (Grubb 2019). This finding is consistent with prices being determined primarily by the constancy in the barter portion of the economy. In colonial economies where there was little technological or productivity change, and where prices were being primarily determined in trades using EB rather than Mp, prices this year should be strongly determined by prices in the immediately prior years. The slight gain in moneyness value associated with bills of credit meant that they displaced

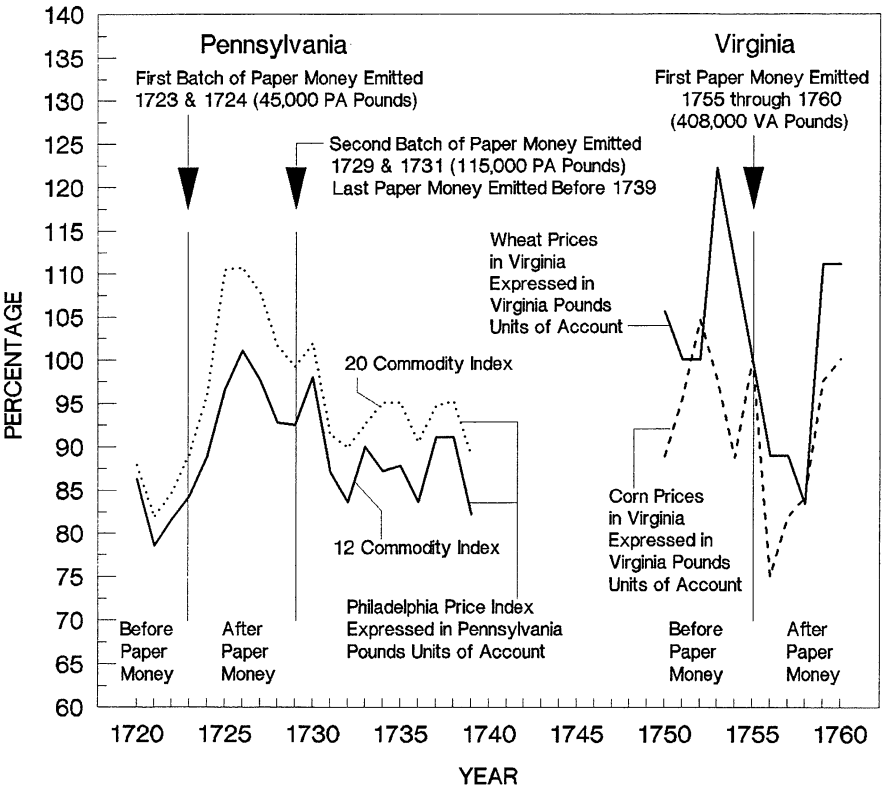


Fig. 8 Prices before and after the first emissions of bills of credit in Pennsylvania and Virginia (Sources: Carter et al. 2006: 685, 687; Bezanson et al. 1935: 433; Grubb 2012, 2016b). The Philadelphia data are arithmetic unweighted price indices with a base year 100 = 1741–1745. The Virginia price data are expressed as the percentage of the price listed for 1755

efficient barter for executing some portion of domestic transactions. Given how under-monetized colonial economies were, there was substantial room to expand the quantities of bills of credit without affecting the underlying magnitude of the medium of exchange in use (M), see Eq. (7).

$$\begin{aligned} &\Delta \mathrm{Mp} = -\Delta \mathrm{EB} \text{ up to some threshold } \alpha \mathrm{EB} \text{ where} \\ &1 \geq \alpha \geq 0. \text{Thus, } \Delta(\mathrm{M} + \mathrm{EB}) \approx 0 \approx \Delta M. \end{aligned} \tag{7}$$

Price movements were driven by random real shocks, such as weather, wars, and trading panics. Figures 7 and 9 show that prices were temporarily driven up during wars and during the 1772 transatlantic trading panic (Sheridan 1960). Finally, the small reduction in transaction costs that accompanied using bills of credit rather than barter structures to execute local trades explains the small increases in output

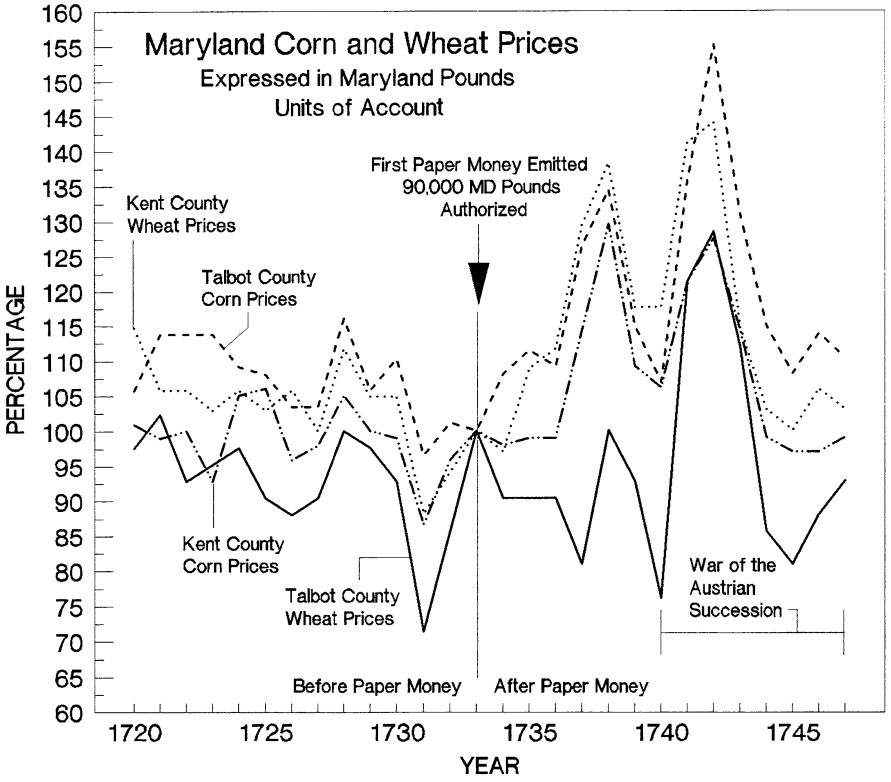


Fig. 9 Prices before and after the first emissions of bills of credit in Maryland (Sources: Clemens 1980: 226–227; Grubb 2012, 2016b). Notes: All prices are expressed as the percentage of the price listed for 1733

associated with increases in the amount of bills of credit in circulation found in some studies (Rousseau 2007).

Colonial legislatures could emit more bills of credit with impunity regarding the effect more bills had on prices. This was a boon to colonial legislatures and colonial economies. It gave them flexibility in how to finance government spending and it enhanced internal trade (Grubb 2016c). While colonial legislatures were not constrained by the expected effect on prices in their considerations over whether to emit more bills of credit, they were constrained by the need for fiscal credibility. Being zero-coupon bonds, the public had to believe that the redemption of the bills by future tax or mortgage payments would be credibly executed. The public's ability to pay taxes and mortgage payments were the primary real-economy constraint on colonial legislatures' ability to emit more bills of credit.

The value and performance of colonial bills of credit and their relation to prices in the colonial economy raises questions about why British merchants, British monetary writers, and the British government objected so strenuously and used

their power to repress and thwart colonial paper money legislation, especially the disallowance of land-bank paper money acts. Future scholars need to reassess these British objections. The polemical rhetoric needs to be seen for what it was and how it was used, and not as a substitute for historical facts. A more nuanced political history of the conflict between the Crown and colonial legislatures over monetary powers is needed.

Cross-References

► Experiments with Paper Money

References

Abel AB (1987) Ricardian equivalence theorem. In: Eatwell J, Milgate M, Newman P (eds) The New Palgrave: a dictionary of economics, vol 4. Macmillan Press, London, pp 174–179

Adams DR Jr (1986) Prices and wages in Maryland, 1750–1850. J Econ Hist 46:625–645

Barro RJ (1974) Are government bonds net wealth. J Polit Econ 82:1095–1117

Bezanson A, Gray RD, Hussey M (1935) Prices in colonial Pennsylvania. University of Pennsylvania Press, Philadelphia

Bezanson A, Gray RD, Hussey M (1936) Wholesale prices in Philadelphia, 1784–1861. University of Pennsylvania Press, Philadelphia

Brock LV (1975) The currency of the American colonies, 1700–1764. Arno, New York

Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G (eds) (2006) Historical statistics of the United States earliest times to the present, millennial edition, vol 5. Cambridge University Press, New York

Celia J, Grubb F (2016) Non-legal-tender paper money: the structure and performance of Maryland's bills of credit, 1767–1775. Econ Hist Rev 69:1132–1156

Clemens PGE (1980) The Atlantic economy and colonial Maryland's Eastern Shore. Cornell University Press, Ithaca

Cole AH (1938) Wholesale commodity prices in the United States, 1700–1861. Harvard University Press, Cambridge, MA

Cutsail C, Grubb F (2017) Colonial North Carolina's paper money regimes, 1712–1774. NBER working paper #23783. http://www.nber.org/papers/w23783

Davis AM (ed) (1964) Colonial currency reprints, 1682–1751, vols 1–4. Augustus M. Kelley, New York

Ernst JA (1973) Money and politics in America, 1755–1775. University of North Carolina Press, Chapel Hill

Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1867–1960. Princeton University Press, Princeton

Gould CP (1915) Money and transportation in Maryland, 1720–1765. vol 33. Johns Hopkins Press, Baltimore, pp 9–176

Grubb F (2003) Creating the U.S. dollar currency union, 1748-1811: a quest for monetary stability or a usurpation of state sovereignty for personal gain? Am Econ Rev 93:1778–1798

Grubb F (2004) The circulating medium of exchange in colonial Pennsylvania, 1729–1775: new estimates of monetary composition, performance, and economic growth. Explor Econ Hist 41:329–360

Grubb F (2005a) State 'currencies' and the transition to the U.S. dollar: reply – including a new view from Canada. Am Econ Rev 95:1341–1348
Grubb F (2005b) Two theories of money reconciled: the colonial puzzle revisited with new evidence. NBER working paper #11784. http://www.nber.org/papers/w11784
Grubb F (2006) The U.S. Constitution and monetary powers: an analysis of the 1787 Constitutional convention and the constitutional transformation of the U.S. monetary system. Financ Hist Rev 13:43–71
Grubb F (2008) Creating Maryland's paper money economy, 1720–1739: the role of power, print, and markets. NBER working paper #13974. http://www.nber.org/papers/w13974
Grubb F (2011) German immigration and servitude in America, 1709–1920. Routledge, New York
Grubb F (2012) Chronic specie scarcity and efficient barter: the problem of maintaining an outside money supply in British colonial America. NBER working paper #18099. http://www.nber.org/papers/w18099
Grubb F (2015a) Colonial New Jersey's paper money regime, 1709–75: a forensic accounting reconstruction of the data. Hist Methods 48:13–34
Grubb F (2015b) Common currency versus currency union: the U.S. continental dollar and denominational structure, 1775–1779. NBER working paper #21728. http://www.nber.org/papers/w21728
Grubb F (2016a) Colonial Virginia's paper money regime, 1755–1774: value decomposition and performance. NBER working paper #21881. http://www.nber.org/papers/w21881
Grubb F (2016b) Is paper money just paper money? Experimentation and variation in the paper monies issued by the American colonies from 1690 to 1775. Res Econ Hist 32:147–224
Grubb F (2016c) Colonial New Jersey's provincial fiscal structure, 1709–1775: spending obligations, revenue sources, and tax burdens during peace and war. Financ Hist Rev 23:133–163
Grubb F (2016d) Colonial New Jersey paper money, 1709–1775: value decomposition and performance. J Econ Hist 76:1216–1232
Grubb F (2017) Colonial Virginia's paper money regime, 1755–1774: a forensic accounting reconstruction of the data. Hist Methods 50:96–112
Grubb F (2019) Colonial paper money and the quantity theory of money: an extension. Soc Sci Hist 43:forthcoming
Labaree LW (ed) (1967) Royal instructions to British Colonial Governors, 1670–1776, vol 1. Octagon Books, New York
McCusker JJ (1978) Money and exchange in Europe and America, 1600–1775. University of North Carolina Press, Chapel Hill
McCusker JJ, Menard RR (1985) The economy of British America, 1607–1789. University of North Carolina Press, Chapel Hill
Middleton AP (1984) Tobacco coast: a maritime history of Chesapeake Bay in the colonial era. Johns Hopkins University Press, Baltimore
Nettels CP (1934) The money supply of the American colonies before 1720. University of Wisconsin Press, Madison
Newman EP (2008) The early paper money of America, 5th edn. Krause Publications, Iola
Officer LH (2005) The quantity theory in New England, 1703–1749: new data to analyze an old question. Explor Econ Hist 42:101–121
Price JM (1973) France and the Chesapeake: a history of the French tobacco monopoly, 1674–1791, and its relationship to the British and American tobacco trades, vol 1. University of Michigan Press, Ann Arbor
Price JM (1980) Capital and credit in British overseas trade: the view from the Chesapeake, 1700–1776. Harvard University Press, Cambridge, MA
Price JM (1995) Tobacco in Atlantic trade. Ashgate Publishing, Brookfield
Rockoff H (1971) Money, prices and banks in the Jacksonian era. In: Fogel RW, Engerman SL (eds) The reinterpretation of American economic history. Harper & Row, New York, pp 448–458
Rothenberg WB (1979) A price index for rural Massachusetts, 1750–1855. J Econ Hist 39:975–1001

Rotwein E (ed) (1970) David Hume writing on economics. University of Wisconsin Press, Madison
Rousseau PL (2007) Backing, the quantity theory, and the transition to the U.S. dollar, 1723–1850. Am Econ Rev 97:266–270
Shell M (2013) Wampum and the origins of American money. University of Illinois Press, Chicago
Sheridan RB (1960) The British credit crisis of 1772 and the American colonies. J Econ Hist 20:161–186
West RC (1978) Money in the colonial American economy. Econ Inq 16:1–15
Wicker E (1985) Colonial monetary standards contrasted: evidence from the Seven Years' War. J Econ Hist 45:869–884

Privately Issued Money in the United States

18

Matthew Jaremski

Contents

Abstract

In recent years, there has been a revival of privately issued money. Due to the general lack of successful or even widely circulating private currency, it can be challenging to get a clear view of its efficiency using modern data. The U.S. historical period, however, offers a unique environment to examine the topic as private bank money made up a sizable portion of the money supply. Moreover, the period presents a wide range of regulation, including spans with and without the presence of a central bank or monetary authority. This chapter begins by highlighting the general history of privately issued money in the United States

M. Jaremski (✉)
Colgate University, New York, NY, USA

National Bureau of Economic Research (NBER), Cambridge, MA, USA
e-mail: mjaremski@colgate.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_20

from 1790 through its elimination in the 1930s. Topics include the rise of state bank notes, the switch to national bank notes, clearinghouse currency, the Aldrich-Vreeland emergency currency associations, and the decline of private currency. It then examines open topics in the literature and provides suggestions for study going forward.

Keywords
Private currency · Banks · Bank notes · Clearinghouses · Financial regulation

Introduction

The recent popularization of Bitcoin and other cryptocurrencies has renewed interest in privately issued money. Despite this rise, currency issued and controlled by governments and central banks still dominates the money supply. In fact, only a couple of private currencies have been issued in the United States since 1935 and most of those (e.g., BerkShares and Ithaca Hours) only circulate within a small geographic area. This, however, has not always been the case. Privately issued bank money made up a large portion of the nation's money supply before the Great Depression. Therefore, this historical period offers a unique environment to understand how private currency could operate in an economy alongside federal currency. This chapter examines the rise and fall of the various historical private currencies as well as the major open questions that are still being examined.

Though the American colonies and Continental Congress issued large amounts of paper money, their record was shabby at best. Most of the currency issued before 1790 circulated at steep discounts preredemption, and some were either repudiated or never fully redeemed. As a result, the nation's earliest politicians had reason to be suspicious of government issued currency, especially paper currency. The Constitution, therefore, stripped states of the ability to issue currency and the Treasury issued relatively little currency itself. Even as late as 1860, all government issued currency consisted of gold or silver coins. The coins were heavy, difficult to carry in large quantities, and often were worth more on the commodity market than as currency due to the bimetallic system which maintained fixed exchange rates of silver and gold until 1900. When Congress gave in and first issued paper money to support the Civil War, the number of notes was constrained. Congress began to allow paper currency to grow with the issue of silver certificates after 1878, but again it was for a larger purpose: appeasing Western political interest groups. It was not until after the creation of the Federal Reserve in 1914 that privately issued money was marginalized and eliminated.

The banking system was responsible for almost all of the privately issued money in the United States before the modern period. Bank notes were collateral-backed notes issued by individual banks and were used alongside federal currency in everyday transactions. The notes were likely a necessity of an undermonetized economy. Alternatively, private clearinghouse associations issued clearinghouse loan certificates (as well as checks and notes) as temporary currency in order to

provide emergency liquidity during panics. The issues were passed back and forth between banks in order to satisfy their clearing obligations but sometimes also circulated more generally. Lacking a central bank until 1914, these certificates allowed banks to protect their limited reserves and prevented them from having to liquidate their assets on an already depressed market.

These private currencies also experienced changes over time. Bank notes in the antebellum period were loosely regulated by individual states, whereas the National Banking Acts and surrounding legislation replaced state bank notes with federally regulated national bank notes. The Aldrich-Vreeland Act of 1908 supplemented the clearinghouse system by allowing currency associations to form and provide emergency currency during panics.

The promise for productive research on privately issued money is greater now than at any time in recent memory. This is because scholars are only beginning to unlock the potential of disaggregated data contained in national, state, and local government documents as well as the archives of contemporary newspapers and financial press. At the same time, the recent interest in understanding the issue and efficiency of private money has led to new models and questions that can be examined using historical data. There is little doubt that the combination of data and tools have created an unprecedented opportunity to uncover lessons from the past that can further our understanding of monetary conditions today.

Rise and Fall of Private Currency in the United States

The slow growth of federal currency provided an opening for privately issued money. Before the Great Depression, the banking system created many different types of currency. This section highlights the various types of private money that were issued in the United States using the most recent research on the period.

Rise of Bank Notes

Deposits during the antebellum period were not a high debt priority or in much demand. States, therefore, gave banks the right to monetize their assets. By promising to pay the full value in specie whenever demanded, banks split large illiquid assets into smaller liquid ones. The notes were denominated in dollars and resembled modern paper currency. Every note provided the denomination, the type of backing assets, and the name and location of the issuing bank. The note thus provided the relevant information for its use in either transactions or redemption. The issuing bank designed its notes uniquely and often depicted the bank president or concepts that were important to the bank (e.g., agriculture, manufacturing, and railroads).

Before 1837, each bank petitioned for a unique charter from its state legislature. The note requirements of these banks (called charter banks) varied greatly over place and time, but generally allowed notes to be backed by any type of asset. Starting

with Michigan in 1837, a series of free banking laws altered this trend by installing a well-defined set of requirements (Hammond 1957; Bodenhorn 2003). Seeking to protect noteholders, the new laws required free banks to purchase state or federal debt as security for each note. In a few states, other assets were allowed such as real estate in Michigan and New Jersey, railroad bonds in Minnesota, and slaves in Georgia (Rockoff 1972; Jaremski 2010). The bank then deposited those bonds with a state representative and received an equal value of bank notes in return. The state representative held the bonds as note collateral and only relinquished them when the bank returned an equal (or lesser) number of notes for destruction. If even a single request for note redemption was unmet by a bank, the state representative would close the bank and liquidate the collateral bonds to redeem any outstanding notes. The state only had to make sure the collateral was worth as much as the outstanding notes to prevent losses, and as a result, they placed little control on the number of notes a bank could issue. A bank could issue more notes as long as they had collateral.

An implication of these note requirements was that the security and acceptance of notes were tied to the government bond price level. If bond prices fell, then the real value of the bank's capital and backing declined while its notes were to be redeemed at the same value. Banks thus had to take losses redeeming notes at par or refuse to redeem the notes and shut their doors. State officials were tasked with making calls on banks for extra bonds when the collateral dropped below a set amount.

The notes easily circulated as currency in a local environment where banks (and their notes) were well known, but mechanisms arose to allow them to circulate more widely. Outside the local area, a merchant first had to determine that a note was legitimate, and then trust that it would be redeemed at face value when presented for repayment. Private note brokers alleviated these issues by identifying and purchasing notes at a percentage discount from their face value (see Dillistin (1949) and Gorton (1989) for a detailed description of brokers and reporters). Similar to modern currency exchangers, brokers advertised the notes they were willing to buy and sell and their corresponding prices. As the number of notes and brokers grew, note reporters began to bundle and publish the prevailing discounts in a city as a periodical. Note reporters also provided each note's originating location, physical description, and a description of any counterfeit notes. A merchant could thus verify a note's provenance and discount by examining the reporter's description, although, as described by Mihm (2007), reporters did not perfectly protect merchants from counterfeits. Identification was on small details, and notes were often faded and worn from use.

In addition to note brokers and reporters, another institution arose to allow notes to circulate more effectively. The Suffolk System (established in 1818) offered to clear the notes of any participating bank at par so long as it held funds in an account at the Suffolk Bank. By holding a large number of other banks' notes, the Suffolk Bank had the leverage needed to force a participating bank to hold a safe portfolio through the possibility of presenting its notes for immediate redemption in specie. The arrangement was so popular that the system allowed nearly all of the bank notes of New England to circulate very close to par. In 1858, the Bank of Mutual

Redemption took over the note clearing from the Suffolk Bank and operated until the mid-1860s.

As a result of these institutions, thousands of notes circulated in the economy by the 1860s. Bank notes dominated the amount of federal currency through 1840 and remained of similar size through 1860. Notes also exceeded bank deposits for most of the period. Banks likely preferred notes to deposits because they circulated as currency and did not quickly return to be cleared. As such, banks could issue large numbers of notes yet only keep a fraction of their value in the vault as cash. The public may also have preferred notes because deposits in most states were not subject to reserve requirements, while bank notes were subject to these (Mitchener and Jaremski 2015). About two-thirds of banks had more circulation than deposits in 1860. Putting this in the aggregate context, private bank notes made up at least a third of M1 before 1860.

Switch to National Bank Notes

The Civil War changed the nature of note issue. Tasked with financing the Union army, Secretary of the Treasury Salmon Chase pushed for national bank regulation after the initial federal bond issues led to the specie suspension of 1861. The resulting National Currency Act of 1863 and the National Banking Act of 1864 adopted the relatively stringent structure of New York's free banking law, and added even more requirements to prevent bank losses and create a market for national debt.

Most of these new requirements affected the issue of bank notes. First, they avoided free banking's attachment to risky state debt by requiring the use of Treasury bonds to back national bank notes at 90% of their value. This change raised the cost of issuing notes and reduced the number of notes that national banks could issue. Second, they placed a $300 million ceiling on national bank note circulation. The effect of the ceiling was felt almost immediately, but since the ceiling was raised in 1872 and eliminated in 1875, it did not have much of a long-term effect. Third, they required national banks to redeem their notes at par in a large city, meaning a merchant only had to find the bank's redeeming agent in the city to be assured of full payment.

In 1863 and 1864, most state banks had converted over to a national charter. Not satisfied with the remaining state banks, Congress passed a 10% tax on state bank notes in 1865. Striking a different tone than the other legislation, Selgin (2000) argues that the tax was intended to drive the remaining high-quality state bank notes out of circulation and limit money supply growth. Hammond (1957) posits the tax was responsible for the collapse of state banking. State banks with high note issues had one of three options: convert to a national bank charter to keep issuing notes, quickly attract a large number of deposits to supplement the loss of bank notes, or close their doors. As a result, over 225 state banks closed after the tax was implemented, leaving less than 250 state banks still in operation at the end of the decade (Jaremski 2013). The majority of closures occurred in the Midwest where population, capital requirements, and deposits were low. The pattern suggests that

many banks did not have the capital necessary to convert to a national charter or deposits necessary to remain a state bank when confronted with the tax.

As expected, the regulation dramatically changed the structure of bank notes in the country. State bank notes declined to zero and national bank notes rose up to take their place. That said, the growth of national bank notes was relatively stagnant and declined relative to GDP over the postbellum period. Growth remained slow even after national banks were allowed to issue bank notes up to the full value of their collateral in 1900.

The legislation also led to the elimination of bank note discounts. Because notes were backed by national debt and redeemable in major cities, merchants no longer had to worry about redemption. As a result, national bank notes circulated at par throughout the country, and most note brokers either started their own banks or went out of business.

Clearinghouse Certificates

Another type of private currency was issued by the banking system: clearinghouse loan certificates. New York City established the first U.S. clearinghouse association in 1853 to lower the costs of clearing checks and bank notes. Before its creation, banks had to send a representative to every other bank in the area in order to redeem their debt. Clearinghouses provided a central location and time to clear debt every day. Each member bank would then only have had to send a representative to the clearinghouse. Following New York's lead, banks in other large financial centers formed their own associations. As a result, clearinghouses had spread across major cities in nearly every state. (For more information on clearinghouses, see Cannon (1900), Timberlake (1984), Gorton (1985), and Gorton and Mullineaux (1987).)

The clearinghouse system evolved to do much more than clear bank debt. As the Panic of 1857 approached, depositors and country banks began to withdraw their deposits. The New York clearinghouse decided to create clearinghouse loan certificates and distribute them to qualified member banks. The certificates were backed by collateral from the receiving bank, but they were drawn on the clearinghouse, making them a joint-liability of the members. If a member defaulted, then the remaining members bore the loss in proportion to their capital stock. Member banks, therefore, could conserve their limited currency reserves by settling their clearinghouse obligations with certificates rather than federal currency or bank notes. Though they did not have a choice of whether to accept loan certificates, lending banks gained interest on the certificates. The banks could thus refrain from selling assets at fire sale prices. That said, the certificates were designed with large denominations to prevent them from circulating outside of the banking system (Gibbons 1859; Calomiris and Schweikart 1991).

Later panics pushed clearinghouses to innovate. The Panic of 1873 ushered in the use of certified checks. These checks differed from loan certificates because receiving banks did not have to give up collateral. The lack of collateral made them easier for banks to issue and they could keep safe assets on their balance sheets.

Clearinghouses created low-denomination clearinghouse notes during the Panic of 1893. Unlike clearinghouse loan certificates and certified checks, these notes could be issued to anyone and operated as a makeshift currency. Some clearinghouses issued notes as low as $0.25 to help make change during the Panic. The use of small denomination notes rose during the Panic of 1907 with some clearinghouses even printing blank checks made out to "bearer" so that banks could quickly fill in the needed amount (Andrew 1908; Cannon 1900).

By the creation of the Federal Reserve, clearinghouse associations were issuing large amounts of emergency funds during panics. For instance, the Comptroller of the Currency (1908, pp. 65–66) reported that at least $248 million in clearinghouse loan certificates were issued during 1907. The public also took notice of the actions of clearinghouses. Newspapers in New York and Chicago both reported on the stabilizing functions of clearinghouses during the Panic of 1907 (Moen and Tallman 2000). The clearinghouse thus had a variety of effects during panics. The loan certificates and certified checks impacted those inside of the banking system, clearinghouse notes helped those outside of the banking system, and the clearinghouse's actions provided a public signal of stability. Cannon (1900, p. 96) calls them "one of the finest examples the country has ever seen of the ability of the people when left to themselves to devise impromptu measures for their own relief."

Emergency Currency Under the Aldrich-Vreeland Act

While only temporary, the issues of clearinghouses were treated as a form of currency in their local areas, propping up the banking system until the market had settled. Hepburn (1924, p. 352) states: "This temporary currency performed so valuable a service. . .in moving crops and keeping business machinery in motion, that the government, after due deliberation, wisely forbore to prosecute [the issue of currency]." The National Banking Acts prevented national banks from issuing currency outside of national bank notes, whereas the state bank note tax prevented state banks from issuing currency without the sizable tax. The government thus could have prosecuted the clearinghouse members. The importance of clearinghouses' action during the Panic of 1907 even inspired Congress to take action. The resulting Aldrich-Vreeland Act of 1908 was a substantial step in financial regulation and created another form of private emergency currency.

First, the Act created the National Monetary Commission that was tasked to provide long-term recommendations on how to secure the banking system and prevent future panics. The Commission produced detailed reports on the banking experiences of the US, Canada, Mexico, and many European Countries. These findings ultimately led to the creation of the Federal Reserve System.

Second, the Act allowed the formation of currency associations that could issue emergency currency when invoked by the Secretary of the Treasury. Any group of 10 national banks or more with at least $10 million in combined capital could form a currency association. The associations lay dormant during most years and did not require the fees or supervision of clearinghouses. However, during a panic,

the members could exchange a broad range of securities (not just government bonds) for national bank notes at the association. The notes were largely indistinguishable from regular national bank notes and circulated as such. The emergency currency was taxed at an increasing rate over time to encourage banks to redeem their issues after the panic subsided.

Despite the period's susceptibility to panics, emergency currency under the Aldrich-Vreeland Act was only issued in 1914. The outbreak of World War I put intense pressure on the world's financial system as investors grew worried and pulled out of stocks. When the London stock exchange was shut down, investors started dumping securities on the New York Market. Treasury Secretary McAdoo shut down the New York Stock Exchange and authorized the use of emergency liquidity through currency associations. In total, currency associations issued $385 million of emergency liquidity in addition to the certificates issued by clearinghouses. As a result, the panic was largely aborted (Silber 2007; Jacobson and Tallman 2015).

Decline of Private Currency

The arrival of the Federal Reserve System in 1914 ushered in many changes to the banking and monetary systems. The Fed was intended to create a flexible currency and facilitate the payments system. With the exception of the clearinghouse issues, the nation's money supply was relatively inflexible before 1914. National bank notes required high quality collateral backing. Therefore during panics and seasonal spikes in demand, the money supply often contracted due to depositor withdraws rather than increasing to provide extra liquidity (Sprague 1910; Wicker 2000). Clearinghouse liquidity was often too little too late. The situation also prevented the government from exerting much control over the money supply.

To help solve these issues, the Federal Reserve began to issue federal currency by way of Federal Reserve Notes and Federal Reserve Bank Notes. This issue augmented the federal currency that was already in circulation. The Fed also created the discount window which allowed member banks to borrow short-term funds simply by exchanging accepted collateral regardless of whether a panic was occurring or not. This mechanism allowed the Fed to inject both seasonal and emergency liquidity into the financial system and banks did not have to wait for authorization by the Treasury or clearinghouse association. The Federal Reserve's founding was associated with reduced seasonal variability of interest rates and increased seasonal variability of lending (Miron 1986).

As national banks were forced to join the Fed, pressure started to build against national bank notes. Banks had some decision-making power to customize their notes before 1914, but began to be pushed towards standardized formats. In the mid-1930s, the bonds that could be used to collateralize national banks notes were retired and national bank notes were consolidated into Federal Reserve notes. The move gave the Federal Reserve more control and coordination over the physical portion of the money supply.

The creation of the Federal Reserve also eliminated the need for emergency currency from clearinghouses or currency associations. Banks no longer had to wait for others to authorize the issue of emergency currency. As a result, the Aldrich-Vreeland Act was allowed to expire and clearinghouses were integrated into the Fed's nation-wide clearing system.

Open Research Questions

This section provides a guide to the open questions and trends in research on historical privately issued money. The survey is selective in its choice of topics and literature coverage due to the wide range of possibilities for productive research in the area. Those seeking to pursue research in these topics would do well to start with the references discussed here and follow up with the literature cited in each of these sources.

Efficiency of Private Money

Given the modern discussions, the most relevant questions regarding the nation's historical experience with private currency concern its efficiency. The U.S. historical period presents a largely unique environment for studying this phenomenon. Not only did these periods contain a tighter and a looser regulatory approach to private bank notes, but they also include stretches when private currency circulated with and without the presence of a central bank. The historical narrative and data thus provide a testing environment for a large number of potential situations and questions.

The literature provides no clear answer as to whether private currency was efficient. Authors have remarked on both sides of subject. For instance, Cagan (1963, p. 20) writes that the United States "could not so easily have achieved its rapid industrial and commercial expansion during the second half of the nineteenth century with the fragmented currency system it had during the first half," whereas Selgin (2000, p. 600) writes the "because state banks of issue could have accommodated market demands that national banks failed to satisfy, consumers might have been better served had state banknotes been allowed to survive." Despite the question's importance, most studies have been theoretical or descriptive in nature. The majority of the econometric or data-driven studies focus either on (1) the determinants of bank note discounts or (2) the relationship between banks and economic growth.

Since over half of all banks closed before 1862, the issue of bank notes could have posed a very real problem for the economy, but the collateral of each note largely minimized the losses. For instance, Rockoff (1974), King (1983), and Rolnick and Weber (1988) find that aggregate losses ranged between 2 and 4 percent per year. At the same time, several banks left behind note losses over 60 percent. The liquidity benefits of notes might simply have outweighed these losses, but it is also possible that the pricing of bank notes shielded noteholders. If discounts

reflected the fundamentals of the issuing bank and the collateral backing, then noteholders would be informed of and compensated for their exposure. Banks might even have been pressured to reduce their risk in order to allow notes to circulate further. The bank note discount literature thus examines the extent that notes were dynamically priced.

Comparing reporters from Cincinnati, Cleveland, New York City, and Philadelphia, Gorton (1996, 1999) and Ales et al. (2008) highlight several stylized discount facts. First, "local" bank notes almost always traded at par, because holders could easily demand payment from a local bank. As banks were required to redeem their notes at par, a local discount could only exist if the bank was closed or suspended. Second, most, but not all, "foreign" notes circulated at a discount and discounts were asymmetric across locations. For example, Indiana notes were discounted in New York City (NYC) and Philadelphia, yet NYC notes were not always discounted in Indiana. Moreover, Indiana notes had different discounts in NYC than Philadelphia. Third, bank-level variation existed. Banks in the same city often traded at different discounts in other locations.

The empirical evidence shows that the discounts reflected the cost to redeem the note and the probability of default. Gorton (1996, 1999) shows that notes from distant banks received an excess discount. Bodenhorn (1998) shows that banks that were about to close often received a high discount, indicating that note brokers were pricing bank-specific risk into their discounts. Linking balance sheet data with discounts, Jaremski (2011) shows that bank discounts were correlated with the size of a bank's note circulation and the market value of its collateral assets. Therefore, even in the absence of modern regulations or reporting, discounts seemed to have mitigated risk. Further work by Selgin (2000) also argues that this discount process drove "bad" state bank notes out of circulation, leaving the relatively safe notes to widely circulate.

The literature makes a convincing case that private bank notes filled the currency needs of the developing nation and market mechanisms were able to increase their circulation and efficiency. However, it neither accounts for all the costs associated with bank notes nor makes a comparison with federal currency. For starters, the thousands of different bank notes introduced frictions to the market, as merchants had to subscribe to a bank note reporter or take time to consult a local note broker. This is not to mention the large number of counterfeit notes or the potential that a bank would close before a note reporter was updated. The important question, therefore, is whether the replacement of private bank notes with federal paper currency would have influenced economic growth in either direction. The recent collection of antebellum balance sheets by Weber (2008) as well as the collections of bank note discounts by Weber (2011) and Gorton and Weber (2011), however, now enable a much deeper study of the liquidity benefits of bank notes.

The analysis of finance-led growth is the second literature that relates to the efficiency of bank notes. Instead of focusing on bank notes specifically, it focuses on the effect of banks in general. Because notes made up a large portion of the bank's liabilities especially in the antebellum period, it is reasonable to connect the effect of banks to the effect of bank notes. Analyzing the period before 1850, Rousseau and Sylla (2005) argue that the "Federalist financial revolution" of the 1790s and its

system of state-chartered banks helped to set the nation on a path of modern economic growth. They examine the effect of the aggregate number of banks on business incorporations. The analysis, however, stops before the creation of the majority of the free banks. As Cagan was particularly critical of free bank notes, the analysis cannot directly disprove his arguments. Separating free and charter banks at the county-level after 1850, Jaremski and Rousseau (2013) find that free banking did not have an immediate impact on local economic growth, yet charter banks had positive effects on both manufacturing and urbanization. They argue that the high risk of free banks was responsible for the difference as short-lived banks were unlikely to provide many loans and individuals were unlikely to hold their bank notes. Another county-level study by Jaremski (2014) suggests that destruction of state banks caused by the National Banking Acts and the state bank note tax positively impacted manufacturing and urban development by encouraging the exit of free banks that were not growth promoting and concentrating new, more stable national banks in certain locations.

The evidence suggests that not all banks were growth promoting. However, none of these studies successfully test for the efficiency of note issue. First, the studies of growth look at whether a bank was present rather than the size and quality of its note issue. Second, the county-level results might not directly translate into nation-wide GDP growth. The studies identify local growth which means that one has to sum up all counties to gain an overall effect. Similarly, the studies focus on the results for manufacturing and urbanization rather than agriculture. As agriculture made up more than half of all economic activity, the analysis misses an important element of GDP. For instance, Jaremski (2014) focuses on the rise of the Manufacturing Belt as a result of the increased banks in certain areas, yet does not separately analyze the effect of decreased state banks in the agricultural regions of the South and Midwest. The removal of state bank notes might have been a particular blow to the economies of agricultural areas. The high capital requirements and restrictions on mortgage lending of national banks prevented many from being established in rural and agricultural areas, and without the ability to issue notes, state banks were slow to fill the gaps. Third, notes circulated beyond their place of issue suggesting that the effect of banks on the local community potentially does not represent the total effect of banks or bank notes.

The competition and eventual removal of national bank notes after the establishment of the Federal Reserve also has received little attention even though it has the clearest implications for the modern period. For instance, how did national bank notes respond to monetary policy and what was the effect of their removal? The Federal Reserve Era also has the benefit of having better data because the Fed began to collect more comprehensive information.

Note Issue Paradox

Another open section of the historical literature focuses on the size of bank note issues. Many authors including Friedman and Schwartz (1963) and Cagan (1965) have argued that national bank notes were under issued. The hypothesis states that

notes were more profitable than the typical loans funded by deposits and capital, and thus banks should have issued as many notes as allowed by the bonds in the system. Based on Friedman and Schwartz's numbers, note issue would have been profitable in every period except the period of temporarily high bond prices in the 1880s. That said, notes still only made up 28 percent of the maximum in 1900 when prices had returned to normal. It was not until after the creation of the Federal Reserve and World War I that note issue had risen to 80 percent of its maximum.

At the aggregate-level, many studies including Cagan and Schwartz (1991), Duggar and Rost (1969), Champ et al. (1992), and Wallace and Zhu (2007) have pointed toward redemption costs as a potential explanation. The explanation, though, is generally built around theoretical discussions or proxy measures, which do not provide a complete picture. Champ et al. (1992), for instance, focus on the amount of national bank note redemptions at the Treasury. They take the finding that half of outstanding notes were redeemed each year as evidence that national bank notes were not considered as valuable as federal currency for reserves. Diving deeper into the data, however, shows that most of the redemptions were either worn out notes, a bank's reduction of its own circulation, or the destruction of the notes of closed banks. Less than a fourth of all redemptions were for other reasons, suggesting the theory is not as powerful as the total number of redemptions might imply.

Another recent explanation for the aggregate picture is the issue of silver certificates. Created as a compromise to western farmers and miners, the government agreed to buy a certain amount of silver with notes redeemable in silver under the Bland Allison Act of 1878 and the Sherman Purchase Act of 1890. Proposed by Champ and Thomson (2006), the argument is that the increased circulation of federal currency reduced the need for bank notes. They find that the circulation of silver certificates took off after Congress started allowing the creation of one, two, and five dollar certificates in 1886. Bank note circulation also experienced its largest declines after 1885. The study, however, relies heavily on aggregate timing that is not exactly perfect. For instance, circulation began to rebound before silver certificates stopped being issued.

Another side of the literature has gained traction in explaining the puzzle by looking at disaggregated measures of bank note issue. James (1978) finds that the under issuance displays a regional pattern with southern and western banks issuing many fewer notes. He argues that the pattern is driven by the high profitability of loans in agricultural and developing areas. That said, the theory does not explain why banks in other regions did not continue to expand their note issues over time.

Calomiris and Mason (2008) disaggregate the data a step further by looking at bank-level data for 1880, 1890, and 1900. They argue that the aggregate picture obscures important dynamics. Specifically, they find an important role for the bank-specific limits put on national bank note issue. After accounting for the requirements, they find that many banks were at either the minimum or maximum notes allowed by their bond holdings and capital. Their regression analysis supports James' findings in that the opportunity cost of lending is an important determinant of whether a bank is at the maximum or minimum circulation allowed. They also find

that high bond prices are associated with lower circulation, and the lack of a bounce back after bond prices fell in the 1890s was due to a compositional shift in banks. New national banks issued few notes and closed national banks often issued the maximum amount. On the other hand, they find little evidence for the redemption cost hypothesis.

The evidence presented by Calomiris and Mason (2008) is convincing, but there are several questions left to be addressed. First and foremost, the authors do not analyze the choice of bonds and capital. Their approach measures note issue relative to the maximum amount of notes a bank could issue based on their bond holdings and capital stock. Banks could still have decided to purchase more bonds or increase their capital in order to take advantage of the profits of issuing more notes. This gap prevents a direct aggregation to the national-level and an immediate connection to studies such as Friedman and Schwartz (1963).

Second, only examining three years of data prevents a clear examination of the trends and dynamics behind bank decisions. Were the changes immediate with declines in bond price and were they at the behest of banks or note holders? The recent work by Champ and Thomson (2006) on silver certificates could also be integrated into a time-series or panel analysis. On a related note, the selected dates miss two critical events. Stopping in 1900 prevents a view of the effect of the Gold Standard Act of 1900 which increased the amount of notes that could be issued for every dollar of collateral bonds and the general profitability of note issue. The Act should have allowed banks to issue a large amount of new notes without buying any more bonds, but the bounce back was small relative to the potential. The period also misses the run up in notes to the near maximum during the early 1900s.

Third, the low bank note issues of Southern and Western banks (as well as urban banks) remains unexplained despite the inclusion of opportunity cost variables. The availability of data on weather, crop selection, or even crop prices, for instance, might help to better identify the effect of opportunity cost changes on bank note issue.

It is also worth noting that the note issue puzzle extends to the antebellum period. Dual studies by Bodenhorn and Haupert (1995, 1996) examine the profitability of bank note issue before 1860. The earlier study shows that New York free banks behaved similarly to national banks in that they issued notes much below their maximum. The authors argue that this finding makes sense because the National Banking Acts were based on the New York Free Banking Act. The later study argues that deposits and loan issue might have been an equally profitable alternative to note issue during the period. The parallels between the two literatures suggest that conclusions drawn on one period might also apply to the other.

Clearinghouse Mechanisms

The role of clearinghouses in panics is quite well known, but our knowledge is based almost entirely on the New York City clearinghouse. The New York City clearinghouse, however, might not be the best to examine as it issued clearinghouse loan

certificates rather than notes or checks. The descriptions of the role of notes and checks mainly come from the few bits and pieces of other clearinghouses described by Cannon (1900) and other authors. Moreover, New York City's role as the largest financial market and central reserve city might prevent conclusions from being generalized to every clearinghouse. The literature has also tended to focus on the bigger picture of clearinghouse actions rather than the detailed mechanics of clearinghouse currency. For instance, Jaremski (2015, 2016) examines the effect of the presence and membership of clearinghouses on bank portfolios and failures, but even though he examines many clearinghouses, he assumes them to be the same and does not account for the specific actions of each. Along these lines, there are several important gaps in the literature.

First and foremost, the literature lacks a general knowledge of the actions of clearinghouses outside of New York and Chicago. Even the Comptroller of the Currency only reported clearinghouse currency statistics for a handful of the major cities. As a result, we do not know how much currency was issued in total or the full extent of the variation across location. We also do not have enough information to calculate whether the issue of clearinghouse loan certificates had a larger or smaller effect than the issue of clearinghouse notes or checks. Specifically, we might expect clearinghouse loan certificates or notes to have a bigger effect on the banking system, but clearinghouse checks to have a bigger effect on the local economy. The increasing number of historical newspapers going up online should help shed light on these topics.

Second, the literature lacks knowledge of how clearinghouse currency was used, circulated, and retired. Most studies have looked at the effect of a bank receiving clearinghouse funds during a panic, yet they do not examine why a bank requested funds or even whether they used them. While all members were mutually liable for the issues, the actual decision on whether to request and use funds is quite important. For instance, in at least one panic, all member banks of the New York City clearinghouse were required to take out clearinghouse loan certificates even if they chose not to use them (Hoag 2016). On the back end, we also know relatively little of how clearinghouse currency was unwound. Each loan certificate, note, and check had to be redeemed in full by the specific bank that issued it. Therefore, before redemption, the bank had to track down all of its currency. There is at least one story of a bank simply declaring that they would not be paying any more interest in order to encourage others to redeem the certificates; however, this might be an exception rather than the rule.

Third, the literature is relatively silent on how regulators and bank examiners felt about clearinghouses and their members, let alone their currency. Did they fear the moral hazard implications of emergency liquidity or did they give members the benefit of the doubt due to their ability to gain liquidity during panics? The availability of national bank examiners reports in the National Archives would allow a much deeper insight into the actions taken by clearinghouses and the thoughts of examiners about those actions.

Conclusion

The early monetary history of the United States is a history of privately issued money. Despite the Treasury issuing federal currency, private money filled in the gaps and circulated across the nation. Seminal works such as Hammond (1957) and Friedman and Schwartz (1963) provide important insights about the broader trends driving this evolution, but inadequate data often precluded these works from moving beyond aggregate statistics and anecdotal evidence. The recent advent and continued collection of micro-level data, however, has opened the door to address key questions in the historical narrative that have never been fully answered. Due to the world's relatively lack of good examples of private currency, the historical period thus provides an important testing ground for understanding the efficiency of privately issued currency.

Cross-References

- ▶ Experiments with Paper Money
- ▶ Money and Prices in Colonial America
- ▶ The Evolution of US Monetary Policy
- ▶ The Historical Evolution of Central Banking

References

Ales L, Carapella F, Maziero P, Weber WE (2008) A model of banknote discounts. J Econ Theory 142:5–27

Andrew AP (1908) Hoarding in the panic of 1907. Q J Econ 22:290–299

Bodenhorn H (1998) Quis Custodiet Ipsos Custodes? East Econ J 24:7–24

Bodenhorn H (2003) State Banking in Early America: A new economic history. Oxford University Press, New York

Bodenhorn H, Haupert M (1995) Was there a note issue conundrum in the free banking era? J Money Credit Bank 27:702–712

Bodenhorn H, Haupert M (1996) The note issue paradox in the free banking era. J Econ Hist 56:687–693

Cagan P (1963) The first fifty years of the national banking system: an historical appraisal. In: Carson D (ed) Banking and monetary studies. Richard D. Irwin, Homewood, pp 15–42

Cagan P (1965) Determinants and effects of changes in the stock of money. Columbia University Press, New York, pp 1875–1970

Cagan P, Schwartz AJ (1991) The national bank note puzzle reinterpreted. J Money, Credit, Bank 23:293–307

Calomiris C, Mason JR (2008) Resolving the puzzle of the underissuance of national bank notes. Explor Econ Hist 45:327–355

Calomiris C, Schweikart L (1991) The panic of 1857: origins, transmission, and containment. J Econ Hist 51:807–834

Cannon J (1900) Clearing houses–their history, methods, and administration. D. Appleton and Company, New York
Champ B, Thomson JB (2006) National bank notes and silver certificates. Federal Reserve Bank of Cleveland Working Paper WP06–22
Champ B, Wallace N, Weber WE (1992) Resolving the national bank note paradox. Q Rev Fed Reserve Bank Minn 16:13–21
Dillistin WH (1949) Bank Note Reporters and Counterfeit Detectors, 1826–1866. Numismatic Notes and Monographs, no. 114. New York: American Numismatic Soc
Duggar JW, Rost R (1969) National bank note redemption and treasury cash. J Econ Hist 29:512–520
Friedman M, Schwartz AJ (1963) A Monetary History of the United States. Princeton University Press, Princeton
Gibbons J (1859) The banks of New York, their dealers, the clearinghouses, and the panic of 1857. D. Appleton and Company, New York
Gorton G (1985) Clearinghouses and the origin of central banking in the United States. J Econ Hist 45:277–283
Gorton G (1989) An introduction to Van Court's Bank note reporter and counterfeit detector. Unpublished manuscript, The Wharton School, University of Pennsylvania
Gorton G (1996) Reputation formation in early Bank note markets. J Polit Econ 104:346–397
Gorton G (1999) Pricing free Bank notes. J Monet Econ 44:33–64
Gorton G, Mullineaux D (1987) The joint production of confidence: Endogeneous regulation and nineteenth century commercial-Bank clearinghouses. J Money, Credit, Bank 19:457–468
Gorton G, Weber WE (2011) Quoted discounts on state bank note discounts in Philadelphia, 1832–1858. Research Department, Federal Reserve Bank of Minneapolis. http://cdm16030.contentdm.oclc.org/cdm/singleitem/collection/p16030coll4/id/3/rec/6
Hammond B (1957) Banks and politics in America: from the revolution to the civil war. Princeton University Press, Princeton
Hepburn A (1924) History of Coinage and Currency in the United State and the Perennial Contest for Sound Money. New York: Macmillan
Hoag C (2016) Clearinghouse loan certificates as interbank loans in the United States, 1860–1913. Financial History Review 23(3):303–324
Jacobson M, Tallman E (2015) Liquidity provision during the crisis of 1914: private and public sources. J Financ Stab 17:22–34
James J (1978) The conundrum of the low issue of national bank notes. J Polit Econ 84:359–367
Jaremski M (2010) Free Bank failures: risky bonds vs. undiversified portfolios. J Money, Credit, Bank 42:1565–1587
Jaremski M (2011) Bank-specific default risk in the pricing of Bank note discounts. J Econ Hist 71:950–975
Jaremski M (2013) State banks and the National Banking Acts: measuring the response to increased financial regulation, 1860–1870. J Money, Credit, Bank 45:379–399
Jaremski M (2014) National Banking's role in U.S. industrialization, 1850–1900. J Econ Hist 74:109–140
Jaremski M (2015) Clearinghouses as credit regulators before the fed? J Financ Stab 17:10–21
Jaremski M (2016) The (dis)advantages of clearinghouses before the fed. Unpublished working paper
Jaremski M, Rousseau P (2013) Banks, free banks, and economic growth. Econ Inq 51:1603–1621
King R (1983) On the economics of private money. J Monet Econ 12:127–158
Mitchener KJ, Jaremski M (2015) The Evolution of Bank Supervisory Institutions: Evidence from American States. J Econ Hist 75:819–859
Mihm S (2007) A nation of counterfeiters: Capitalists, Con Men and the Making of The United States. Harvard University Press, Cambridge
Miron JA (1986) Financial panics, the seasonality of the nominal interest rate, and the founding of the fed. Am Econ Rev 76:125–140

Moen J, Tallman E (2000) Clearinghouse membership and deposit contraction during the panic of 1907. J Econ Hist 60:145–163
Rockoff H (1972) The free banking era: a reexamination. Dissertations in American history (rev. Ph.D. diss., University of Chicago). New York; Arno Press
Rockoff H (1974) The free banking era: a reexamination. J Money, Credit, Bank 6:141–167
Rolnick AJ, Weber W (1988) Explaining the demand for free Bank notes. J Monet Econ 21:47–71
Rousseau PL, Sylla R (2005) Emerging financial markets and early US growth. Explor Econ Hist 42:1–26
Selgin G (2000) The suppression of state banknotes: a reconsideration. Econ Inq 38:600–615
Silber WL (2007) When Washington shut down wall street: the great financial crisis of 1914 and the origins of America's monetary supremacy. Princeton University Press, Princeton
Sprague OMW (1910) History of crises under the National Banking System. Senate Document 538, 61st Congress, 2d Session, National Monetary Commission, Washington, DC
Timberlake R (1984) The Central Banking Role of Clearinghouse Associations. J Money Credit Bank 16:1–15
Wallace N, Zhu T (2007) Float on a note. J Monet Econ 54:229–246
Weber W (2008) Balance sheets for U.S. Antebellum State Banks. Research Department, Federal Reserve Bank of Minneapolis. http://cdm16030.contentdm.oclc.org/cdm/about/collection/p16030coll5
Weber, Warren E (2011) Quoted discounts on state Bank notes in New York, Cincinnati, and Cleveland, selected dates, 1827–1858. Research Department, Federal Reserve Bank of Minneapolis. http://cdm16030.contentdm.oclc.org/cdm/singleitem/collection/p16030coll4/id/1/rec/5
Wicker E (2000) Banking panics of the gilded age. Cambridge University Press, Cambridge

Money, Prices, and Payments in Planned Economies

19

Michael Ellman

Contents

Abstract

This chapter is an account of the role of money in planned economies, which sets the subject in both theoretical and historical frameworks. Theoretically, money in planned economies is considered to fit into the state theory of money. Historically, the changing role of money, from hyperinflation to monetary stabilization, to the initial stages of the administrative-command economy, to monetary reform, to stable money, to chronic shortageflation, to acute shortageflation and loss of monetary control, is surveyed. Attention is paid both to the similarities (medium of

M. Ellman (✉)
Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

Cambridge Journal of Economics, Cambridge, UK
e-mail: ellman@xs4all.nl

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_40

exchange, unit of account) and to the differences (passive money, partial equivalence, role of the banking system) between money in planned economies and money in capitalist economies. It is concluded that cash money in planned economies was an instrument, analogous to propaganda or repression, to incentivize the population to work hard to achieve the goals of the national leadership, and that noncash money was an instrument to facilitate and check plan fulfillment.

Keywords
State theory · Passive money · Partial equivalence · Shortageflation · Chartalism

Introduction

Although the title of this chapter refers to planned economies in general, the chapter concentrates on the Soviet economy. This is justified because the USSR was the prototype for the planned economies (or administrative-command economies, as they were derisively called in the perestroika period) and the model from which they were derived. ("Administrative-command" is a literal translation from the Russian. A better English translation would be "bureaucratic-command." However, "administrative-command" has become the conventional translation and is therefore used in this chapter.) The role of money in administrative-command economies differed from that of money in capitalist economies because of a number of specific features explained below, such as the two monetary circuits, its partial equivalence, and the role of banks. Money in administrative-command economies had an important role, but that role only partly overlapped with that in capitalist economies and was partly specific to the administrative-command system. Hence, its role differed substantially from that analyzed in theories specific to the capitalist system, such as the Marxist or post-Keynesian. However, the role of money in administrative-command economies did fit into the state theory of money (Knapp 1924; Goodhart 1998) in that it was created by the state for purposes of the state.

Marxist-Leninist Theory

In the period between the two World Wars, currencies were expected by Western governments, businesses, and much of the public to have a gold "backing." Soviet money, however, was inconvertible and did not have a gold "backing." How did Marxist-Leninism explain this "strange" situation of a currency without any gold "backing"? The answer was given by Stalin in a speech of 1933 (published in English in Stalin 1955: 209). He explained that:

> The stability of Soviet money is ensured, first of all, by the vast quantity of goods held by the state and put into commodity circulation [i.e. sold to the public] at stable prices. What economist can deny that such a guarantee, which exists only in the USSR, is a more real guarantee of the stability of the currency than any gold reserve. Will the economists in

> capitalist countries ever understand that they are hopelessly muddled in their theory of a gold reserve as the "sole" guarantee of the stability of the currency?

Stalin was correct in thinking that economists in capitalist countries would eventually come to realize that stable currencies did not need gold "backing." However, well-informed economists outside the reach of Soviet state security would not only have denied but thought completely absurd Stalin's argument about the stability of Soviet money. At the time when Stalin spoke these words (January 1933), the economic situation in the USSR was marked by acute shortages of all goods, rationing of both consumer and producer goods, high inflation (in the market price of food products), and mass starvation (the famine of 1931–1934 killed about as many people as the Holocaust). There was no "vast quantity of commodities" available, and there were no "stable prices" at which all the population could buy food. Subsequently, in 1937, 1950, and 1961, official gold exchange rates were fixed for the rouble. Their only functions were to serve as an accounting device to convert foreign trade transactions into roubles and to attempt to raise the prestige of the rouble, but they had no influence on monetary policy or international trade. The rouble remained an inconvertible fiat currency. Gold (of which the USSR was a considerable producer) became something to accumulate as part of its reserves or to export as just another primary product.

It should be noted that the survival of money in a country officially building first socialism, then achieving it, and then building communism was something of an anomaly which required explanation. Socialists had always thought of socialism as a moneyless system, where consumer goods would be distributed on the basis of work done or of need, and production operate on the basis of direct exchange of goods for goods without the mediation of money. Nevertheless, buying and selling, and the money to make this possible, continued to exist in the USSR. Why was this? An answer to this question was given by Stalin in 1952 in his work *Economic problems of socialism in the USSR*. There he explained that it was only temporary and resulted from the existence and importance of the collective farm sector which required payment from the state for its deliveries of agricultural products. Stalin saw the survival of commodity circulation (i.e., the transfer of goods from organization to organization or person to person by means of sale and purchase for money) in the USSR as a necessary evil, resulting from the persistence of the non-state agricultural sector, which would disappear during the transition from socialism to communism.

> Of course, when instead of the two basic production sectors, the state sector and the collective-farm sector, there will be only one all-embracing production sector, with the right to dispose of all the consumer goods produced in the country, commodity circulation, with its "money economy", will disappear, as being an unnecessary element in the national economy. But so long as this is not the case, so long as the two basic production sectors remain, commodity production and commodity circulation must remain in force, as a necessary and very useful element in our system of national economy.

This may seem a rather strange argument, given the role of coercion in extracting resources from the collective farms in the Stalin period and the demonstrated

usefulness of money in the nonagricultural sector of the Soviet economy. However, it was fairly conventional in Soviet Marxism, and a very similar position had previously been advocated by economists in the 1930 discussion about the role of money in a planned economy (Davies 1989: 175–176). Stalin also suggested in *Economic problems of socialism in the USSR* that, in the case of a country with a very large foreign trade sector, such as Britain, foreign trade might be an argument for the persistence of commodities and money even after the victory of socialism there.

In the post-Stalin period, less attention was paid in the USSR to the views of the Marxist classics about money, and more attention was paid to the lessons of experience. The economic usefulness of money even in the period of "developed socialism" (i.e., the Brezhnev period) was officially recognized. As the 1972 Soviet Encyclopedia of Political Economy put it (Ekonomicheskaya 1972: 416):

> The whole historical experience of the USSR and other socialist countries proves the necessity of money for the planned management of the economy, for the building of socialism and the organisation of the socialist economy. The well-known attempt in 1921 to organise the exchange of industrial and agricultural goods without money was unsuccessful.

Soviet Monetary History

Soviet monetary history can be divided into seven periods: hyperinflation (1918–1922), monetary stabilization (1922–1925), the initial stages of the administrative-command economy (1926–1947), monetary reform (1947), stable money (1948–1965), chronic shortageflation (1966–1986) (i.e., the combination of shortages and inflation), and loss of monetary control (1987–1991).

Hyperinflation

The Bolsheviks inherited a situation of rapid inflation. Between the February (1917) and October (1917) Revolution's prices roughly quadrupled. Just between July (1917) and October (1917), the price of a meal with meat in a typical St. Petersburg factory canteen rose from 15 to 58 kopecks and without meat from 8 to 42 kopecks. This rapid inflation was caused by a combination of monetary financing of a large part of government expenditure (mainly the war) and a decline in the availability of goods resulting from strikes, disruptions to transport, and a decreasing willingness of the peasants to exchange their output for rapidly depreciating money. The failure of the provisional government to control the monetary situation was one of the reasons why the Bolsheviks were able to overthrow it.

The Bolsheviks converted this brief rapid inflation into a prolonged hyperinflation (1918–1922). This was caused by the civil war and the resulting inability of the new Soviet government to cover its military and civilian expenditures with taxation or borrowing and hence its recourse to paying its bills with newly printed paper money, the collapse in production, and the unwillingness of the peasants to

sell their products for almost worthless money. The hyperinflation was unintended. However, since it happened anyway, some Bolsheviks made a virtue of necessity and drew attention to its positive aspects. The emission of paper money played a very important role in financing the Soviet government. Even rapidly depreciating money enabled the state to finance part of its military expenditures and part of the food and fuel needs of the towns in Soviet territory. (The part it could not finance in this way was mainly filled by force – the conscription of soldiers and the requisitioning of agricultural goods in the villages.) As the Bolshevik economist Preobrazhensky (2006/[1920]: 600) put it in 1920, the printing press was "that machine-gun of the Commissariat of Finance which fired into the rear of the bourgeois system through its monetary system, turning the laws of monetary circulation of the bourgeois regime into a means of destroying that regime and a source of financing the revolution." Preobrazhensky also noted in this book that the large-scale monetary financing and the resulting rapid fall in the value of money was a kind of tax which, under Russian conditions, mainly fell on the peasantry. Other Bolsheviks treated the hyperinflation as a positive move in the direction of the socialist goal of the abolition of money. Keynes met Preobrazhensky at the Genoa International Monetary Conference of April–May 1922 and adopted Preobrazhensky's argument about the inflation tax in his writings of 1922–1923 (Kleiman 2000: 240–246).

Monetary Stabilization

The monetary stabilization of 1922–1925 ended the hyperinflation and introduced a socialist money that had, to a large extent, the properties of money in the capitalist system. The monetary stabilization began with the issue of a new bank note (the *chervonets*) equivalent to the prerevolutionary 10 rouble gold coin. Initially this new note circulated at the same time as the rapidly depreciating existing Soviet roubles (contemptuously known as *sovznaki*, literally "Soviet notes"), so that for a time there were two parallel legal moneys. According to the decree which authorized it, the issue of the new notes – which began at the end of 1922 – had to have at least 25% backing in precious metals and hard currency. This was intended to prevent the new notes being used for large-scale monetary financing of the deficit of the state budget and to ensure the public that this new money would not go the way of the existing Soviet money (the *sovznaki*). The issue of this new currency seems to have been the idea of the Russian banker V. Tarnovsky; it was implemented because of the disorganization caused by hyperinflation, the rapid decline in the yield of the inflation tax, and Lenin's personal support; and it was positively evaluated by Keynes when he heard about it prior to its implementation (Goland 2006: 160). It was politically acceptable because the Communists had adopted a New Economic Policy (NEP) in 1921 which largely restored a market economy, and the latter needed stable money. The political credit for it went to Sokol'nikov who had political responsibility for it as the People's Commissar for Finance (i.e., Minister of Finance).

In 1923 Gosbank (the state bank) also began issuing gold *chervonets* coins (mainly for use in foreign trade). The *chervonets* circulated besides the rapidly depreciating *sovznaki* until 1924, when Gosbank began withdrawing the *sovznaki* from circulation and established the *chervonets* as the sole legal tender, equal to 10 new roubles. Other denomination notes and coins for these new roubles were also issued. (Whereas the *chervonets* was issued by Gosbank, the smaller notes and coins were issued by the People's Commissariat of Finance.) The new stable Soviet rouble was a freely convertible currency both internally and externally. Internally, private citizens could, within limits, exchange their *chervontsy* in banks for foreign currency needed to travel abroad or to pay for parcels received from abroad. In addition, they could sell their *chervontsy* for foreign currency or gold at the informal markets in major cities, where agents of the currency and securities department of Gosbank traded in gold and foreign exchange (Goland 1994: 1258–1259). Their transactions stabilized the internal market value of the *chervonets*. Secondly, from the spring of 1924, the *chervonets* began to be quoted on a number of international foreign exchange markets, initially in the Baltic countries and later in Turkey, Persia, China, and Italy. (In Turkey, Persia, and China, the *chervonets* was used in informal cross-border trade.) Italy was the only major capitalist country where the *chervonets* was quoted. On these markets *chervontsy* were convertible. Stabilization transactions on these markets by Gosbank maintained the exchange rate.

The *chervonets* and other new roubles were issued by the state, and primarily served as a means of exchange, but also to some extent as a unit of account, a store of value, and a means of settling debts. As a means of exchange, the new money was used in the retail market and for transactions between enterprises and stimulated both. Its existence also encouraged peasants to expand their production and sell their surpluses on the market, thus increasing urban food supplies. It only partly served as a unit of account, since many economic calculations were conducted in prewar roubles. It only partly served as a store of value, since Soviet roubles were less trusted by much of the public than prewar gold coins or foreign currency. It was the legal means of settling debts and enabled personal, business, and tax debts to be settled. The success of the stabilization was largely a result of a drastic reduction in government expenditures (there was, e.g., a huge demobilization of the Red Army) and an increase in government revenue. The latter resulted from the increase in production after the end of fighting and famine and the government adoption of the NEP, an important constituent of which turned out to be the introduction of stable money. The reduction in expenditures and the increase in taxation resulting from the increase in production greatly improved the budget situation. In each of the 6 years 1924/1925–1929/1930, the budget – according to Soviet accounting – was in surplus (Nakamura 2017: 195). Hence there was no need to finance the budget with newly printed money. Considerable confidence was established in Soviet money. The tax in kind which the peasants had to pay out of their output was gradually replaced by monetary taxation, and this naturally increased the demand for money with which to pay the tax.

The Initial Stages of the Administrative-Command Economy (1926–1947)

The internal convertibility of the new Soviet roubles ended in the spring of 1926. The informal internal gold and foreign exchange markets were closed. There was a sharp increase in the restrictions on the access of private citizens to gold or foreign exchange. The official of the People's Commissariat of Finance who had very competently implemented the policy of trading on the internal markets to stabilize the value of the *chervonets* (Lev Volin – the head of the special section of the currency administration of the People's Commissariat of Finance) was arrested and subjected to extrajudicial execution. The external convertibility ended in the summer of 1926 with the end of official stabilization transactions on the foreign exchange markets of foreign countries and the ban on exports of *chervontsy*.

The end of the stable currency policy mainly reflected its incompatibility with a policy of rapid industrialization focused on branches of industry which did not produce consumer goods. Most Bolsheviks preferred the rapid development of heavy industry to the maintenance of financial stability (Carr 1958: Chap. 9). As Dzerzhinsky, candidate member of the Politburo, People's Commissar for Internal Affairs and chairman of the Supreme Council of the National Economy, explained in February 1926 (Dzerzhinsky 1926):

> Therefore when it is said that because of the shortage of resources we should halt our investment projects, or reduce them to a certain level, then I assert that I, as chairman of the Supreme Council of the National Economy, will struggle against such an opinion to the end as it is fundamentally incorrect.

Initially, the 1929 breakthrough, i.e., the decision to replace market relations by coercion (the aim to "liquidate the kulaks as a class," the mass deportation of peasant families, and the forced collectivization of agriculture) and economic planning (the 5-year plan) was expected by Communists to usher in a moneyless economy. In 1930 it was generally assumed in official circles that the new planned economy would be a moneyless one (Davies 1989: 173–178). Monetary relations, it was thought, would be replaced by relations in kind. However, experience showed that even in a planned economy, money had its uses. Besides its traditional roles of incentivizing workers, and allowing inter-enterprise transactions to be settled, it also had a useful role in regulating the national economy and in planning. Access to the archives of the Soviet state has shown that many key decisions (e.g., the allocation of resources to investment) were made in monetary terms (Gregory and Harrison 2005). This resulted from money's usefulness in aggregating the large number of goods produced in an economy (Garvey 1977: 42).

Within a few years, from about 1930 to 1931, by a process of trial and error, a new, nonmarket, monetary system had come into existence. This had the following features:

1. There was a distinction between two monetary circuits, noncash money (used for transactions between state enterprises) and cash money (used for paying

wages and in retail trade and also for inter-household and illegal transactions). Noncash money existed in the accounts of the state bank and was used for payments between state-owned enterprises in accordance with the plan. It is generally referred to as "passive money" (Peebles 1991: 40–41 and earlier authors there cited) because its transfer followed a plan decision and did not initiate it. The basic idea of the credit reform of 1930–1931 was that money flows should passively follow material resource flows instead of money determining the flow of material resources as in a market economy. The importance of noncash money could be seen as a move toward the Marxist goal of an economy in which money did not play a leading role.

Cash money, on the other hand, could be drawn by enterprises from Gosbank in accordance with their wages plans and Gosbank's balance of cash inflows and outflows and emission plan (i.e., Gosbank's plan for cash emission which was fixed by the political leadership as part of national economic planning). Once drawn, this cash money could be held by enterprises for future wage payments or issued to their workers (wages were paid in cash), who used it for purchases in retail trade or for other payments. (They could also use it for purchases of state bonds or to deposit in personal accounts with the state saving bank.) Cash money was active money. Its use initiated an economic transaction.

There was some leakage between these two circuits. For example, enterprises could get loans from the state bank to cover the working capital needed to fulfill the plan and then use some or all of these loans for wage payments. This was one of the causes of inflation. However, according to the calculations of Nakamura (2011: 1145), "the division between cash and noncash money was largely effective in the sense that increases in noncash money did not directly cause increases in cash money."

2. Both noncash money and cash money were not convertible into gold or foreign exchange. They were fiat money. Foreign trade was conducted in foreign currencies.
3. Noncash money was not officially convertible into producer goods (which were rationed) without official documents stating that the holder was entitled to receive a certain quantity of the goods concerned. Frequently, the acquisition of producer goods necessary for production required outside-the-plan barter. It might also require the intervention of local Party officials. Sometimes cash money could facilitate the unofficial acquisition of producer goods.
4. Cash money for most of this period (1929–1935 and 1941–1947) was not convertible into consumer goods at state prices without ration coupons. (In 1935–1940 consumer goods were available at state prices without ration coupons.)
5. Soviet money served as a means of exchange, but its role in this respect was severely limited by the rationing of producer and consumer goods, the allocation of many important consumer goods and services on a nonmonetary or heavily subsidized basis (housing, education, medical care, public transport), and the payment in kind, rather than money, to collective farmers for their work for the collective farms. Noncash money alone was insufficient to acquire producer

goods. Cash money's role as a medium of exchange was also limited by the existence of special (*torgsin*) shops (in 1931–1936) where consumer goods (such as bread) could be bought for gold or foreign exchange but not for Soviet money. Nevertheless, the existence of the free market for food (the collective farm market) from 1932, where trade took place in cash money, increased the importance of cash money as a means of exchange. Its role as a means of exchange was sufficient for cash money to serve as a way of incentivizing paid work for the state by the population. So important and successful was this that after World War II, a wage system was even introduced in the Gulag (the network of forced labor camps) with cash payments for good work.

6. Despite the limited role of Soviet money, as part of the bargaining between enterprises and the higher bureaucratic bodies, the managers of Soviet enterprises strove not only to be assigned easy plans but also to obtain additional money for their enterprises and for themselves (Harrison and Kim 2006). They used their information advantage to pad their cost and gross value of output plans and hence receive official noncash payments for their products in excess of their real costs as well as additions to their wages fund (which was often calculated as a percentage of gross value of output) and hence an official entitlement to extra cash money. They could sometimes also receive off-the-books cash payments from their customers. They used the additional money mainly to assist their enterprises but sometimes also for personal enrichment. Extra cash money was useful to provide side payments to their workers and to unofficial supply agents and to "siphon" goods off the retail market. "Siphoning" meant using the money of enterprises, and their barter possibilities, to obtain consumer goods intended for the retail market which were supposed to be bought only by state employees and collective farmers with cash money. The goods obtained by siphoning could be distributed to the enterprise's own workers (e.g., in canteens) or used for barter. This siphoning was widespread and increased shortages of consumer goods for the general public in state trade.
7. Soviet money in this period experienced rapid inflation. Between 1928 and 1947, producer goods prices seem to have risen only by a relatively modest 180% (although changes in the composition of output make calculating this figure theoretically and practically complex). However, state consumer goods prices rose by about 3800% and prices on the collective farm markets by about 11,400% (Holzman 1960). (The very high increases on the collective farm markets mainly took place during the war and reflected the extreme shortage of food products available for purchase then. Many urban inhabitants had to grow (some of) their own food.) The immediate cause of this inflation on the consumer goods market was the excess of the rate of growth of money incomes (at a time of very rapid growth of the urban labor force and an accommodating monetary policy) over the rate of growth of the availability of consumer goods (mainly food). Hence, looked at from a monetarist point of view, this inflation was a monetary phenomenon. However, looked at from a deeper perspective, it was simply a by-product of the Soviet industrialization strategy and the extreme food shortages during the war. The causes of inflation on the producer goods

market were twofold: first, the rapid industrialization program, which generated shortages of all the inputs that the program required and, secondly, the rapid increase in labor costs resulting from the increase in wages as a result of food price increases and competition between enterprises for workers and from the low labor productivity of the rural migrants to the towns under Soviet conditions of industrial production.

8. Soviet money in this period was not a very useful unit of account because of the rapid inflation. Hence much economic accounting was done not in current roubles but in "unchanged 1926–1927 prices." This was a problematic unit of account; since in the 1930s and 1940s, many new goods were produced that had not been produced in 1926–1927.
9. Because of the rapid inflation, Soviet money in this period was not a good store of value. Individuals who wanted to store value used tangible assets such as gold (coins or jewelry), foreign currency, rare books or works of art, privately owned housing (mainly traditional peasant housing), etc. There was a state savings bank, but rapid inflation made deposits in it unattractive in this period. Stocks and bonds of private companies did not exist since the economy was state-owned. State bonds were "sold" to the public but their purchase was obligatory, their interest and repayment terms unattractive, and in fact their issue was a form of taxation, economically similar to wage cuts. Because of low living standards and wartime destruction, few individuals were in a position to hoard, although the high prices on the collective farm market – especially during the war – did give some people the opportunity to acquire large quantities of rouble notes and looting in Germany at the end of World War II provided considerable opportunities for acquiring valuables.
10. One important function that Soviet noncash money did play in this period was "control by the rouble." This meant the ability of the monetary – and security – authorities to use data on the financial transactions of state-owned enterprises, as registered in their accounts with Gosbank, to check whether all their transactions had been in accordance with the state plans. This was analogous to the role of auditing and forensic accounting in capitalist economies.
11. Soviet money was a debt of the state. Noncash money was a promise to provide producer goods (subject to possession of the relevant plan documents) to the holder. Cash money was partly a direct promise to provide consumer goods at state prices (if the holder was also in possession of the relevant ration coupons when consumer goods were rationed). However, it was partly permission to buy goods at market prices on the free market for food products (i.e., the collective farm market). In this use it was a permit allowing the holder to attend these markets and to exchange it with other private individuals for food. This was not an obligation of the state. However, indirectly, even in this case, Soviet money was a debt of the state because the reason it was accepted by sellers was because it could be used to buy goods in state retail trade (or make tax payments).
12. Money played an important role in planning and policy-making. It was used for aggregating the diverse goods produced in the economy. For example, in July 1932, when Stalin decided to reduce investment (even in the military sector) to

stabilize the economy, he fixed the target for reduction in roubles – at least 500–700 million roubles (Stalin i Kaganovich 2001: 235). (He did not confine himself to monetary targets and also indicated some sectors which should be exempted from the cuts – light industry, ferrous metallurgy, and the railways.) Targeting a monetary goal was the only way of aggregating cuts in various sectors of the economy. This was a general procedure. Archival investigation has shown that: "The most important control figures decided by the Politburo, the annual investment plan and the defense budget, including the plan for military procurements from industry, were fixed in roubles. So were most supply quotas that were binding on production ministries and enterprises" (Harrison and Kim 2006: 7). Nevertheless, despite this Politburo planning in monetary units, there was also Politburo planning in physical units. For example, in 1930–1936 (Davies 2001: 68): "The Politburo approved annual and quarterly economic plans, which included many specific targets for particular products, given in physical terms." These physical plans had to be coordinated using material balances and input norms, the calculation and implementation of which was the core of the planning process.

In addition to national economic planning in both monetary and physical units, the national leadership engaged in micromanagement in physical terms. An example is the Magnitogorsk steel plant, one of the priority projects of the first five-fear plan. In January 1929 the Council of People's Commissars approved the project, with a planned capacity of 650,000 tons. In November 1929 this was raised to 1,100,000 t. Shortly thereafter it was raised to 1,600,000 t. In February 1930 the Politburo increased it to 2,000,000 t. These micromanagement decisions naturally led to major changes in the design of the plant. Another example is the late 1940s restoration of the factory Zaporozhstal' (Ellman 1992: 111). At a planning meeting presided over by Voznesensky, then a Politburo member and chairman of Gosplan (the State Planning Commission), the Minister of Finance kept remarking that there was no money for the project. "Voznesensky reacted angrily. He ordered the finance minister to leave the room and sit in the adjacent waiting room while the planners discussed the "real" issues (the availability of capacity, labour and materials)." For Voznesensky, as for Dzerzhinsky, monetary factors were subordinate to the plan. In addition, the top national leader made decisions about specific weapons systems.

Monetary Reform

In 1947 the Soviet government implemented a successful monetary reform which inaugurated a period of relative monetary stability. The reform was initiated by Stalin in 1943 (Zverev 1973: 231–233, 2012: 169–174) and finalized after long discussions within the People's Commissariat (later Ministry) of Finance and within the economic and political leadership. The aims of the reform were both economic and political. Economically, it had three goals: first, to reduce the cash money in the hands of the public so as to reduce inflationary pressures; secondly, to unify price

levels (prior to the reform there were three price levels, low ones for rationed goods, higher ones in "commercial" state shops, and still higher ones on the collective farm market); and, thirdly, to enable rationing of consumer goods to be ended. (The war was partly financed by issuing cash, as a result of which the cash money in circulation quadrupled, rising from 18.4 bn roubles on 1 June 1941 to 73.9 bn roubles at the end of 1945 (Denezhnoe 2010: 128)). Politically, it aimed to deprive peasants, traders, spivs, and black marketeers (known collectively in the USSR as "speculators") of the large sums some of them had been able to accumulate during and immediately after the war because of the then high prices on the collective farm markets. This political goal was widely supported by the population, since such wartime gains, at a time when millions of people were fighting (and often dying) for their country, or working for long hours in industry producing weapons for their country, sometimes under starvation conditions, were generally considered as unfair profiteering from the country's desperate situation. Both the economic and political goals of the reform were similar to those of other countries after the end of World War II. For example, they were similar to the goals of the Dutch 1945 monetary reform (Barendregt 1993).

The reform had two parts. First, existing bank notes (which ceased to be legal tender) were exchanged (in 1 week in December 1947 – except in some remote regions where 2 weeks were allowed) for new ones at a rate of 10 to 1 (a rate chosen by Stalin personally and higher than the Minister of Finance thought necessary on purely economic grounds). Secondly, small balances (up to 3,000 roubles) in the state savings bank and state bank were exchanged for balances in new roubles at a rate of one to one, and balances in excess of 3,000 roubles into balances in new roubles at discounts of only $33\frac{1}{3}\%$ to 50%, depending on their size, compared with 90% for cash. This much more favorable treatment of deposits in the state savings bank than cash holdings was a political move to discriminate against "speculators" and in favor of those with money the origin of which was not problematic. At 1 January 1947, 84% of savings bank accounts held less than 3,000 roubles, but they represented only 21% of the total value of savings bank deposits. Forty-six percent of the value of the accounts was in accounts with more than 10,000 roubles. However, this large sum was held by only 5% of accountholders (Denezhnaya 2010: 580). Hence, the decision to convert accounts with less than 3,000 roubles on a one-to-one basis, and accounts with more than 10,000 roubles at a much less attractive rate, protected the assets of the overwhelming majority of depositors and was simultaneously very redistributive. Most state loans were exchanged on a three-for-one basis (i.e., the holders lost two-thirds of their nominal value), and their interest rates were reduced. The combined result was to reduce the nominal value of the internal state debt by about 63% and save the state five billion roubles per annum in interest and prizes. About 44% of the gains to Gosbank's balance sheet by reducing its liabilities to the public were used to cancel old Gosbank loans to the Ministry of Finance (Denezhnaya 2010: 633). Nominal wages and other incomes remained unchanged. Simultaneously, the rationing system

(for consumer goods) was ended. This was made possible by the monetary reform, since after it there was much less money available to buy unrationed goods. Without it, the end of rationing would have meant extreme shortages in state trade. In addition, prices in state "commercial" trade were drastically reduced, and the price of bread reduced below the previous ration price. Hence real wages rose significantly (where bread and other goods were actually available at the new prices and purchases of government bonds did not increase). However, some social benefits were reduced (payments to mothers with several children) or abolished (payments to families of soldiers killed in the war). This was a help to the state budget but naturally a blow to the former recipients of these benefits.

Although the reform was supposed to be strictly secret until it was officially announced, to prevent people with large amounts of cash from spending it or depositing it in the savings bank, this proved impossible to realize. Even in the dictatorial Stalinist state, news of the impending reform leaked out. Rumors began spreading quite widely. This led to unusually high retail trade volumes and sharp market price increases in Moscow at the end of November – more than 2 weeks before the beginning of the reform. Similar events took place in other cities, e.g., in Kyiv. Some officials took advantage of the reform to enrich themselves (Denezhnaya 2010: part III). Some trade officials were able to record sales of goods as having taken place in old roubles, when in fact they took place in new roubles and pocket the difference. Some financial officials accepted bank deposits in old roubles after the deadline for this, in return for a payment from the depositor. Other officials withdrew large sums from the bank, divided them into numerous relatively small amounts, and then deposited them in the names of relatives, friends, or subordinates. In this way they hoped to protect their money from confiscation by taking advantage of the more favorable exchange rates for small deposits. By 1 May 1948, the Ministry for Internal Affairs had uncovered about 100,000,000 roubles of these illegal deposits. Other officials used their authority to exchange old cash roubles for new ones at a favorable rate. As a result of his failure to preserve secrecy and prevent the misuse of their position by numerous officials, the Minister of Finance (Zverev) was demoted to Deputy Minister (Sitnin 1993: 14).

Despite the usual misuse of their positions by officials, and the continued shortages, particularly in provincial towns (in some places there were initially long queues for bread and even less bread sold to each customer than the former bread ration – Denezhnaya 2010: 664), the reform can be considered a success. It was an example of a sensible and well-executed Stalinist economic policy. It greatly reduced disequilibrium on the consumer goods market and enabled rationing to be abolished – much earlier than in some other European countries.

Analogous monetary reforms were implemented in Eastern Europe after World War II, in North Korea in 1947, in China in 1949–1951 and 1955, and in Vietnam in 1985. There was also a monetary reform in North Korea in 2009–2010 (Frank 2010). However, unlike its Soviet predecessor, this did not lead to the abolition of rationing but to its revival.

Stable Money (1948–1965)

During this period the value of money was not entirely stable. In the late Stalinist period, some prices were reduced, thus leading to a deflationary increase in the value of money. During this period there were frequent shortages of particular goods in particular places, which could be interpreted as signs of suppressed inflation. Nevertheless, by past Soviet standards and by international standards, this was a period of relative monetary stability. The leaders paid attention to the budget and the desirability of balancing it or running a surplus. For example, in December 1948 the Minister of Finance (Kosygin) sent a report to Stalin outlining the tax increases and expenditure reductions necessary to ensure that the 1949 budget would be in surplus (Denezhnaya 2010: 493–500). It would have won the warm approval of the IMF, the ECB, and the creditor countries in the EMU. Its proposal to double rents (frozen at the 1926 level) would no doubt have struck them as a positive "structural reform."

Budget deficits were met by borrowing or issuing additional money. In the Stalin period, borrowing was a considerable source of budget income (e.g., it accounted for 6.7% of the income of the state budget in 1947 and 5.8% in 1948 – Denezhnaya 2010: 630). This borrowing was treated at the time as income, although Western accountants would have treated it as a means of financing the deficit. As a result of this Soviet accounting, the 1947 budget showed a surplus although using Western accounting, it showed a deficit. Unlike the bonds sold in capitalist countries mainly to banks and other financial institutions, this borrowing was from the general public (about 70 million people subscribed to the state loan in 1946), and purchase of the state bonds was de facto compulsory. Repayment only began in the mid-1970s, when many of their original owners were already dead. Compulsory subscription to these bonds was ended in 1958 as part of Khrushchev's destalinization policy. This was de facto a wage increase. However, the maturity of the outstanding bonds was extended by 20 years which reduced the state's debt servicing costs but was a loss for their holders. Some government bonds continued to be sold on a voluntary basis, e.g., prize bonds, which paid their lucky owners prizes and were sold for cash by the state savings bank. In addition, up to 1962 the budget continued to count the regular annual increases in deposits with the state savings bank as income rather than partial financing of the deficit. (In 1963 the savings bank, which had been part of the Ministry of Finance, became part of Gosbank.) Another source of finance was Gosbank loans to the Ministry of Finance and purchase of Ministry of Finance bonds. (These were also counted as income rather than financing of the deficit.)

An additional source of financing the budget deficit was monetary emission, both cash and noncash. Cash emission (Denezhnoe 2010: 113–114) was positive in every year from 1948 to 1991 (except for 1960 – the year in which it was announced that on 1 January 1961 existing roubles would be exchanged for new ones at a rate of 10 for 1, with a corresponding adjustment of all wages and prices). Seen from an operational perspective, cash emissions were the difference between the cash provided by Gosbank to pay wages and the cash received by Gosbank, mainly from sales in state retail trade. Looked at from the perspective of the balance of the money incomes and expenditures of the population (the main monetary planning

instrument), cash emission resulted from an excess of money incomes (less noncash saving) over state retail trade (and other cash payments to the state). Cash emission in excess of that required to service the increasing volumes of goods available in state retail trade (and other payments to the state) naturally worsened shortages in state trade, where prices were fixed by the state, independent of supply and demand, and fairly stable.

In the Stalin-Khrushchev period, Gosbank was sometimes only able to adhere to its quarterly emission plans by refusing to allow some enterprises to withdraw the cash they needed to pay wages. This led to delays in paying wages and was a serious problem for the workers affected. For the state it was a necessary cost of preserving monetary discipline. In the Brezhnev period, priority was given to paying wages on time at the cost of monetary discipline (Sitnin 1993: 18). This was part of the general slackening of discipline which characterized the Brezhnev era.

Chronic Shortageflation (1966–1986)

The years 1966–1986 were marked by a worsening of the macroeconomic situation. Growth rates of output declined; military expenditures substantially increased; food subsidies rapidly increased; the siphoning of consumer goods by enterprises increased (Kim 2002: 112–113); the incomes of the population often exceeded plan levels (Denezhnoe 2010: 576); and a substantial share of the budget was financed by loans from Gosbank to the Ministry of Finance and government bond purchases by Gosbank (Nakamura forthcoming). As far as military expenditures are concerned, there are no reliable data. However, as the Minister of Finance pointed out in 1989 (Pavlov 1989):

> It would be naive to think that, let us say, the budget of the USA is in deficit because of huge military expenditures, but that is not the case for us. We always said that our total military expenditures were 17–18 billion roubles per year. However, it is clear that military parity could only be maintained by equal efforts. For a weaker economy this requires much greater efforts than for an economy with greater productive potential.

The excess demand these phenomena generated manifested itself partly in price rises, mainly on the free market for food products. However, since state retail prices were fairly stable, the excess of demand over supply on the state retail market manifested itself primarily in increasing shortages but also in growing second economy activity (Ellman 2014: 39–40; Nuti 1986; Grossman 1986). The economy suffered from chronic shortageflation (Kolodko and McMahon 1987). The same macroeconomic disequilibrium under market economy conditions would have led to open inflation. The shortageflation was an irritant for the general population, but senior officials were protected from it by special facilities which provided quality goods at low prices. Nevertheless, total inflation in these years (allowing both for price increases and for increases in shortages) was modest by contemporary Western standards (in the UK in 1970–1979 inflation averaged 12.5% p.a. and in

1981 reached 16%). According to calculations by the USSR bank institute, total inflation in 1968–1987 averaged just 4% p.a. (Bazhan 1990: 15). It was chronic but not yet acute.

Loss of Monetary Control (1987–1991)

The loss of monetary control was a result of a combination of political reforms, economic reforms, ill-conceived policies, political leaders with no understanding of economics, the political constraints limiting the actions of the leadership, the actions of the highest political leader in the country, an adverse shift in the terms of trade, and secrecy. Gorbachev was committed to an increasingly radical reform of the country (perestroika) and paid little attention to calls for monetary stabilization because he considered that perestroika was a revolution and, that in a revolution, "stability" is the slogan of the counterrevolutionary forces (Ellman 2000).

Gorbachev's 1988 removal of the Party from the economy was very destabilizing, since it was the authority of the Party at all levels that cut through bureaucratic tangles and restricted the role of disloyal behavior by subordinates. Its absence led to a widespread disregard for orders from above and the national interest. Political liberalization also led to the leadership's attempt to increase its popularity by increasing social expenditures. The economic reform of 1987, especially when implemented under conditions of political reform, was also destabilizing. Its increased independence for state enterprises led to increased wage payments, a decline in the share of enterprise profits going into the state budget, increased prices, and increased leakage between the two monetary circuits. The decline in the siphoning of goods from the retail market by state enterprises, which had been taking place for several years, seems to have reversed, and siphoning seems to have increased in 1988 and 1989 (Kim 2002: 113). All these factors naturally increased shortageflation.

Important ill-conceived policies concerned agriculture. A destabilizing policy introduced as a by-product of the 1965 economic reform was food subsidies. They rose 15-fold in 1965–1984, burdening the budget and worsening food shortages. In 1980–1982, before Gorbachev became the supreme leader, he was the person within the national leadership responsible for agriculture (he moved to the national leadership in Moscow from a position as head of a largely agricultural region). As such, he was the head of the national agricultural lobby, and the food program he advocated in that capacity in 1980–1982 was an expensive one which combined increases in procurement prices (paid by the state to the farms) with substantial state investment in the agro-industrial sector. The Ministry of Finance opposed the former and Gosplan the latter. When holding the highest office in the land, not only did he increase expenditure on the traditional food subsidies, but he rapidly increased the additional subsidies paid (from 1983) to economically weaker farms. Hence, according to the estimates of Kim (2002: 108–110), by the late 1980s, just a few years after he had taken over the supreme leadership, agricultural and food subsidies amounted to about 20% of budget expenditure and about 10% of GNP! Although

some officials argued for reducing the subsidies by increasing food prices, the leadership did not do this till very late because of a major political constraint. They were afraid of the political consequences in view of events in Novocherkassk in 1962 (Kozlov 2002: Chaps. 12 and 13). In June 1962 a national increase in meat prices led in Novocherkassk – a town in the North Caucasus – to a strike, blocking of the railway line, riots, and demonstrations under the slogan "Meat, Butter and an Increase in Wages." Top Moscow leaders arrived to pacify the demonstrators but to no avail. The restoration of order required troops to shoot into a crowd of unarmed demonstrators. In all 23 people were killed and many more injured. A show trial held 2 months later sentenced seven participants to death and others to prison sentences of between 10 and 15 years. Former finance minister Pavlov (1995: 97) ironically described the Soviet leaders as suffering from a "Novocherkassk syndrome," i.e., an unwillingness to raise food prices out of fear of precipitating strikes and demonstrations. The Soviet leaders may also have been influenced by events in neighboring Poland. In 1970 food price increases in Poland led to demonstrations and strikes. They were dealt with by force. The army fatally shot about 40 workers and wounded about a thousand. As a result, the Party leader (Gomulka) was forced from office. In 1980 a similar episode took place. Food price increases led to discontent and protests, and the Party leader (Gierek) was forced from office.

The food queues and shortages were so conspicuous that it seemed obvious that more investment in agriculture, grain imports, and financial support for farms were necessary to increase food output. However, there was an alternative price and output policy that would have been much better for the budget and balance of payments. This would have been a structural price and output reform which reduced the demand for food by increasing the price (and/or availability) of nonfood products, such as housing, durables, and services. This was repeatedly advocated by Gosbank officials but was ignored. A similar argument for Poland was made by the Polish economist Podkaminer (1982, 1988) but that was also ignored. Podkaminer's argument was confirmed by relative price movements in Poland after the price liberalization of 1989–1990 (Bell and Rostowski 1995).

Another ill-considered policy was the big increase in investment in the engineering sector in 1985–1987. This lengthened the building time for many projects, worsened the balance of payments, and worsened shortages of producer and consumer goods (because it increased the demand for producer goods and increased incomes generated by activities that did not produce consumer goods). Military expenditures also seem to have increased in 1985–1988.

An additional destabilizing policy was the anti-alcohol policy. This led to a decline in the state's receipts from alcohol sales and a big increase in the sale of illicitly distilled spirits (known in the US prohibition period as moonshine and in the USSR as *samogon*). This transferred a substantial part of what would have been state budget income into additional demand on the retail market (via the purchases of the sellers and producers of *samogon*). This naturally worsened both the budget deficit and shortages.

The USSR was hit in the mid-1980s by a sharp fall in the world market prices for its chief exports – oil and natural gas. This badly hit the revenue from foreign trade

which the state budget received from the difference between world market prices and domestic prices for these products. Revenue from foreign trade fell from 18.2% of budget revenue in 1985 to only 13.6% in 1989 (Kim 2002: 117). The decline in the terms of trade was exacerbated in 1988–1991 by a sharp decline in the quantity of oil production and exports.

The secrecy concerning the real situation of the budget (not only was the published budget at the beginning of perestroika in surplus, but the actual situation was not known even to most Politburo members) delayed political attention to its state. This same secrecy delayed serious discussion of the consequences of prolonged and growing monetary financing of the deficit.

This combination of increased expenditures and reduced revenue naturally resulted in a sharp increase in the budget deficit. Data on this is set out in Table 1.

The huge and rapid increase in the budget deficit shown in Table 1 was partly financed by monetary emission. Data on cash emission is set out in Table 2.

Another source of finance was borrowing from Gosbank. Data on internal borrowing is set out in Table 3.

The USSR also tried to balance the budget by foreign borrowing, but the amounts raised in this way were insufficient, and servicing external debts became an increasing burden (leading to default in 1991). At the same time, production began falling, as a result of Gorbachev's economic and political reforms and the conflicts between the Union Republics (in particular Russia) and the central government. GDP fell in 1990. In 1991 Russian GDP – according to the revised World Bank-Goskomstat estimate in Russian (1995: 94) – fell by 5%. Shortages worsened dramatically and market prices rose rapidly. To deal with the shortages (partly resulting from hoarding) in 1990, local rationing of some food products was widely introduced. By the end of 1990, some towns were even unable to supply the amounts to which ration-coupon holders were entitled. By March 1991 bread rationing existed in many regions (Gaidar 2006: 329). By the autumn of 1991, the USSR had moved from the stage of chronic shortageflation to acute shortageflation. Shortages and queues

Table 1 USSR budget deficit as % of GNP

Year	Deficit
1985	1.8
1988	9.2
1989	8.6
1990	4.1[a]
1991	15.0[b]

Source: For 1985–1990 these are the official figures in Narkhoz 1990: 16 and 5. For 1991 the figure is from Abalkin (1992: 169). There are a variety of figures for the budget deficit in this period, but the sharp increase in the early perestroika period and the crisis in 1991 are generally agreed. According to the IMF et al. (1991: 55), the budget deficit as a percentage of the GDP rose from 2.4% in 1985 to 11.0% in 1988. For a discussion of the financial and monetary situations in this period, see Khanin (2010: Chap. 6)

[a]According to Abalkin (1992: 169), the deficit in 1990 as percent of GNP was 5.8%

[b]Estimate

Table 2 Net cash emission in the USSR (billion roubles)

Year/s	Net cash emission
1975–1984 (annual average)	3.3
1987	5.9
1988	11.8
1989	18.3
1990	28.4
1991	120.7

Source: For all years up to and including 1990, Denezhnoe (2010: 114). For 1991, Abalkin (1992: 175). In that year, cash money in circulation rose by 92% Abalkin (1992: 175)

Table 3 State internal debt as percent of GNP

1985	18.2
1986	20.3
1987	26.6
1988	35.6
1989	43.1
1990	56.6

This debt was probably mainly borrowing from Gosbank (but some of it was probably government bonds held by the public). Part of it consisted of deposits in the savings bank lent by Gosbank to the Ministry of Finance. (Almost all the assets of the savings bank consisted of deposits with Gosbank.) A large part of the rest was just monetary financing
Source: Narkhoz 1990: 19

were very extensive everywhere. There were even bread shortages in Moscow. The increase in Russian retail prices in 1991 was officially estimated to be about 160% but may well have been more. In response to the loss of monetary (and political) control, the black market value of the rouble rapidly declined. In 1990 it fell from about 15 to the US dollar at the beginning of the year to about 24 at the end of the year. By the end of 1991, it had fallen to about 140 to the US dollar.

It is a mistake to regard the crisis as exclusively a monetary crisis or a fiscal crisis or a political crisis, since it was simultaneously all three. On the surface it seemed to be a monetary crisis, and indeed the immense monetary emission did destabilize the economy and society. However, the monetary crisis was the result of a fiscal crisis – the inability of the state to finance its expenditures in a noninflationary way. In turn, the fiscal crisis resulted from the weakness of the state, the destabilizing actions of the country's leader, the decline in its export earnings, the unreliability of its officials, the actions of its republics, and the unwillingness to use force.

In 1991, the monetary crisis, the fiscal crisis, and the collapse of the state were intertwined phenomena. The monetary crisis was both a cause and a result of state collapse. It caused it by contributing to the acute shortageflation which exacerbated public dissatisfaction with the status quo, undermined support for the USSR, and strengthened separatist sentiment in Ukraine and Russia. At the same time, it resulted from the inability of the state to collect sufficient taxes (and savings by the population and foreign loans) to balance its budget. The combined result of economic

reform and political liberalization was to reduce the share of enterprise profits going to the state budget from 56% in 1985 to 36% in 1990 (Narkhoz 1990: 21). The state's revenues were frequently intercepted by the republics which wanted them for themselves. The state's tax and customs officials were increasingly unable, and often unwilling, to collect all the taxes due or to control illicit international trade and the profits it produced. In the general chaos, a large number of officials became more concerned with their own income and assets than with serving the state. All this reduced budget income, which increased the need for monetary emission and contributed to state weakening and ultimately to state collapse.

Despite an attempt to raise food prices in 1990, a partial monetary reform in 1991, and an increase in state retail prices in 1991, the monetary system was not stabilized, and the monetary system was not brought under control. In fact, it was out of control. This was one of the factors which led to the collapse of the USSR, alongside the following: the removal of the Party from the economy and other political and economic reforms; Gorbachev's unwillingness to use force to stabilize the situation; the failure of the August 1991 attempt to halt the weakening of the USSR as a unitary state; the overwhelming majority for independence in the Ukrainian referendum; and the desire of the Russian leadership for independence. The restoration of monetary control in the successor states of the USSR was a difficult task which was an important part of their state-building and in some of them took years.

Money as a Partial Equivalent

Capitalist money is what Marx (1867) termed a universal equivalent. It can be used to buy a very wide range of goods and services and also assets such as businesses, real estate, and securities. Socialist money was only a partial equivalent. Some important goods and services (such as education, medical care, housing, and public transport) were allocated free or for nominal amounts. (This was an important reason for the shortages of the products that had to be paid for.) An important example of the different roles of money under capitalism and in the administrative-command economy concerns higher education. Under (current Anglo-American) capitalism, students finish their studies with a large monetary debt, which they are supposed to repay from their future earnings. In the administrative-command system, they finished with a nonmonetary debt to society, which they were supposed to repay by working for a period wherever they were allocated.

The ability of noncash money to buy producer goods was limited by their rationing. The ability of cash money to buy consumer goods was limited by the distribution of many goods and services on a nonmonetary or heavily subsidized basis, closed distribution (the allocation of consumer goods – sometimes at reduced prices or for free – to a privileged group rather than the general public), and the state monopoly of foreign trade which restricted the ability to obtain foreign consumer goods. It was also limited by the widespread shortages of consumer goods, which were a characteristic feature of the administrative-command system (Kornai 1980), although their extent fluctuated sharply over time and between places. All land, all

productive enterprises, and nearly all urban housing were legally or de facto owned directly or indirectly by the state and were not for sale.

This partial equivalence of socialist money was a very important feature and an important difference between socialist money and capitalist money. Some Marxists considered that because it was not a universal equivalent, Soviet money was not "real money." Similarly, some non-Marxist economists considered that noncash money, because it was passive, was not "real money."

Banking

The administrative-command economy had banks, mainly the state bank (Gosbank – sometimes known in the Western literature as the Monobank to allow for the non-Soviet administrative-command economies), the state savings bank (often a part of Gosbank), and domestic and overseas foreign-trade banks. The four major functions of Gosbank were to act as the state's bank, to provide credit to state enterprises to enable them to finance working capital (work in progress, stocks, purchased inputs, goods delivered but not yet paid for), to provide a payments system for state enterprises, and to act as a control mechanism. As the state's bank, Gosbank held the balances of budgetary organizations and made loans to the Ministry of Finance as required. As the bank for enterprises, Gosbank held deposits by enterprises, and made loans (mainly short-term) to them, subject to the appropriate plan and other documents. All state enterprises held accounts with Gosbank, and payments were made to and from these accounts and cleared by Gosbank. As a control mechanism, Gosbank checked the financial transactions of state enterprises to ensure that they were in accordance with the plan.

Banks in administrative-command economies were not free to create money in accordance with the wishes of potential borrowers (other than the Ministry of Finance) or their own lending decisions. As Schumpeter long ago observed (Bertocco 2007), in a socialist economy, the crucial role of the banking system in providing finance for entrepreneurial investment is replaced by the central authority (in the Soviet case, the Party leadership) that decides on the allocation of investment. Investment in the USSR was mainly financed by budgetary grants. Gosbank was expected to check that these grants were only used to finance planned investment. In general, Gosbank was intended only to lend money to state-owned enterprises to enable them to meet their plan targets (Sitnin 1993: 33–34; Nuti 1992). This was a form of nonmarket money creation. In principle, all money was created by decision of state organizations in order to fulfill state goals.

The settlement of transactions between enterprises by transfers of noncash money in Gosbank's books was essentially a state giro system. For a chartalist such as Knapp, such a system represented the abolition of money for the payments concerned, since money consisted of tokens (coins, notes) and noncash settlement was a nonmonetary means of payment (Knapp 1924: 156–157). However, Knapp's first edition was published in 1905, and for a long time now, bank balances have been included in the definition of money. Furthermore, in modern capitalism a large

proportion of payments are made by bank transfer (wire transfer) so such payments were not specific to the administrative-command economy. What was specific to it was that, first, the payments were all made to, and cleared by, a single state institution; secondly, noncash money could not be turned into cash; and, thirdly, transfers of noncash money were a result of instructions from above about the acquisition of goods and not commercial decisions based on the availability of money prior to the transaction.

However, it was difficult to adhere fully to this simple model, as study of the 1930s has shown (Gregory and Tikhonov 2000; Sitnin 1993: 40–41). In that period, there was extensive provision of commercial credit between enterprises (as is normal under capitalism). This partly resulted from substantial payment arrears. These arrears were often dealt with by mutual offsetting of payments due, with Gosbank credits covering any discrepancy. There was also the issue of money surrogates by enterprises that lacked sufficient official money to meet their commitments. These money surrogates were debts of the (state) enterprises which issued them, rather than of the central government or Gosbank. Enterprises and ministries acted as de facto banks, extending credit, and clearing mutual transactions. These unexpected developments partly resulted from the impossibility of Gosbank actually checking all transactions and handling the huge burden of paperwork this would have entailed. It lacked the necessary information and the clerical staff to process carefully (in the pre-IT era) all the data required to meet the intentions of the simple model in an economy marked not just by instructions from above but also by opportunism at all levels below the top. They partly resulted from the failure of the authorities to permit the bankruptcy of state enterprises and the resulting loss of output (and jobs for their workers). A 1932 decree provided a procedure for declaring enterprises insolvent, and it was taken seriously for a short time, but it seems that no enterprises were actually declared insolvent. Their output and employment were valued higher than the maintenance of financial discipline. The lack of financial discipline (the soft budget constraint) was one of the factors generating shortages.

Gosbank's balance sheet for the last pre-Gorbachev year is set out in Table 4.

Table 4 shows that Gosbank combined what in the capitalist world are two different types of banks, commercial banks and central banks. Gosbank's main assets were loans to enterprises. These were mainly short-term loans intended to cover working capital. In the 1930s Gosbank received from the government quarterly credit plans which listed the planned loans to each ministry for that quarter (Sitnin 1993: 33). The role and volume of Gosbank's short-term loans to enterprises (and their superiors such as ministries) were increased by government decisions aimed at increasing efficiency, made in 1954 and 1965, the latter introducing an overdraft-type facility. Lending increased, but the hoped-for increase in efficiency does not seem to have materialized (Nakamura 2017: 108–109). When it made a loan to an enterprise, Gosbank simultaneously increased both sides of its balance sheet. On the assets side, its stock of loans to enterprises increased. On the liabilities sides, its stock of enterprise deposits increased by the same amount. The loan would be intended to facilitate plan fulfillment and thus increase output. Increasing output was the Soviet remedy for shortages. However, in the short run, the loan might also

Table 4 Gosbank's balance sheet for 1984 (billion roubles)

Assets	
Gold, foreign exchange, and balances with foreign banks	15.2
Loans to enterprises	405.2
Loans to the government	115.6
Loans to foreign governments[a]	67.6
Total assets	603.6[b]
Liabilities	
Foreign borrowing	31.6
Enterprise deposits	156.3
Government deposits	128.0
Public deposits	202.3
Cash in circulation	65.3
Own resources[c]	20.1
Total liabilities	603.6[b]

Source: Nakamura (2017: 194). Data on Gosbank's balance sheets for the whole Soviet period with explanations can be found in English in Nakamura (2017: Chap. 5 and 192–194) and in Russian in Kashin and Mikov (2010)

[a]These were probably mainly loans to CMEA countries (CMEA stands for Council for Mutual Economic Assistance, sometimes known as Comecon)

[b]The figures for total assets and liabilities in the source differ slightly from the sum of the constituents given in the source and copied in the table

[c]This corresponds to equity in capitalist banks. This figure was not explicit in the source and is just the difference between assets and liabilities

increase shortages of producer goods. Furthermore, if any of it were to leak into cash money, or be used for siphoning goods from the retail market, when the loan had not produced any extra consumer goods, it would increase shortages on the retail market and prices on the free market for food products.

In addition to lending to state enterprises to provide working capital, Gosbank also made loans to the Ministry of Finance to finance the deficit of the state budget. These loans (and Gosbank's purchase of Ministry of Finance bonds) grew rapidly from the mid-1960s and exploded in 1985–1990 (as was shown in Table 3). When Gosbank made a loan to the government, it increased its assets (loans to the government) and its liabilities (government deposits) by the same amount. If the government spent some of these increased deposits on cash outlays (e.g., pensions or salaries), this would be reflected in Gosbank's balance sheet by a decline in one liability, government deposits, and an increase in another one, cash in circulation (unless the extra income was deposited in the savings bank, in which case it would increase the public deposits, or spent in state retail trade on goods which would otherwise have remained unsold, in which case the decline in government deposits would be reversed). Unlike the situation in capitalist countries in a depression or recession, additional government expenditure financed by borrowing from the banking system could not produce additional output, because administrative-command economies were supply-constrained rather than demand-constrained. Hence additional government borrowing and the expenditures it financed normally

merely increased construction periods, worsened the balance of payments, and added to shortageflationary pressures (especially since it generally did not produce additional consumer goods). Unsurprisingly, Gosbank officials made numerous proposals to reduce the deficit and the need for loans to finance it (e.g., Denezhnoe 2010: 561–565) but were ignored.

Another type of shortageflationary bank loan concerns agriculture. Bank loans to agriculture rose rapidly in the post-Khrushchev period. In 1970–1986 short-term bank loans to the collective farms rose 15 times and to agriculture as a whole tenfold (Narkhoz 1987: 635). They also received long-term loans. A large part of these loans was wasted (e.g., on the purchase of farm machinery which was of poor quality and frequently immobilized because of shortage of spare parts). Such loans were often not repaid and were written off and were similar to nonperforming loans or loan losses/write-offs under capitalism. They increased demand more than supply.

When these Ministry of Finance and agricultural loans were combined with the destabilizing policies of 1987–1991, the worsening terms of trade, and the decline in output, the resulting monetary disequilibrium led to acute shortageflation and contributed to the collapse of the USSR and the initial problems of its successor states.

Money in Trade Between Administrative-Command Economies

The emergence of new socialist countries in Eastern Europe created the need for money to use in trade between them. From 1950 the "clearing rouble" (an inconvertible unit of account) was used for bilateral trade between the USSR and the other socialist countries. In 1963 a special bank (the Bank for International Economic Cooperation) was created for the purpose of multilateral clearing of trade between CMEA members. Accounts in it were kept in "transferable roubles." (The transferable rouble was an inconvertible unit of account and was basically just another name for the "clearing rouble.") The new name signified a recognition, in principle, of the desirability of moving toward multilateralism and convertibility. In the agreement setting up this bank, it was stated that "within one year from the foundation of the Bank . . . the Board will study ways of introducing into the scope of its operations transacted in transferable roubles the possibility of a conversion into gold or freely convertible currencies." This goal was never achieved. The "transferable" rouble remained an inconvertible unit of account for bilateral trade. This caused numerous difficulties in intra-CMEA trade. Nevertheless, in 1963–1990 the transferable rouble was useful in settling bilateral trade balances, and trade between CMEA members grew quickly. In 1990 the USSR proposed that from January 1991, trade within CMEA should be at world market prices and settled in hard currency. This was Soviet recognition, not only of its own hard-currency payments crisis but also of the difficulties resulting from the use of the transferable rouble and its limitations as a means of exchange in international trade. In the changed political circumstances of the time, this proposal played a major role in the dissolution of the CMEA in 1991 (Ellman 2014: 353–354).

Conclusion

Despite the expectations of Marxists, money persisted in planned economies. However, the operation by the state bank of what was essentially a state giro system for making and clearing inter-enterprise payments did represent the abolition of money for those payments from the perspective of a chartalist such as Knapp for whom not all means of payment were money. The 1918–1922 hyperinflation was unintentional and contributed to an economic disaster. It was ended by standard monetary and fiscal policies. The 1930 attempt to move to a moneyless economy was abortive. However, the role of money was distinctive as a result of the specific features of the economic system. All legal money was created by agencies of the state with the intention of facilitating the achievement of state goals. Monetary transactions by state enterprises were, in principle, checked to ensure that they conformed to the plan. Money was only a partial equivalent. There were two monetary circuits.

Nevertheless, because of its usefulness in aggregating goods, money played an important role in economic policy and planning decisions. Furthermore, because of its usefulness in buying food, drink, clothing, and consumer durables, it played a useful role in incentivizing the population. Despite its limited usefulness in obtaining both producer and consumer goods, it also provided an incentive for enterprises to obtain it (e.g., by delivering their output, raising their prices, overstating their costs, or requiring above-plan or off-the-books payments for their products).

In planned economies the role of money was much more restricted than in capitalist economies because of the distinctive features of the former. All producer goods were rationed. Many important consumer goods and services were distributed free or at nominal prices. The political system made position in the political hierarchy more important for obtaining many goods and services than mere possession of money. Because of privileged access to underpriced (relative to market equilibrium) goods and services, in administrative-command economies, unlike the situation in capitalist economies, (uncorrected) Gini coefficients or other indicators of differences in *money incomes* are a very inadequate guide to *real income* inequalities. Economic flows in planned economies were mainly determined by decisions of the national leadership, and the reaction to them at lower levels, rather than by autonomous monetary flows. Cash money in the planned economies was primarily an instrument (similar to propaganda or repression) intended to enable the people at the top of the political hierarchy to control the country and assist in disciplining the population and stimulating them to work hard to achieve the goals set out by that leadership. Noncash money was an instrument to facilitate, and check, plan fulfillment. Like other instruments used in hierarchical systems, money encountered the usual principal agent problems. Opportunism abounded and the center lacked the information to evaluate the real situation of all economic agents. Furthermore, not all national leaders were strict disciplinarians. Some were slack ones, whose failure to enforce strict monetary discipline had serious economic consequences.

The role of the monetary policy and the monetary system was always important but varied sharply in the Soviet period (1917–1991). The issue of *sovznaki* in

1918–1922 helped the new regime to obtain resources during the civil war and economic collapse. The initial stages of the administrative-command system (1926–1947) enabled a monetary system to be developed that served that system. The monetary reform of 1947 eliminated the excess money issued to help finance the 1941–1945 war and enabled rationing (of consumer goods) to be ended. It initiated almost two decades of relative monetary stability. Similar reforms in Eastern Europe and China had similar effects. In 1966–1986 the combined effect of declining economic growth, increased military expenditures, the growth of food subsidies, increased siphoning of consumer goods by enterprises, the growth of money incomes in excess of plan levels, and monetary financing of the budget deficit led to chronic but relatively modest, shortageflation. The loss of monetary control in 1987–1991 was both a result of state weakening and one of the causes of the collapse of the USSR in December 1991 and the difficult legacy it left to its successor states.

The (Soviet) administrative-command economy was conceived in a monetary crisis (1917–1922) caused primarily by state weakness and collapse and died in another monetary crisis (1990–1991) also primarily caused by state weakness and collapse. This illustrates the state theory's link between money and the state. Stable money (usually) needs a strong state, and a strong state needs stable money.

Cross-References

- ▶ Money, Law, and Institutions
- ▶ Money in Wars
- ▶ The Historical Evolution of Monetary Policy in Latin America

References

Abalkin L (1992) K tseli cherez krizis. Spustya god. Luch, Moscow

Barendregt J (1993) The Dutch monetary purge. The monetary consequences of German occupation and their redress after liberation, 1940–1952. Tinbergen Institute research series LI: Thesis Publishers, Amsterdam

Bazhan A (1990) Osobennosti inflyatsii v SSSR. Vestn Statistiki (6)

Bell J, Rostowski J (1995) A note on the confirmation of Podkaminer's hypothesis in post-liberalisation Poland. Eur Asia Stud 47(3):527–530

Bertocco G (2007) The characteristics of a monetary economy: a Keynes-Schumpeter approach. Camb J Econ 32(1)

Carr EH (1958) Socialism in one country 1924–1926, vol 1. Macmillan, London

Davies RW (1989) The Soviet economy in turmoil, 1929–1930. Macmillan, Basingstoke

Davies RW (2001) Making economic policy. In: Gregory P (ed) Behind the facade of Stalin's command economy. Hoover Institution Press, Stanford

Denezhnaya (2010) Denezhnaya reforma v SSSR 1947 goda. Dokumenty i materialy. ROSSPEN, Moscow

Denezhnoe (2010) Denezhnoe obrashchenie Rossii, vol 3. Interkrim-Press, Moscow

Dzerzhinsky F (1926) Zaklyuchitel'nogo slovo tov. Dzerzhinskogo, Leningradskaya Pravda 14 Feb

Ekonomicheskaya (1972) Ekonomicheskaya Entsiklopediya. Politicheskaya ekonomiya, vol 1. Sovetskaya Entsiklopediya, Moscow
Ellman M (1992) Money in the 1980s. From disequilibrium to collapse. In: Ellman M, Kontorovich V (eds) The disintegration of the soviet economic system. Routledge, London
Ellman M (2000) Perestroika economics from the inside. In: Maskin E, Simonovits A (eds) Planning, shortage, and transformation. MIT Press, Cambridge, MA
Ellman M (2014) Socialist planning, 3rd edn. Cambridge University Press, Cambridge, UK
Frank R (2010) Money in socialist economies: the case of North Korea. Asia-Pacific J 8(2):1–13
Gaidar E (2006) Gibel' imperii, 2nd edn. ROSSPEN, Moscow
Garvey G (1977) Money, financial flows, and credit in the Soviet Union. NBER, Cambridge, MA
Goland Y (1994) Currency regulation in the NEP period. Eur Asia Stud 46(8):1251–1296
Goland Y (2006) Diskussii ob ekonomicheskoi politike v gody denezhnoi reformy 1921–1924. Ekonomika, Moscow
Goodhart C (1998) The two concepts of money: implications for the analysis of optimal currency areas. Eur J Polit Econ 14:407–432
Gregory P, Harrison M (2005) Allocation under dictatorship: research in Stalin's archives. J Econ Lit XLIII(September):407–432
Gregory P, Tikhonov A (2000) Central planning and unintended consequences: creating the Soviet financial system, 1930–1939. J Econ Hist 60(4):1017–1040
Grossman G (1986) Inflationary, political, and social implications of the current economic slowdown. In: Hohmann H, Nove A, Vogel H (eds) Economics and politics in the USSR: problems of interdependence. Westview, Boulder
Harrison M, Kim B-Y (2006) Plans prices and corruption: the Soviet firm under partial centralization. J Econ Hist 66(1):1–41
Holzman F (1960) Soviet inflationary pressures, 1928–1957: causes and cures. Q J Econ 74(2): 167–188
IMF et al (1991) A study of the soviet economy, vol 1. OECD, Paris
Kashin I, Mikov V (eds) (2010) Balansy gosudarstvennogo banka SSSR, 1922–1990 gg (Po stranitsam arkhivnykh fondov Tsentral'nogo banka Rossiiskoi Federatsii, vypusk 9). Central Bank of Russia, Moscow
Khanin G (2010) Ekonomicheskaya istoriya Rossii v noveishyeye vremya, Ekonomika SSSR i RSFSR v 1988–1991 godakh, vol 2. NGTU, Novosibirsk
Kim B-Y (2002) Causes of repressed inflation in the consumer goods market, 1965–1989: retail price subsidies, the siphoning effect, and the budget deficit. Econ Hist Rev 55(1):105–127
Kleiman E (2000) Early inflation tax theory and estimates. Hist Polit Econ 32(2):233–265
Knapp G (1924) The state theory of money. Macmillan, London. 1st German edition 1905
Kolodko G, McMahon W (1987) Stagflation and shortageflation: a comparative approach. Kyklos 40(2):176–197
Kornai J (1980) Economics of shortage, vol A & B. North Holland, Amsterdam
Kozlov V (2002) Mass uprisings in the USSR. M.E. Sharpe, Armonk
Marx K (1867) Das Kapital, vol 1. Otto Meissner, Hamburg. There is an English edition, Marx, K. (n.d.) Capital, vol 1. Progress, Moscow. The relevant passage are on pages 73–75, 90, 92–96, & 108
Nakamura Y (2011) Did the Soviet command economy command money? A quantitative analysis. Eur Asia Stud 63(7):1133–1156
Nakamura Y (2017) Monetary policy in the Soviet Union. Palgrave Macmillan, New York
Nakamura Y (forthcoming) The relationship between the real and the financial economy in the Soviet Union: an analysis of government debts using newly available data, Explorations in economic history
Narkhoz (1987) Narodnoe khozyaistvo SSSR za 70 let. Finansy i statistika, Moscow
Narkhoz (1990) Narodnoe khozyaistvo SSSR v 1990g. Finansy i statistika, Moscow
Nuti D (1986) Hidden and repressed inflation in Soviet-type economies. Contrib Polit Econ 5:37–82

Nuti D (1992) Socialist banking. In: Newman P, Milgate M, Eatwell J (eds) The new Palgrave dictionary of money and banking, vol 3. Macmillan, London
Pavlov V (1989) Pochemu my voshli v finansovyi shtopor? Kak budem iz nego vykhodit'? Ekonomicheskaya gazeta no.44. Reprinted in V. Pavlov, Finansy – nasha glavnaya zabota. Finansy i statistika, Moscow (1990). p 291
Pavlov V (1995) Upushchen li shans? Terra, Moscow
Peebles G (1991) A short history of socialist money. Allen & Unwin, Sydney
Podkaminer L (1982) Estimates of the disequilibria in Poland's consumer markets, 1965–1978. Rev Econ Stat 64(3):423–431
Podkaminer L (1988) Disequilibrium on Poland's consumer markets. J Comp Econ 12(1):43–60
Preobrazhensky E (2006/1920) Arkhivnye dokumenty i materialy: 1886–1920 gg. Glavarkhiv Moskvy, Moscow. Originally published in 1920 as Bumazhnye den'gi v epokhu proletarskoi diktatury. Gosizdat, Moscow
Russian (1995) Russian federation: report on the national accounts. World Bank & Goskomstat, Washington, DC
Sitnin V (1993) Sobitiya i lyudi. Luch, Moscow. Copy provided by the International Institute for Social History, Amsterdam
Stalin J (1955) Works, vol 13. Foreign Languages Publishing House, Moscow
Stalin i Kaganovich (2001) O.V. Khlevnyuk et al. (comps) Stalin i Kaganovich. 1931–1936 gg. ROSSPEN, Moscow
Zverev A (1973) Zapiski ministra. Politizdat, Moscow
Zverev A (2012) Stalin i den'gi. Algoritm, Moscow

Complementary Currencies

20

Massimo Amato and Luca Fantacci

Contents

Abstract

Recent decades have witnessed the proliferation of monetary instruments that differ, in many respects, from the official currency issued by the central bank and by the regulated banking system, such as local currencies, corporate barter, and mutual credit systems. Today, complementary currencies tend to appear as a bizarre exception to the rule of one single, all-purpose currency for each country (or currency area). However, in historical and comparative perspective, it is monetary plurality that prevails: different monies coexist side by side, serving different purposes, in most economies throughout most periods. The most significant and pervasive distinction in premodern economies was between internal and external money: one currency for the domestic economy and a different one for foreign trade. Only the rise of modern territorial states in the seventeenth

M. Amato · L. Fantacci (✉)
Department of Social and Political Sciences, Università Bocconi, Milan, Italy
e-mail: massimo.amato@unibocconi.it; luca.fantacci@unibocconi.it

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_50

century established the uniformity of the currency in each jurisdiction and the monopoly of coinage as a prerogative of sovereignty. In this chapter, we provide a broad overview of the various forms of complementary currencies throughout history, we analyze more recent experiments, and we review current proposals to introduce parallel currencies, before we turn to discuss how an enquiry into monetary complementarity can help shed light on the very nature of money itself.

Keywords
Complementary currencies · Parallel currencies · Alternative currencies · Clearing systems

Introduction

Recent decades have witnessed the proliferation, under a variety of names and guises, of monetary instruments that differ, in many respects, from the official currency issued by the central bank and by the regulated banking system (Zelizer 1998, Peacock 2014). Just to mention the most relevant examples (more broadly described by Lietaer 2001 and Kennedy et al. 2012), we may quote local exchange trading systems (LETS, SEL); time banks, used by individuals to exchange goods and services on the basis of the time required to produce them; local currencies issued in the form of paper money to support local economies (time dollars, Ithaca Hours, Chiemgauer); mutual credit systems between businesses aimed at reducing the cost of working capital and enhancing sales (such as WIR and Sardex); and targeted currencies designed to provide access to specific categories of goods or services (Fureai Kippu) – not to speak of the plethora of privately issued monies (qv).

These experiments, variously designated as complementary, social, parallel, or community currencies, may have very different characteristics, but, and very different objectives: environmental protection (Seyfang 2001), the creation of social capital (Richey 2007), the resistance against globalization (Pacione 1997, Tibbet 1997), the development of a moral economy (Peacock 2006). However, they all share one common feature: they do not have the purpose of substituting the official currency, but of complementing it, by circulating within appropriately delimited exchange circuits (Blanc 2000, 2011), and according to different rules (Maurer 2005).

The recent emergence of monetary plurality stands in stark contrast with a secular trend that seems to point toward a relentless monetary unification: from a situation characterized by the coexistence of different monies for different purposes to a system ruled by a single currency for all. Today, complementary currencies tend to appear as a bizarre exception to the rule of the single currency. However, in historical and comparative perspective, what prevails is monetary plurality: different monies coexist side by side, serving different purposes, in most economies throughout most periods.

In fact, plurality appears to be a constant and pervasive feature of premodern monetary systems. This has been widely acknowledged by anthropological research:

> "our money is 'aII-purpose' money. [. . .] Early money is, as we saw, special-purpose money. Different kinds of objects are employed in the different money uses; moreover, the uses are instituted independently of one another. The implications are of the most far-reaching nature. There is, for instance, no contradiction involved in 'paying' with a means with which one cannot buy, nor in employing objects as a 'standard' which are not used as a means of exchange." (Polanyi 1957: 264–266)

Monetary plurality implies a distinction between monetary functions, which is not merely conceptual but operational: different monies perform different functions.

The most significant and pervasive distinction within premodern monetary systems is between internal and external money: one currency for the domestic economy and a different one for foreign trade. Analogous distinctions may be found in the most disparate settings, from the dichotomy between market type A and B in early modern Europe described by Fernand Braudel (1979: 49–51) to the distinction between "taonga" and "oloa" in the Polynesian culture studied by Mauss (1990 [1925]: 10–13). A similar distinction is drawn by Aristotle since the beginning of Western reflection on money.

It is the rise of modern territorial states in the seventeenth century that established the uniformity of the currency in each jurisdiction and the monopoly of coinage as a prerogative of sovereignty. The concentration in the hands of the sovereign (monarch or parliament) of the *ius cudendae monetae* is one of the most important chapters of the making of the modern (Weberian) state.

Hence, since the early modern age, the suppression of monetary plurality has been regarded as a direct political consequence of strong centralistic territorial sovereignty (Helleiner 2003). In the perspective of the modern state, plurality is the sign of a monetary disorder. This has also been the interpretative tendency of many monetary historians vis-à-vis the monetary multiplicity characterizing the premodern world. Many, but not all: authors such as Einaudi (1953) and Bloch (1954) have emphasized the economic significance of the specialization of monies and the distinction of monetary circuits that permanently characterized monetary architectures until the establishment of the gold standard.

By subjugating even national currencies to one single physical standard, the international gold standard (qv) clearly epitomizes the single-currency system. Indeed, between the end of the nineteenth and the beginning of the twentieth century, gold claimed the status of a global money, circulating throughout the world and serving as the ultimate means of payment to all intents and purposes.

The demise of the last vestige of the gold standard, with the suspension of dollar convertibility in 1971, ushered in a new era of monetary plurality. However, and interestingly enough for our purposes, it did not bring in its wake the end of the single-money doctrine, according to which one currency is better than many: first, because the dogma of the "single currency" continues to hold within the borders of each state – and sometimes in even broader areas, as in the case of monetary unions (qv) – and second, and perhaps more importantly, because even in the international sphere, where monetary plurality prevails, different national currencies do not perform different complementary functions, but they compete for the same function: that of a global reserve currency.

Monetary plurality, therefore, does not necessarily imply complementarity. In fact, it may involve a very different form of relationship, which can be described as substitutability. Complementarity designates a "division of labor" between different currencies that circulate in separate exchange circuits, performing different tasks. Substitutability entails a competition between all-purpose monies that coexist within the same sphere, contending to gain market share (as in the regime advocated by Hayek 1990). Complementary currencies are normally tied to one another by a system of adjustable pegs. Substitutable currencies are traded at a variable exchange rate, which is a market price reflecting the conditions of their relative supply and demand. Complementarity allows for a distinction between different monetary functions and in particular for the separation and articulation of the functions of unit of account and means of exchange. Substitutability implies the collapse of all functions in the same money, which is at the same time unit of account, means of exchange, and store of value.

Even from these preliminary remarks, it should be clear that *complementarity between monies* is a broader and deeper phenomenon than the more recent social and political movement of *complementary currencies*: indeed, it touches upon the functions and nature of money. Therefore, the clue for a correct understanding of the meaning of the more recent phenomena, as well as of their relevance in both factual and theoretical terms, relies on a historical and theoretical understanding of monetary complementarity. This is what we set out to achieve in the following sections, where we provide a broad overview of the various forms of complementary currencies throughout history, we analyze more recent experiments, and we review current proposals to introduce parallel currencies, before we turn to discuss how an enquiry into monetary complementarity can help shed light on the very nature of money itself.

The Complementary Currency System of Early Modern Europe

Since the beginning of its monetary history, Europe has been characterized by a plurality of monetary architectures. Different units of account and different coined monies were used not only in different areas but also in the same areas for different purposes. Historically, therefore, the process of gradual unification of functions and monies is very recent, and, in a long-term perspective spanning over centuries, it represents rather the exception than the rule.

The first, fundamental feature that characterized European monetary systems for centuries is the distinction between the unit of account and the means of exchange. What is striking, and difficult to understand from a modern standpoint, is that this was not merely a theoretical, but a practical distinction: for an entire millennium, from Charlemagne to Napoleon, the money that was used to count differed from the money that was used to pay. The pound (together with its subunits, the shilling and penny) was not coined money, but a pure name, an abstract unit of account, used to denominate prices and debts. Actual coins made of metal served as means of exchange, but their value in terms of the unit of account (extrinsic value) was not fixed, and it was not necessarily proportional to their metal content (intrinsic value).

The most tangible evidence of this is the fact that, until the franc germinal of 1805, virtually no European coin had the one feature that today is considered indispensable in money, whether it is made of metal or paper: a number designating its value inscribed in the body of the currency. In fact, the value of each coin was determined by public authority through a decree called "tariff" and would vary from time to time (Fantacci 2005, 2008).

Why such an apparent complication? The regulation of the tariff by the monetary authority was not arbitrary, but was intended to pursue designated policy goals and to follow specified rules. The main purpose was to regulate, through the tariff, the power of money to discharge debts: by raising the value of the coins, *all* debts were made lighter and *general* insolvency crises could be averted. At the same time, the increase in the nominal value of the coinage could serve to stave off the deflationary tendencies of a metal standard, where the money supply would have been otherwise constrained by the availability of precious metal. Moreover, the separation between the unit of account and the means of exchange implied the absence of the third function that is now commonly ascribed to money, namely, the store of value: since the coins did not have a fixed value, they could not serve, strictly speaking, as a store *of value*; they were merely a store *of metal*, the value of which could change according to the tariff. This feature contributed to contrast the hoarding of money and to incentivize its circulation.

However, a further explanation for the distinction between unit of account and means of exchange has to do with the possibility of preserving, through the regulation of their relationship, a second form of monetary complementarity between an internal money for the domestic economy and an external money for foreign trade. Europe had a dual currency system based on this distinction for over 500 years.

The recovery of long-distance trade in Europe, starting from the thirteenth century, fostered a gradual commercial unification in a context of political fragmentation. The Florentine florin and the Genoese ducat, which were coined starting from 1250 in two of the most prominent trade centers of the period, were the first gold coins to be minted in Europe after centuries of economic and political decline. Together with the Venetian sequin, they were continuously minted for centuries with a content of precious metal that was extraordinarily stable in terms of weight and fineness. On this ground, they were received throughout Europe, and beyond, by the cosmopolitan class of merchants, who were clearly interested in the certainty of the intrinsic value of the coins that they received and gave in exchange for their precious goods. Large coins of gold and silver were thus the money of international trade.

At the same time, local economies relied on a different type of currency: small, "dark" coins of copper and cheap alloy. Unlike the large coins used for foreign trade, the small coins contained a low and variable quantity of precious metal. This occurred because their intrinsic value was ultimately irrelevant, as they were used as a pure intermediary in the local exchanges, as a compensation for work and as an entitlement to a share of primary goods, hence as a token money that was received and dispensed at its face value. Therefore, small coins maintained stable over centuries their extrinsic value, regardless of their exiguous, and ever declining, metal content.

Large and small coins were truly complementary: they had different characteristics because they were intended to serve different purposes, to circulate in two distinct but communicating exchange circuits that may be best described by referring to the classical distinction, made by Fernand Braudel (1979: 49–51), between public and private markets. The public market was the place for local exchanges between producers and consumers and between artisans from the town and peasants from the country, where primary goods were traded according to the rules set by the public authority, in view of preserving distributive justice. The appropriate currency for this type of exchange was small coins, characterized by an irrelevant substance but a fixed value (guaranteed within the state by the "prince") for the payment of prices and debts. The private market was the open space of long-distance trade, dominated by specialized intermediaries exchanging mostly luxury goods on the basis of private norms and regulations (the *lex mercatoria*) essentially aimed at preserving commutative justice. Such form of exchange was best served by large coins that were received on the basis of their reliable and unalterable metal content, regardless of the value that local authorities might wish to assign them.

As appreciated by Einaudi (1953), this dual currency system allowed to reconcile two objectives that tend to be conflicting in more recent monetary regimes: a stable external currency to ensure the required certainty in commercial and financial relationships with foreigners and a managed internal money to sustain full employment and regulate credit in the local economy. However, as all institutional settings, also this peculiar monetary system required to be preserved in its delicate balances by a guarantor, in the person of the monetary authority. The system functioned quite well until governments were in the position of acting as impartial arbiters between debtors and creditors, and not as major debtors themselves, i.e., until the prince had a treasury. The rise of the modern state – with its standing army and bureaucratic administration – transformed the government into a prominent economic actor and the treasury into a permanent debt, eventually undermining the impartiality in the regulation of the currency. The system was thus eventually distorted from its original purposes and twisted to serve the funding requirements of insatiable governments, constantly at war with one another.

The dual currency system started to be dismantled in England with the so-called financial revolution, which was also, and perhaps even more significantly, a monetary revolution that led from the complementarity between internal and external money to the substitutability between paper money and gold. In order to stop the arbitrary monetary mutations orchestrated by governments for the sole purpose of financing their military expenditure, in 1696, the Great Recoinage irrevocably fixed the relationship between unit of account and metal, inaugurating a silver standard (turned into a gold standard in 1717). At the same time, the establishment of the Bank of England, with the privilege of issuing paper money, created a potentially unlimited source of liquidity to finance both trade and war. Formally the distinction between two types of money remained: banknotes circulated within the country to sustain commerce, while gold was used for international settlements and to pay the troops abroad. However, the relationship between the two currencies changed dramatically: it was no longer a relationship of complementarity, sanctioned and

regulated by the tariff, but a relationship of substitutability, relying on the promise of convertibility.

With the adoption of the metal standard by an increasing number of countries, the dual currency system gradually disappeared. It survived, however, even in the European periphery, until the late nineteenth century. A country that had recently gained its independence from the Ottoman Empire, like Bulgaria, exercised its monetary sovereignty without even actually minting any coin of its own, by allowing foreign coins to circulate within its territory at a designated parity in terms of an abstract unit of account.

Unofficial, Low-Denomination Coins: Tokens, Méreaux, and Pseudo–Monies

Other forms of complementary currencies in premodern Europe were characterized by an even more restricted and informal sphere of circulation compared to small coins. It was the case of tokens, also called *méreaux* (from the Latin word *merere*, to deserve) or *jetons* (from *jectare*, to throw). These were coin-like objects that, however, were differentiated from official money by three distinctive features. First, they were not issued by a public authority, but by an intermediate body, such as a religious community or a guild. Second, they were made of a material with hardly any intrinsic value, such as copper, aluminum, lead, or tin, but even leather or cardboard. Third, they were characterized by a targeted use and eventually gained circulation but only in a limited context.

The first *méreaux* appeared reportedly in France in the mid-thirteenth century and perhaps not by chance. This was the period when Louis IX attempted to unify the royal coinage at a national scale and revoked minting rights previously enjoyed by local authorities. Due to the centralization of money supply, and the inefficiency of the "top-down" transmission of monetary policy, certain local economies could suffer from a scarcity of means of exchange. This opened the way to the "bottom-up" introduction of complementary currencies.

The need for complements to the official currency was enhanced, in subsequent centuries, by the gradual diffusion of monetary transactions. Different types of *méreaux* were introduced for the most various purposes (Labrot 1989).

The most important example is provided by communion tokens, which were originally handed out in churches to those who participated in celebrations. The tokens gave access to the distribution of food, beverage, or clothes from a common fund. The *méreaux* of the chapter of Boulogne significantly bore the inscription *moneta distributiva*, "distributive money." Even lay confraternities and guilds issued tokens to the participants of meetings and celebrations who could then use them to access specific goods (typically bread) or services (e.g., healthcare services). Other similar pseudo-coins were issued expressly for specific types of beneficiaries: charity tokens would be assigned to the poor; remuneration tokens would be used to pay the local population for public works, as the feudal system of the *corvée* was gradually phased out (Ibid.).

In periods and places that suffered from a scarcity of coins, these tokens would then gradually gain circulation in the local economy, by being used to pay workers or suppliers. In some areas, they even occasionally became the only type of money in circulation. This occurred particularly in territories at the border between two jurisdictions. Between the sixteenth and seventeenth century, in Roussillon, a contested region between France and Spain, the only means of exchange for local trade was provided by the church in the form of brass tokens called *pallofes* (from the catalan *pell*, skin). A particularly significant case is represented by the city tokens, issued by municipal authorities, particularly under siege. Quite appropriately, these were called "necessity money" (*monnaies de nécessité*) or "emergency money" (*Notgeld*). Similarly, in the eighteenth and nineteenth century, in Britain, token coins and scrips were issued by merchants, industrialists, or even publicans to meet the needs of local exchanges in times of war or economic hardship (Cameron and Neal 1989: 171).

All these forms of unofficial money gradually disappeared when territorial states eventually managed to enforce their monopoly over the issuance of money and to supply national economies with adequate means of payment. In certain isolated communities, however, they survived well into the twentieth century, like in the Balearic Islands.

Almost completely ignored by economic historians, they are known to numismatists under the heading of *exonumia*, or *paranumismatica*. Most of them still await to be studied from an economic perspective and to be appreciated for their role in complementing official coins in the fulfillment of the monetary functions required by local communities.

However, recent research in the fields of economic history and anthropology has contributed at least to question the traditional approach to monetary systems that portrays all-purpose money as the ultimate point of arrival of an evolutionary trajectory having the force of a universal law. In particular, the existence of complementary currency systems has been proved to be pervasive both in Africa (Guyer 2012) and in Asia (Kuroda 2008b).

Interwar Period: Depression Scrips and Local Currencies

During the Great Depression, local economies suffered from a new version of the problem of monetary scarcity that had characterized previous centuries. The problem then was the paradox of "poverty in plenty," i.e., the coexistence of unused resources and unsatisfied needs, due to a lack of means of exchange. The answer was sought, rather spontaneously and in a variety of circumstances, in the issuance of a local currency.

The notion had originally been conceived by the German businessman and self-trained economist Silvio Gesell at the beginning of the century and had started to find significant applications in Germany and Austria after the outbreak of the Great Depression, in 1931. The inscription on the banknotes issued in the Austrian town of Wörgl gave a clear statement of their purpose: "Lindert die Not, gibt Arbeit und

Brot" (lightens the need, gives work and bread). The scrip was intended to overcome the state of deprivation of local communities, by bridging the gap between potential supply and potential demand. The scrip was typically issued by a municipality that would use it to hire unemployed workers. An agreement with a number of local merchants would allow the scrip to be spent and to remain in circulation for the payment of supplies that could be satisfied locally, until it was eventually redeemed by the municipality at its face value at a given date (Fantacci 2013).

In 1932, stamp scrip started to spread across the United States. The otherwise orthodox American economist Irving Fisher, after having surveyed the most significant examples, came to the conclusion that, at certain conditions, it could actually provide relief to economic distress. He identified the essential feature of a successful local currency in the application of a tax to be paid periodically by the holder of the currency as a percentage of its face value: Fisher called it an "ambulatory tax," but it is more appropriately described as a "demurrage," since it is applied to the currency that is hoarded (demurred), rather than to the currency that circulates (ambulates) – circulation being rather the deliberate purpose of the tax. Thanks to this expedient, stamp scrip could be deprived of the function of a store of value.

Precisely because it could not serve as a store of value, stamp scrip had an advantage over ordinary money: it circulated more rapidly. Each holder had an incentive to spend the scrip as soon as possible to avoid the tax: "In Stamp Scrip we have, for perhaps the first time in history, a tax which the taxing authority wants to see avoided - by the maximum number of people. For, by passing the tax on to the next fellow you speed the scrip, and that is the chief purpose" (Fisher 1933: 14). Fisher calculated that, thanks to the tax, stamp scrip could achieve a velocity of circulation that was 12 times the speed of the dollar during the depression (Ibid.).

A further benefit of the tax was that it allowed the issuer to guarantee the ultimate redemption of the scrip. In fact, the scrip was a liability of the issuer, and, in order to enhance its credibility and acceptance, it had to be redeemed in money at a specified date. Following Fisher's example, suppose the payment is due 1 year after the issuance, and suppose that the tax is equal to 2 percent of the face value of the scrip and is to be paid every week. Since the tax is paid in money to the issuer, the latter would over a year collect 104 percent of the value of the scrip, enough to allow for full redemption and to cover the costs of the operation. In other terms, the scrip would be a self-liquidating debt and would perhaps even allow for enhanced tax revenues, thanks to the multiplier effects produced by the monetary expansion on local economic activity.

Thus, thanks to the tax, the scrip would not be inflationary, despite the fact that it was created out of nothing. The money collected through the tax would provide a backing to the scrip and eventually ensure its convertibility in actual money at a par. Moreover, the tax would act as an automatic mechanism to reabsorb idle balances: to the extent that the scrip was not spent, it was gradually withdrawn from circulation. Thus, what ultimately provided a backing for the scrip was the goods that were offered in exchange for it.

Fisher conceived of stamp scrip as a temporary integration to official money: not as a substitute but as a supplement. As an example, to "prime the pump" (Fisher

1933: 66): when official money should eventually start circulating again, there would be no need for stamp scrip. However, Fisher envisaged the possibility of applying the principles embodied locally in stamp scrips also at the level of the national monetary system, to support public spending, give work to the unemployed, and reflate the economy (Fisher 1933: 59).

Following the "free money" doctrines of Gesell, another type of complementary currency was created in Switzerland in the mid-1930s, as a countermeasure to contrast the economic crisis and as a way to promote a sustainable economy at the local level. In 1934, a group of cooperators founded the WIR cooperative in Zurich. WIR is the German word for "us," indicating the cooperative character of the initiative, and it is the first syllable of "Wirtschaftsring," meaning business circle. The businesses that joined the cooperative were granted an interest-free overdraft facility denominated in a complementary currency also called WIR. The currency WIR was not issued in the form of a scrip, but existed only as a pure unit of account, to denominate the balances of the members with the cooperative. Each time a member had to pay another member, the buyer's account would be debited and the seller's account would be credited by an equivalent amount. Each member could run a negative balance up to a predetermined amount, corresponding to its creditworthiness and to its capacity to buy and sell locally, i.e., within the business circle of the cooperative. Negative balances would have to be reabsorbed by selling to other members. Positive balances could not be withdrawn in cash, in official currency, but would have to be spent within the circle, purchasing goods from other members. In this way, the circle would provide cheap funding and enhanced sales opportunities to all members. The number of businesses involved, and the value of goods exchanged, grew steadily since the inception: after a few years, the cooperative was transformed into a bank, and subject to Swiss banking regulation, but continued to provide interest-free credits in complementary currency. After having reached over 70.000 participants in the early 1990s, the WIR Bank is still active today throughout the Swiss confederation, particularly in the Alemannic cantons, with total loans over 4 bio Swiss francs, roughly one fifth of which are denominated in WIR (Stodder 2009).

Recent Developments: Community Currencies, Mutual Credit Schemes, and Corporate Barter

Crises, by making evident the limits of the existing monetary framework, often pave the way to the propagation of monetary experiments, most of which, however, are designed and implemented by improvised practitioners. Among these, it is important to discern, following a distinction by John Maynard Keynes, the manifold "monetary cranks" from the rare "brave heretics" (Ingham et al. 2016). Among the latter, Keynes most certainly included Silvio Gesell, whom he considered an "unduly neglected prophet," from the spirit of which "the future will learn more [. . .] than from that of Marx" (Keynes 1936: 353-355 passim).

As discussed above, Gesell provided the inspiration in the 1930s for two types of complementary currencies, exemplified by the Wörgl shilling and the WIR. Now, the distinction between these two types of complementary currency is fundamental also to frame the contemporary experiments that have proliferated after the 2007 financial crisis and that can be roughly divided in two groups: those that are backed by official currency and those that are used as units of account within a clearing system.

Interestingly enough, in recent years, Gesell's idea of negative interest rates has been applied even to official money, more specifically by the European Central Bank starting from July 2014, as a way to contrast the propensity to hoard, to avoid the liquidity trap, and to alleviate deflationary pressures (Coeuré 2014). Keynes himself stated that the idea of demurrage was sound. What has received much less attention is the fact that for Keynes this was only a half-baked solution, because money "is not unique in having a liquidity-premium attached to it" and it has only "a *greater* liquidity-premium than any other article" (Keynes 1936: 357–8). It is then not sufficient, for him, to remove the liquidity character of money to force people to spend it. What is required, in his view, is an adequate design of the monetary circuit in order to provide actual occasions of expenditure and thus to boost the velocity at which the currency circulates.

The idea of creating a currency that serves as a pure means of exchange and not as a store of value is a common feature of the two types of complementary currencies, which differ in terms of how they pursue this objective. The principle underlying the first type of complementary currencies, which are also called "community currencies" or "local currencies" (*monnaies locales complémentaires* in France, where their number is steadily growing), is precisely that a *redenomination* of the currency in order to restrain its circulation to a given territory can boost its velocity of circulation (whether or not it is reinforced by demurrage). The rationale for the second type is that, within a community, a large part of one's debts toward certain members of the community can be offset by one's credits toward other members of the community. In this case, we can speak of *mutual credit schemes* or of multilateral clearing systems.

As mentioned above, France has witnessed the widest proliferation of local currencies backed by official currency (today, in 2018, their number is around 50). Normally they are small experiments that intend to remain small, with a strong grassroots component, even if in some cases the associations that manage them are financed with public funds, or led by local authorities, as in the case of the SOL Violette in Toulouse.

Following the example of WIR, a wave of mutual credit schemes started in 1983 with the first LETS (local exchange trading system) experiment launched in British Columbia by Michael Linton. Since then, LETS schemes have proliferated in all continents with a peak in the 1990s. In Europe, it is the UK and France (where they are called SEL, *systèmes d'échange local*) that recorded the most widespread diffusion of the model. The architecture of LETS/SEL is the same as WIR: a central counterpart keeps account of the exchanges between participants; participants benefit of an overdraft facility that allows each of them to buy before having started to sell; balances are kept in an internal unit of account and are not convertible in official

money; the unit of account is normally pegged to the official currency at a fixed rate, but it can also be linked to the time required to offer a given good or service, as in time banks.

The use of time as a unit of account normally reduces the range of the exchangeable goods within a circuit. However, this is frequently not an unintended side effect, but rather a goal of such schemes. This applies also to the backed currency circuits. The idea in this case is that the delimitation of the geographical scale goes hand in hand with a reduction in the scope of market exchanges and with a restriction of the goods and services exchangeable within the circuits. In this case, the socio-ethical dimension prevails over the economic dimension.

In other cases, the goal of a transformation of the economic environment, which is almost always present in this sort of experiments, does not take the form of a mere territorial demarcation but of a general rearticulation of economic and social relationships within the local community. If "integral cooperation" was the explicit aim of the founders of WIR in 1934, something similar is also the declared goal of the founders of Sardex, a mutual credit system launched in Sardinia, Italy, in 2010. The idea that has guided the development of Sardex is that all the economic actors ought to be involved in the clearing scheme, in order to make it possible for them to reinterpret their position in the economic process. Accordingly, the Sardex circuit has not been restricted to SMEs, for which it represents a way to overcome the difficulty to access financing from banks and financial markets; but it has been extended to include also employees, in order to redefine the relationship between businesses and their workers, who in the local economy are also citizens and consumers. The aim of including the final demand has led to extend the circuit also to public administrations. This sort of integrated clearing circuit aims to achieve social ends via the empowerment of economic relations that in turn embody an alternative vision of how the economy ought to work (Amato and Fantacci 2014: 118–144).

Pure B2B clearing systems can be observed, however, even without any reference to sociopolitical goals of transformation of the economy: the so-called corporate barter is a widespread form of trade credit among businesses, and it is represented by an international organization (IRTA, International Reciprocal Trade Association). The global weight in terms of transactions of the corporate barter is far from negligible (City of London 2011).

Proposals for the Eurozone: Hard ECU, Parallel Currencies, and Grassroots Alternatives

The monetary history of what is now the European Union appears to be a good illustration of what is at stake with complementarity. Otmar Issing's 1996 short book on the project of the euro was entitled *Europe, political union through common money?* It is worth noting that a *common currency*, i.e., a currency that various countries use to settle their reciprocal commercial and financial relationships, need not also be a *single currency*, i.e., a currency that replaces national currencies even in the domestic economies. However, for Europe a decision was

taken, and the euro became, for the countries who adopted it, not only a common but also a single currency.

The complementarity between the European national currencies and an international currency common to all member states proved to work well under the European Payments Union (1950–1958). The EPU was a clearing system along the lines of Keynes's proposal for an International Clearing Union. Each member state had an account and an overdraft facility granted in proportion of its trade with other European countries; the dollar, not being the national currency of any of the participating states, was able to work as a pure international unit of account, exactly as the bancor in Keynes's plan. The overdraft facilities provided by the EPU were a major source of funding for intra-European trade and played a decisive role in promoting postwar growth and the integration of the European economy. Moreover, unlike the EMU, the EPU avoided the buildup of major balance of payments disequilibria among member countries. A key factor in this respect was precisely the complementarity between the national currencies and the international unit of account, within a system of adjustable pegs that allowed to devalue (or revalue) the exchange rate of a member country in the face of persistent deficits (or surpluses).

The EPU was thus able to promote a balanced growth of intra-European trade, at the same time leaving each member state free to pursue the prosperity of the domestic economy through an autonomous monetary policy. The reconciliation of these two potentially conflicting goals was indeed facilitated by the separation and articulation (i.e., by the complementarity) between an international money of account and national currencies (Amato and Fantacci 2012: 110–120). Even after the EPU experience, there was a strong case against replacing national currencies with a single European currency: the lack of convergence of the economies that should have been involved.

It is indeed quite plausible, as an increasing number of people today are ready to admit, that the European unification process may have needed a *common* currency, but not necessarily a *single* currency. However, in 1990, the title of the report of the Commission of the European Communities was, incontrovertibly, *One Market, One Money.*

In fact, the step that preceded the institution of the euro as a single currency was compatible with an evolution of the European monetary system in the direction of complementarity: the European Currency Unit (ECU, an acronym evoking the *écu de marc* of the renaissance trade fairs, an international and private unit of account complementing and not replacing local currencies) could have developed in two different ways. The first would have been to use ECU indeed as a unit of account within a "European Clearing Union" and to link to it all the national currencies through a system of adjustable pegs, in a sort of resumption of the EPU (Day 1954: 212).

The second option had a chance to become a viable plan in the form of a "hard ECU," i.e., of a currency to be issued *in parallel* with the national ones, and not in competition with them, which were deemed to remain in circulation (Phelan 2014). While the proposal of the hard ECU was eventually dismissed by the British government itself, and has not been reconsidered, the theoretical case for a European Clearing Union has been reevaluated after the Eurozone crisis.

With its pretention to achieve *as an effect* what the theory of optimal currency areas indicates as a *prerequisite*, i.e., perfect factor mobility, the single currency has actually wiped out, and without much discussion, all the possible alternatives, surreptitiously insinuating the idea of its irreversibility. After the "sovereign debt crisis," the debate has been polarized between accepting the monetary union as it is or exiting the euro and thus de facto also the EU.

Yet, some proposals aiming to remain within the euro by reforming its architecture have been aired. The additional difficulty is the formally irreversible character of the euro, which makes the reintroduction of local-national currencies in a single currency framework more difficult than it would have been to introduce a common currency in a national currency framework.

The general idea of similar proposals is to introduce parallel national currencies, but in the form of supplementary means of payment pegged to the euro and backed by future tax revenues. The theoretical precedent is Irving Fisher's tax anticipation scrip. Before being used for tax discount (thus decreasing the tax revenues by the same amount), say in 3 years, these vouchers are supposed to enhance the effective demand with a multiplier effect on domestic GDP, so as to increase future tax revenues over and above the value of the tax discounts (for different versions of these projects, see Amato et al. (2016) and Théret (2016)).

What is crucial in all these proposals is the correct calculation of the extension and timing of the multiplier effect, on one hand, and, on the other hand, the provision for correction mechanisms in the case that domestic expansion should lead to a deterioration of the current account. A device like a European Clearing Union could provide adequate mechanisms of symmetric adjustment of (both positive and negative) imbalances in the same way as the EPU or ICU. National currencies could serve as units of account within national (or regional) clearing systems, whereas the euro could act as a common unit of account to finance intra-European imbalances (Kregel 2017).

The abovementioned programs of parallel currencies would be implemented at a state level, but the principle of subsidiarity that the EU at least formally recognizes might lead to the compresence of a multiplicity of complementary layers of money circulation, starting from the bottom (local complementary currencies) up to the top (euro as a common currency). The existence of several community-funded programs supporting the elaboration of community currency programs (CCIA, D-cent, DigiPay4growth) seems to testify the interest of the EU in exploring the perspective of monetary complementarity.

National and International Money between Complementarity and Substitutability

As mentioned above, the first and most significant form of complementarity is between national (local) and international (global) money. This articulation is strategic since it allows a degree of freedom in domestic economic policy, by avoiding

Mundell's trilemma, which, in a context of free capital movements, implies incompatibility between fixed exchange rates and autonomous monetary policies.

Historically this sort of complementarity has applied both to the unit of account and to the means of payment. The "écu de marc" was the unit of account of the international Lyon fairs in early modern Europe, and it was distinct from the local units of account of individual nation states and even of Lyon as a French commercial place. Coins of different metals served different exchange circuits: gold coins international trade, silver coins national commerce, and copper coins local exchange (Boyer-Xambeu et al. 1994).

With the financial revolution, which fixed an immutable correspondence between the unit of account and a certain amount of precious metal, the gold standard introduced the store of value function but abolished the distinction between international and national money. Gold became the ultimate money (means of settlement) for all sorts of payments at all levels.

Since the gold standard, strictly understood, implies the demonetization of all other metals and the immutability of the metal parity of the unit of account, it is a structurally deflationary monetary system. There is in fact no guarantee at all that the growth of the stock of gold will go hand in hand with the growth of real transactions. Moreover, the abolition of the distinction between internal and external money implies that, while a commercial surplus can be sterilized through hoarding, an external negative imbalance must lead to internal deflation and unemployment. A degree of freedom is thus lost.

The deflationary character of the gold standard has been observed, and viewed as a shortfall, since the beginning of the system, and it has been dealt with by the introduction of convertible banknotes. The simultaneous invention of the gold standard and of paper money, with the English Financial revolution, created a new dual currency system based not on complementarity but on *substitutability*. The system was extended on a global scale first informally with the pre-World War I "gold-sterling standard" (De Cecco 1974) and then formally with the gold exchange standard in the 1920s and finally with the gold-dollar standard in the Bretton Woods period.

It is in relation with the latter that a dilemma has been highlighted, which lays bare the relationship of substitutability between gold and exchange. In fact, if convertible paper money were fully backed by gold, the *substitution* with gold would be perfect, but the deflationary character of gold would be transferred to paper; if instead the backing were not full, (national) paper money would circulate as a*parallel currency* to (international) gold but bearing the risk of an incomplete substitutability.

It was Robert Triffin, one of the advocates and inventors of the European Payments Union, who brought to light the trade-off, and hence the potential conflict, inherent in a system based on the identification of international money with a national currency. And it was this conflict that exploded in 1971 with massive arbitrages between a formally convertible dollar and by then insufficient gold reserves: this brought to the end of the Bretton Woods system. The formal identity between assets in gold and assets in dollars, guaranteed by *substitutability*, first

prevented but finally unchained *substitution*, opening a new era characterized by a national fiat money serving as international money, but at the cost of an increasing instability. In fact, the decoupling of the dollar from gold, while it cuts off at the source the trade-off, increases the risk of structural global imbalances.

The history of international money in its relationship with national money seems to draw a univocal trajectory that goes from complementarity to substitutability and finally to indistinctness. But there is nothing univocal nor irreversible in it. The discoverer of the dilemma, Triffin, was also the advocate of the reintroduction of a separation between international and national money, through the introduction of the "special drawing rights" (SDRs). SDRs may be described as a reserve asset, i.e., as an international currency, issued by the International Monetary Fund.

Their potential challenging role vis-à-vis the dollar has severely limited their use as international money, even if in 2009, in the wake of the global crisis, a more systematic and massive use of SDRs has been advocated by the governor of the central bank of China, Zhou Xiaochuan. His stance was at the origins of the decision to proceed to a new general allocation of SDRs on August 2009 (bringing their total amount from around 20 to 204.1 billion). Nonetheless, the role of SDRs as an international currency has remained marginal and is far from threatening the status of the dollar asglobal money.

In his speech of 24 March, Zhou did not only mention SDRs, but he also evoked the Triffin dilemma as a major risk for global stability and growth, deriving from the use of a national currency as an international currency. He even explicitly mentioned the Keynes plan of an International Clearing Union at Bretton Woods as a more radical and, most importantly, as a cleverer way to deal with the dilemma itself.

Indeed, if SDRs can be seen as a way to *exit* the dilemma, the ICU would have been a way *not to enter* it, since it was wholly grounded on a strict complementarity between the national currencies of member countries and the international unit of account of the Clearing Union (bancor). The complementarity was strict because of their mutual inconvertibility. The bancor was to be created by overdraft facilities granted by the ICU and to be destroyed by the actual use of bancor balances to purchase goods on the international market. As already said, most of the features that now characterize local clearing unions were projected by Keynes for the ICU, namely, the growing symmetrical charges on assets and liabilities in bancor (what he significantly called the "Gesell clause"). The complementarity between bancor and national currencies is further strengthened by the peculiar exchange rate regime designed by Keynes for the Clearing Union: the exchange rates were supposed to be fixed but adjustable. Even here, a perfect symmetry was supposed to hold between persistent deficits (that would be cured through devaluations) and persistent surpluses (that would eventually be corrected through revaluations).

These features should have allowed, in the intention of Keynes, to regain that degree of freedom in domestic policies that the other modern systems had lost, without renouncing the objective of a sustained and balanced growth in international trade (Amato and Fantacci 2014: 92–95).

Cryptocurrencies Versus Complementary Currencies

As already observed, in modern history, it is in periods of financial crisis that monetary heterodoxy gets stronger, as a consequence of a fall in the confidence in the monetary dogmas. But the candy-colored army of reformers hosts cranks as well as brave heretics. What they have in common is the need of an alternative solution to the problems raised by orthodoxy. What makes the difference is the way in which they imagine the concrete feasibility of the alternatives they promote.

If the dream of many utopists has often been a society without money, the goal of many contemporary radical "alternativists" is a money issued and circulating without the operation and control of third parties, i.e., *lato* sensu without any state monopoly, i.e., a private money.

Even if most complementary currencies involve some form of decentralization and democratization of the issuing procedure, complementarity and alternativism in this narrower sense are not the same thing. Yet, today, it is "cryptocurrencies" that tend to monopolize the debate or at least to hegemonize it. At present, almost 1000 cryptocurrencies are competing in a rather opaque market, even though opacity did not prevent huge investments from being made in cryptocurrencies and blockchain technology. There are clearly many differences between cryptocurrencies, not only in terms of economic weight but also as to their design and functioning.

If, however, we look at the most famous and widespread cryptocurrency, bitcoin, what we see is a currency that claims to be a radical alternative to official, publicly issued currencies and that at the same time is closely linked to them: in other words, a currency that is not complementary but substitutive.

As a peer-to-peer money, in which payments are no longer intermediated by third parties, bitcoin has the presumption of representing a radical alternative to official money, i.e., legal tender issued by central banks. However, the fact that bitcoin is traded against national currencies on exchanges according to a fluctuating market price, reflecting relative supply and demand, implies that bitcoin is not a complementary, but a substitutive currency.

The true novelty of cryptocurrencies, therefore, lies not in the logic, but in the technology that regulates their issuance. The operation of an algorithm substitutes the action of the central banker, and the use of a distributed ledger for transactions in bitcoin rests upon the operation of "miners," who compete to assure the smooth and secure operation of the payments system, avoiding fraud and double spending (Nakamoto 2008). The "blockchain," as a bookkeeping and payment technology, strictly intertwines the technological and economic aspects of bitcoin, with the perverse effect of inextricably linking together an unsafe conception of money and a revolutionary technology. Most of the rather unproductive efforts made after the publication of Nakamoto's protocol are to decouple technology and economic meaning of cryptocurrencies.

Bitcoin is by construction a volatile "item." That means that its exchange rate with official currencies can neither be fixed by agreement, since there is no public issuer, nor be referred to a "fundamental," since the issuing has no countervalue but the activity of mining. These two negative features undermine the possibility of

seeing bitcoin as money proper or even as a financial asset. As to this second feature, in fact, bitcoin is an asset to which no liability corresponds, i.e., no private issuer whose solvability should work as a fundamental. It has been said that this characteristic assimilates bitcoin to gold. But in the gold standard, the relationship between gold coins and abstract money of account was *a publicly managed one*. Thus rather than electronic currency, rather than an asset class, bitcoin resembles a "virtual tulip." But in a sense, this is a positive characteristic, if bitcoin has to be regarded as a pure object of speculation: something that is acquired not for its use value but for its exchange value, i.e., uniquely with a view to selling it to obtain a surplus. In this sense, bitcoin is the perfect embodiment of the notion of "mobile capital" (Amato 2018).

We can easily see that all these characteristics distinguish radically this sort of alternative currency from complementary currencies. Being the perfect (ideal) commodity, bitcoin is on the opposite scale compared to complementary currencies, the aim of which is not to be a commodity, but that which, by its circulation, makes it possible to have real transactions in real commodities. If there is anything that characterizes all complementary currencies, it is the fact that they are not intended to be compatible with speculation, i.e., that they are connected with a more or less radical reduction of their function as a store of value, i.e., as a financial asset.

Another way to make this point is by referring to the quantity. In order to be independent from any "political" third party, in the assumption that they are all arbitrary, bitcoin, as all cryptocurrencies, is issued in a predetermined quantity. This feature, however, sets bitcoin away from the needs of a real economy, in which the growth of the monetary mass should accommodate the growth of transactions, and it makes it increasingly scarce, strengthening its commodity character. The aim of complementary currencies is, on the contrary, to accompany the rhythm of real exchanges. In other terms, while bitcoin claims its monetary status on the basis of its (scarce) quantity, complementary currencies aim at performing their monetary function in virtue of their circuit velocity. In most forms of complementary currencies, and certainly in the case of clearing circuits, the creation and the destruction of money are determined by the course itself of real exchanges.

Instead, it is clear that what at least complementary clearing circuits and cryptocurrencies have in common is the need to rest on the electronic traceability of transactions. But this a question of technology and not of institutional construction of money.

Conclusions and Perspectives for Further Research

Complementarity has its roots in the symbolic nature of money. Money is a complex institutional structure, and not simply a convention or an evolutionary phenomenon, and the way it is constructed affects its economic and social functioning. Therefore, complementarity – just as its logical opposite, namely, substitutability – depends on the institutional construction of money. Different currencies circulating side by side

are complements or substitutes according to how those currencies are conceived and established.

To put it bluntly, money may be conceived in two radically different ways. It may be apprehended (indeed, quite literally, grasped) as a commodity, as a store of value, as a form of wealth, as something worth being acquired and accumulated for its own sake, as if it had an intrinsic value. Or it may be understood and treated as "a mere intermediary, which passes from hand to hand, is received and dispensed, and disappears, when its work is done, from the sum of a nation's wealth" (Keynes 1923).

If money is conceived as a commodity, it necessarily acquires the feature that, according to the standard economics textbook definition, characterizes all commodities as such, namely, their substitutability. If, instead, money is conceived as a mere intermediary for the circulation of actual goods, it is characterized by complementarity.

Therefore, the complementarity between different currencies is strictly related to the complementary nature of money with respect to the goods that it can be exchanged with, i.e., to the fact that money *faces* the goods and services produced in order to be sold against money (Fantacci 2008: 72n). Consequently, a well-constructed money is intended not to be hoarded, but to circulate, literally to disappear in circulation. This is the basic meaning of the quantitative equation of money and, less explicitly but not less substantially, of the equilibrium conditions in a decentralized monetary economy.

On the other hand, however, both economic theory and economic practices tend to deal with monetary plurality in terms of free negotiations on a market. Being a commodity, money has its own market, governed by the same rules of all other market. In this sense, market competition holds sway in the field of monetary plurality.

In fact, the distinction between complementarity and substitutability is not so clear-cut as our previous arguments may seem to imply. The ambiguity is already in the underlying notion of money: how long may a means of exchange be withheld from circulation before it becomes a store of value? It has been correctly stressed that the sharp opposition between complementarity and substitutability hinders a correct understanding of what is at stake with monetary plurality, and it has been suggested that this opposition should be "unpacked" (Blanc 2017).

Here we propose to unpack the opposition by stating that, although substitutability does indeed exclude complementarity, it may nonetheless appear to preserve a relationship with it: indeed, substitutability may be viewed as the symptom of the malfunctioning of complementarity.

Complementarity is an effective way to deal with monetary plurality when it deploys itself as a plurality of *non-superposing* monetary circuits. Hence, the way the circuits are designed makes a real complementarity possible and viable. This has been observed for many non-European cases, but it is also the normal case for Western monetary history until the rise of the gold standard (Kuroda 2008a). A qualitative differentiation of monies, differing either by purpose or by locality, implies that, in order for the monetary system to be well regulated, there is not a

single optimum quantity of money, but each money has to be adequate to the circulation needs of each circuit.

The normal case, of course, admits abnormal functioning: there is the possibility, better the risk, that a single complementary currency may not be adequate to the requirements of the economy in which it is supposed to circulate. Frictions may arise either when a single money is too scarce or too abundant. The result is the emergence of (often illegal) *substitution practices* that give place to a "secondary market" along the lines of the Gresham rule (qv). Substitutability appears historically as a degeneration of complementarity, caused either by necessity (as in the case of the over-issuing of vellon in sixteenth-century Spain) or by political frictions (the friction between large coins and small coins in the case of the Ciompi riot).

In the present-day context of a growing plurality of local and targeted currencies, this tension is the main criterion to assess the soundness of a complementary architecture. Once again, the significance of a complementary currency depends strategically on the way the circuit in which it has to *circulate* has been designed. Here we find, once again, the "equilibrium clause": behind the soundness of the complementarity between monies lies the complementarity between money and the goods and services that it is supposed to buy.

Cross-References

- ▶ European Monetary Integration
- ▶ International Currencies in the Lens of History
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Monetary Regimes: The Gold Standard
- ▶ Monetary System of the "Ancient Régime" (Third to Eighteenth Centuries)
- ▶ Privately Issued Money in the United States

References

Amato M (2018) The bitcoin or the reality of a waking dream, in Pixley J, Flam H (eds) Critical junctures in mobile capital. Cambridge University Press, Cambridge, pp 268–276

Amato M, Fantacci L (2012) The End of Finance, Polity Press, Cambridge

Amato M, Fantacci L (2014) Saving the Market from Capitalism, Polity Press, Cambridge

Amato M, Fantacci L, Papadimitriou DB, Zezza G (2016) Going forward from B to A? Proposals for the eurozone crisis. Economies 4:18

Blanc J (ed) (2000) Les monnaies parallèles. Unité et diversité du fait monétaire. L'Harmattan (Economiques), Paris

Blanc J (2011) Classifying "CCs": community, complementary and local currencies' types and generations. Int J Commun Curr Res 15(spl issue):11–16

Blanc J (2017) Unpacking monetary complementarity and competition: a conceptual framework. Camb J Econ 41(1):239–257

Bloch M (1954) Esquisse d'une histoire monétaire de l'Europe. Cahiers des Annales, vol 9. A. Colin, Paris

Boyer-Xambeu MT, Deleplace G, Gillard L (1994) Private Money and Public Currencies: The Sixteenth Century Challenge, ME Sharpe, Armonk NY
Braudel F (1979) Afterthoughts on material civilization and capitalism. Johns Hopkins University Press, Baltimore
Cameron R, Neal L (1989) A concise economic history of the world: from Paleolithic times to the present, Oxford University Press
City of London (2011) Capacity Trade and Credit: Emerging architectures for commerce and money. Report prepared for the City of London Corporation, ESRC and Recipco, December, City of London Economic Development, London
Cœuré B (2014) Life below zero: Learning about negative interest rates, Presentation at the annual dinner of the ECB's Money Market Contact Group, Frankfurt am Main
Day ACL (1954) The Future of Sterling, Oxford University Press
De Cecco M (1974) Money and empire: the international gold standard, 1890–1914, Blackwell, Oxford
Einaudi L (1953) The theory of imaginary money from Charlemagne to the French revolution. In: Lane FC, Riemersma JC (eds) Enterprise and secular change: readings in economic history. George Allen and Unwin, London, pp 229–261
Fantacci L (2005) Complementary currencies: a prospect on money from a retrospect on premodern practices. Financial Hist Rev 12(1):43–61
Fantacci L (2008) The dual currency system of renaissance Europe. Financial Hist Rev 15(1):55–72
Fantacci L (2013) Reforming money to exit the crisis: examples of non-capitalist monetary systems in theory and practice. In: Pixley J, Harcourt J (eds) Financial crises and the nature of capitalist money: mutual developments from the work of Geoffrey Ingham. Palgrave Macmillan, Basingstoke, pp 124–147
Fisher I (1933) Stamp scrip. Adelphi, New York
Guyer J (2012) Soft currencies, cash economies, new monies: past and present. Proc Natl Academy Sci 109(7):2214–2221
Hayek FA (1990) Denationalisation of money: an analysis of the theory and practice of concurrent currencies, 3rd edn. Institute of Economic Affairs, London
Helleiner E (2003) The making of national money. Territorial currencies in historical perspective. Cornell University Press, Ithaca/London
Ingham G, Coutts K, Konzelmann S (2016) Introduction: "cranks" and "brave heretics": rethinking money and banking after the great financial crisis. Camb J Econ 40(5):1247–1257
Kennedy M, Lietaer B, Rogers J (2012) The people money: the promise of regional currencies. Triarchy Press, Devon, UK
Keynes JM (1923) Tract on Monetary Reform, The Collected Writings of John Maynard Keynes, Vol. IV, Cambridge University Press, 1971
Keynes JM (1936) The General Theory of Employment, Interest, and Money, The Collected Writings of John Maynard Keynes, Vol. VII, Cambridge University Press, 1973
Kregel I (2017) A two-tier eurozone or a euro of regions? Public Policy Brief, No. 144, September 2017
Kuroda A (2008a) What is the complementarity among monies? An introductory note. Financial Hist Rev 15(1):7–15
Kuroda A (2008b) Concurrent but non-integrable currency circuits: complementary relationships among monies in modern China and other regions. Financial Hist Rev 15(1):17–36
Labrot J (1989) Une histoire économique et populaire du moyen age: les jetons et les méreaux. Errance, Paris
Lietaer B (2001) The future of money. Century, London
Maurer B (2005) Mutual life, limited: Islamic banking, alternative currencies, lateral reason. Princeton University Press
Mauss M (1990) The gift. Routledge, London English translation of *Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques*, L'Année Sociologique (1923–24)
Nakamoto S (2008) Bitcoin: A Peer-To-Peer Electronic Cash System, bitcoin.org

North P (2007) Money and liberation: the micropolitics of the alternative currency movement. University of Minnesota Press, Minneapolis
Pacione M (1997) Local exchange trading systems as a response to the globalisation of capitalism. Urban Stud 34(8):1179–1199
Peacock MS (2006) The moral economy of parallel currencies. Am J Econ Sociol 65(5):1059–1083
Peacock MS (2014) Complementary currencies: history, theory, prospects. Local Econ 29(6–7):708–722
Phelan (2014) The Road Not Taken - The Hard ECU; Britain's alternative to the euro, London School of Economics and Political Science
Polanyi K (1957) The Economy As Instituted Process, in Polanyi K, Arensberg CM and Pearson HW (eds) Trade and Market in the Early Empires, The Free Press, Glencoe, pp 243–269
Richey S (2007) Manufacturing trust: community currencies and the creation of social capital. Polit Behav 29(1):69–88
Seyfang G (2001) Community currencies: small change for a green economy. Environ Plan A 33(6):975–996
Stodder J (2009) Complementary credit networks and macroeconomic stability: Switzerland's Wirtschaftsring. J Econ Behav Organ 72:79–95
Théret B (2016) Réutiliser la monnaie fiscale (avec Wojtek Kalinovski). Les Dossiers d'Alternatives Économiques 6:53–54
Tibbett R (1997) Alternative currencies: a challenge to globalisation? New Polit Econom 2(1):127–135
Zelizer VA (1998) The proliferation of social currencies. In: Callon M (ed) The laws of the markets. Blackwell, Oxford, pp 58–68

Part VI

Asian Monetary Systems

Monetary System in Ancient China

21

Yohei Kakinuma

Contents

Abstract

What was the nature of property exchange in ancient China and how can we describe its historical background? How have people researched the monetary economy of ancient China? To what kinds of topics have we paid attention? How and where can we do our own research on this topic? Based on this chapter, there was not only a continuity but also a great transformation from the Eastern Han period to the Three kingdoms period, so what influence did it have on the whole society? This chapter focuses on these questions and submits five tasks from now on.

Keywords

Coin · Gold · Textiles · Price · Communication

Y. Kakinuma (✉)
Teikyo University, Tokyo, Japan
e-mail: yohey0806@yahoo.co.jp

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_5

Introduction

In a broad sense, communication among people dates back to prehistoric times. Society is constructed over time through various communication methods such as verbal communication, soul communication, and communication through material exchanges. Consequently, if we attempt to reveal the history of a place within a given period, it is necessary to study the specific means of communication in that context. This perspective on history is known as communication history. Among the various communication methods, property exchange is one of the most direct ways of such communication, thus becoming an important clue in the study of history. Moreover, because communication is an ongoing process in modern society, it appears that studying its specific content is conducive not only to understanding the communication characteristics in different times and places and their related social structure but also to making important comparisons among them. Then, what was the nature of property exchange in ancient China and how can we describe its historical background?

Turning Points
15 B.C.E. The Shang period
11 B.C.E. The Western Zhou period
6 B.C.E. The Eastern Zhou period (The Spring and Autumn period)
4 B.C.E. The Eastern Zhou period (The Warring States period)
221–206 B.C.E. The Qin period
206 B.C.E.–220 C.E. The Han period
220 C.E.–280 C.E. The Three Kingdoms period
265 C.E.–316 C.E. The Western Jin period

The answers to these questions are extremely important historically, economically, and sociologically. Much research has been done in regard to these questions, and although previous studies in European languages have been limited, e.g., Tierry (1997, 2017), Peng (2000), and Sheidel (2009), considerable relevant research has been conducted in Chinese and Japanese. As a result, our understanding of the monetary economy in ancient China has rapidly and dramatically changed along with the relatively recent and rapid increase in the excavated historical texts available. For instance, we now have statistics detailing excavated cowry shells prior to the Warring States period (Kakinuma 2011) and excavated bronze coins in the Spring and Autumn and Warring States periods (Mǎ 1988, 2002; Emura 2000, 2011; Huáng 2001), as well as the results of all-inclusive analyses of excavated and nonexcavated historical texts concerning bronze coins, gold, and hemp and silk textiles in the Warring States, Qin, and the Han periods (Kakinuma 2011, 2018). Under these circumstances, how have people researched the monetary economy of ancient China? To what kinds of topics have we paid attention? According to previous research, how and where can we do our own research on this topic? This chapter focuses on these questions.

The Rise and Fall Theory of the Monetary Economy of Ancient China

Accumulation of fundamental studies on ancient Chinese history in academic circles began around the turn of the twentieth century. At the same time, people started to research the monetary economy of ancient China on the basis of not only the academic traditions of epigraphy and numismatics continuing from the time before the Qing period, or evidential research studies developed in the Qing period, but also new methodologies, such as studies on oracle bone inscriptions, wooden or bamboo strips, and archeology. These studies developed rapidly during the twentieth century, and multiple micro-historical studies were conducted based on previous research methods, e.g., lectures between 1925 and 1929 by Kato (1991) and the book by Yoshida (1933).

These basic studies still had problems, because they were based on epigraphy and numismatics that had originated from unscientific personal hobbies and on new unsophisticated methodologies. It is also because they are based on immature studies on oracle bone inscriptions and inscribed bamboo or wooden strips. However, these accuracies had been gradually improved under the influence of the Qing tradition of evidential research studies, with slogans of "Seek Truth from fact" and "Study hard, and think deeply." The Western tradition of source-based history introduced by Leopold von Ranke that was imported to East Asia also supported the new historiographies. Concerning the academic background of Qing dynasty textual research and the Ranke's positivist historiography, there are significant differences of philosophy (Hamaguchi 1994; Iggers 1993), but it cannot be doubt that both of them provided a basis for research on the monetary economy of ancient China.

Additional activities included the Chinese *Shihuo* School in 1930s, which claimed to concentrate on collecting and collating historical materials in detail without any interpretations (Táo 1934), and historical studies on unearthed texts especially developed after the 1970s. As a result, they have further promoted fundamental studies on the monetary economy of ancient China.

On the other hand, many scholars also attempted to integrate the previously mentioned basic research and restore a broad stream of history of the monetary economy of ancient China. They attempted to reconstruct the general evolutive history of the ancient Chinese monetary economy. This should be called "macro historical research," whereas the previously mentioned basic research should be called "micro historical research" which began in China and Japan at the beginning of the twentieth century. It was just an audacious rough sketch, but it could be recognized as a paradigm for young scholars' reference. The paradigm can be called the "Rise and Fall Theory of the Monetary Economy of Ancient China" (hereafter, the "Rise and Fall Theory").

The essence of "Rise and Fall Theory" is that a monetary economy appeared before the Zhou period (especially before the downfall of the Eastern Zhou) approached its peak around the first half of the Western Han period and declined after the latter half of the Western Han period or after the Three Kingdoms period.

This theory generated significant attention, and it was regarded as dominant for a long time, mainly in Japanese historical circles after 1940 (Yoshida 1933; Odake 1935; Quán 1941; Wáng 1947; Makino 1950; Péng 1958; Miyazaki 1964; Lao 1976; Yamada 2000).

From this rough sketch, cowry shell is widely recognized as the oldest money and originated in the Shang period (Hamada 1912; Wáng 1921, 1957; Yoshida 1933; Jiǎ 1976; Lao 1976; Dài 1981; Fù 1980; Xiao 1984; Cài 1996; Yáng 2003; He 1948), before the Shang period (Zhū 1984; Liú and Lǐ 1995; Huáng 2001), from the end of the Shang period to the beginning of the Western Zhou period (Péng 1958; Guō 1933), in the first half of the end of the Western Zhou period (Kondo 1998), or in the Spring and Autumn and the Warring States periods (Egami 1967). Some also recognized bronze ingot as an original means of economic exchange (Matsumaru 1992).

On the other hand, it is widely said that gold and bronze coins were used as money after the Zhou period. Moreover, gold mainly started to be used as vessels and decorations with the spread of Buddhism after the Eastern Han period and the productivity of bronze decreased; thus, instead of gold and coin, bolts of textiles and grain became main money after the Wei and Jin periods (Yoshida 1933). In addition, Quán (1941) recognized the decrease of copper as a reason for the decline of the monetary economy and explained its details as follows. The monetary economy reached its peak in the first half of the Western Han period and declined after the end of the Eastern Han period due to four reasons: (1) economic confusion due to wars; (2) decrease of population; (3) decrease of copper supply; and (4) increase of copper consumption in making statues of Buddha (Quán 1941). Makino also stated a similar theory from a different perspective. In the *Treatise on Food and Money in Hanshu*, there is an economic policy submitted by Li Kui during the Warring States period that is premised on ordinary people calculating their living expenses by bronze coins without which they could not live. Therefore, the monetary economy deeply permeated into farmers' lives during the Warring States period (Makino 1950). However, according to Makino, the monetary economy during the Han period was influenced by a shortage of copper. Especially after the latter half of the Western Han period, people started to hoard coins, the productivity of coins decreased and, as a result, the monetary economy declined (Makino 1953). Similarly, Miyazaki (1964) realized the Warring States, Qin, and Han as periods of economic boom and the periods from Wei to Tang as periods of economic recession. This is because bronze and gold (especially gold) flowed out from the dynasty from the end of the Western Han period. This resulted in money shortage and a regression of the monetary economy occurred (Miyazaki 1964).

This Rise and Fall Theory based on the shortage of bronze or gold after the Eastern Han period has been supported by some scholars thus far. Some still exaggerates that the monetary economy declined due to the shortage of copper (Inaba 1984, 1985), whereas some believe that the monetary economy declined because of the shortage of gold (Péng 1958; Lao 1976). Some even believe that the shortages of both copper and gold caused the decline in the monetary economy (Yamada 2000), which explains the following: bronze coins, gold, and bolts of

textiles were recognized as money in the Qin state during the Warring States period, but bolts of textiles lost their monetary function later and the parallel standard system in which gold and coins had a fixed exchange rate started to function. Under these circumstances, gold was gathered in the hands of the emperor, the emperor's family, and merchants because it is too expensive to use in daily life for ordinary people. In the latter half of the Western Han period, the shortage of coins and outflow of gold became more serious and people attempted to collect coins with goods. Such people illegally shaved the coins and used the shavings of bronze to mint new poor-quality coins. Bolts of textiles again circulated as money instead of coins and gold. Under these circumstances, the duality of monetary circulations arose, i.e., upper-class people used bolts of textiles, gold, and silver while ordinary people used poor-quality coins. Wang Mang fixed the ratio between gold and coins, collected gold, and attempted to stabilize the value of coin on the basis of the gold stock. Moreover, he proceeded with the nominal monetization of coins and minted more coins to resolve the shortage problem, but failed. The Eastern Han resurrected the collection of taxes by coins but it was not enough to recover the monetary economy. As a result, the monetary economy, which reached its peak in the first half of the Western Han period, gradually declined after the latter half of the Western Han period. Concerning the reason why gold disappeared in the Eastern Han period, there are other hypothetical explanations, such as Fù (1980), but Fù (1980) is just a variation of Rise and Fall Theory.

These theories jointly own the model or at the least a part of the model that stated that the monetary economy reached its peak in the first half of the Western Han period and then gradually declined. This can be evaluated as a pioneering experiment by which the importance of cowry shells in the Shang and Zhou periods is known, the importance of coin and gold in the Warring States, Qin, and Han periods are pointed out, and related historical texts are gathered and analyzed in order to grasp the historical change of the monetary economy in ancient China. However, this model also has a lot of points we need to reconsider.

Criticism Against the "Rise and Fall Theory"

A first criticism against the "Rise and Fall Theory" weighs on whether the monetary economy declined after the middle of the Western Han period. Regarding this, two theories have appeared: (1) the monetary economy still flourished during the Eastern Han period, and (2) there was a natural economy before the nineteenth century.

The first theory can be called the "Eastern Han monetary economy theory." According to this theory, a monetary economy or commercial activities flourished not only in the Western Han period, but also in the Eastern Han period (Tada 1965; Ebrey 1986; Kamiya 1993). However, these previous researches neither consider how its monetary economy worked in detail nor reexamine the economic situation after the Eastern Han period. Therefore, the dominant perspective still focuses on a natural or barter economy after the Eastern Han period. Kawakatsu (1982) also discussed that the duality of monetary circulations arose in the Southern dynasties

period, i.e., upper-class people used bolts of textiles, gold, and silver while ordinary people used poor-quality coins; however, there is still the absence of research of monetary economy in the Wei and Jin periods, i.e., between the Eastern Han period and the Southern dynasties period. For this reason, Kakinuma (2018) recently examined the actual conditions of the Chinese monetary economy between these periods in detail. It will be more discussed later.

The second can be called the "pre-modern natural economy theory." It was mainly insisted by scholars in Chinese historical circles. It proposes that the pre-modern Chinese economy was a natural economy. Here, the so-called "natural economy" is a notion that includes the self-sufficient and barter economy and opposes the commodity and monetary economies. Although there are many theoretical variations in Chinese historical circles, most argue that there were no significant economic differences among the Han, Three Kingdoms, and Jin periods (Ye 1994). Moreover, like the Japanese historical circles, they especially underestimate the monetary economy after the Western Han period.

These two theories are completely different ideas but both are critical against the previous theory, which exaggerated by far the difference between the economy in the Western Han and the economy after the Eastern Han period. The first theory urges us to reconsider the Eastern Han economy, whereas the second theory has become dominant in Chinese academic circles as it was is based on Marxism, which has been a political thesis of the People's Republic of China.

In Chinese academic circles, however, people also started to re-examine the notions of natural and commodity economies after the 1980s. For instance, some criticized the theoretical ambiguity that the so-called "natural economy" includes the self-sufficient and barter economy. They redefined natural economy as sufficient economy, and recognized the whole premodern feudalist society as not natural economy but commodity economy (Jing 1987). According to Jing (1987), here the notion of commodity economy should include barter economy and monetary economy. This way of criticism is similar to the way in which Max Weber criticized Bruno Hildebrand. Hildebrand (1864) attempted to explain the Western economic history by evolutional schema from natural economy to monetary economy, whereas Weber (1924) criticized its conceptual ambiguity of "natural economy." Lín (1997) approved to distinguish self-sufficient economy into barter economy, but he disagreed Jing's idea of "barter economy + monetary economy = commodity economy = the feature of the whole pre-modern feudalist society." According to Lin, if Jing's theory stands up, the whole premodern feudalist society should be recognized as commodity economy, which includes both barter and monetary economy. Then, when is the period of natural economy or self-sufficient economy? Can we really find such a period in Chinese history? Lín does not think so, because ancient historical evidences of barter and commodity exchange are found everywhere. Thus, Lín (1997) proposed that natural economy should be defined as "self-sufficient economy supplemented by commodity exchanges." This idea has been developed by some scholars (Fāng 2004; Guō 1998; Huáng 2005).

As can be seen from the previous discussions, in Chinese academic circles, the premodern natural economy theory has a wide variety. However, it is also

obvious that the premodern natural economy theory and its variants are still dominant in China.

So, which theory is the best? Here, it should be noticed that notions such as natural economy and monetary economy were conceived in the Western academic tradition and, as such, they tend to categorize economic situations in comparison with Western economic history. From this perspective, every single society in Chinese history prior to capitalism must be understood as an immature monetary economy. Thus, it is difficult to problematize the special characteristics of the economy of ancient China. To resolve this criticism, five theories recently became powerful. They include both empirical and theoretical criticism against the "Rise and Fall Theory" and the "Pre-modern natural economy theory." Each theory was born on different dates, including some before the 1950s. However, the following five theories include new perspectives to overcome the dichotomy between the natural economy and the monetary economy, and they are still tenable until now.

First, the idea that the monetary economy of the Han period declined due to a shortage of copper is dubious. For instance, Kimura and Adachi admitted that ancient commerce declined during the Wei and Jin periods but believes that the reason behind this was the abolishment of a tax system by coins and not the shortage of copper. This premise of this theory is that people used coins during those days because the government urged people to use them. In other words, coin was just a means of national settlement (Kimura 1955, 1960; Adachi 1990). This theory emphasizes that the circulation and use of bronze coins is not an expression of spontaneous commodity fetishism, but a result of state compulsion, e.g., collection of poll tax in the form of coins. If so, these coins are not the so-called "money" used by people because of its economic liquidity and convenience. By pointing out essential differences between the ancient Chinese coin and the modern coin, this theory criticizes the previous theory, which had restricted the rise and fall of the monetary economy based solely on quantity levels. Here, the new problems are (1) whether the copper shortage really occurred and (2) how did the productivity of copper relate to the monetary economy. We should examine these topics on the basis of historical texts.

Second, the "Rise and Fall Theory" does not look at the relation among the different types of money. Some scholars focus on the relationship between different types of money such as coins and gold. While looking into the relationship among various co-exiting forms of money, it not only discusses the ups and downs of the monetary economy, but also reveals the essential change during its period base on the peculiar variety of money. For instance, Sekino (1956, 2005) pointed this perspective out and explained the historical change of the monetary economy in the Warring States, Qin, and Han periods as follows: bronze coin with a material value corresponding to the exchange value was introduced in the Warring States period. However, in c.3, or 4 B.C., when tools and goods made by wrought iron appeared, the demand for tools and goods made of bronze decreased, and the material value of bronze gradually declined. The nominal monetization of bronze coin also occurred, which furthered its decline in value. For this problem, the Western Han government monopolized gold under the institution of bimetallism of

gold and coins, through which the government attempted to maintain the monetary value of coin but failed. Thus, the Han government minted coins, banned private coin, and attempted to stabilize the monetary economy through these policies (Sekino 1956, 2005). This theory premises that there was adequate copper available at the beginning of the Han period and, in this sense, this theory can be recognized as an anti-Makino theory. However, as with Miyazaki (1964), it also paid attention to the shortage of gold and thus it can be recognized as a variation of the "Rise and Fall Theory." However, it should be noticed that Sekino (1956, 2005) did not simply examine the quantities of coin and gold but grasped the relation between them and explained it as follows: "nominal monetization of bronze coin→depreciation of nominal bronze coin→reconstructing the monetary economy by the governmental monopolization of gold→the shortage of gold→depreciation of nominal bronze coin→recurring to bronze coin whose material value is the same as the nominal value." Here, Sekino (1956, 2005) does not conclude the decline of the whole monetary economy after the latter half of the Western Han period. As Kakinuma (2011) discussed, it is dubious that there was bimetallism with a fixed exchange rate between gold and coin and thus Sekino's idea is not tenable anymore. However, the perspective to the relation among multiple types of money is important. In this sense, Sekino (1956, 2005) left a large fortune to us.

Third, the "Rise and Fall Theory" does not acknowledge regional differences between economies. He Ziquan's research about the economy in the Wei, Jin, and the Northern and Southern dynasties periods states that a self-sufficient farmer-oriented economy flourished in northern China, whereas the monetary economy developed in southern China (He 1948). In Japan, Jitsuzo Kuwabara (1925) is a pioneering figure to point out the regional difference between northern and southern China, and Fumio Okazaki mentioned the monetary economic difference between the north and south from the Wei to Sui periods (Okazaki 1932, 1935). Yoshio Kawakatsu (1982) also explained the development of a monetary economy in the southern dynasties, which resulted in two problems: the shortage of copper and inflation. Such a regional difference of economy from the Wei to the Sui periods has been hypothesized by many scholars; however, theories after that of Hé (1948) especially have a doctrinal significance, i.e., he used his theory as a criticism against the "Rise and Fall Theory." In recent years, Wang Yichen has minutely portrayed the periodical and regional differences of the monetary economy from the Wei to the Sui periods (Wáng 2007). This so-called "theory of regions of monetary history" has been considered important by scholars of the Spring and Autumn and the Warring States periods as well. This research method is mainly to analyze the shape of coins, characters on the surface of coins, and the ingredients of coins by using numismatics and to find out how coins were discovered (Mǎ 1988, 2002; Emura 2000, 2011; Huáng 2001; Chén 2006, 2008). These methods made great strides, especially in terms of the Spring and Autumn, Warring States, Wei, Jin, and the Northern and Southern dynasties periods because one feature of these periods is regional divisions; therefore, it is relatively easy for scholars to examine the economic regional differences of these periods. However, it gradually becomes easy for us to depict the regional differences in the Qin and Han periods, when the unified empire governed,

because the number of excavated texts detailing regional histories has rapidly increased.

Fourth, the "Rise and Fall Theory" realizes coins and gold as the main components of the monetary economy and does not count the existence of non-metal money as money. This criticism considers not only bronze coin, gold, and other metal money, but also cloth (linen fabric), silk, grain, and other tangible goods as elements of a monetary economy. It argues that the previous "Rise and Fall Theory" of ancient Chinese money only bases its calculations on the increase or decrease of metal money. The same criticism on Western historiography was submitted by Marc Bloch, who recognized not only metal money but also black pepper circulating in the market as money during the Middle Ages in Europe. This new understanding became dominant and surely deconstructed the dichotomy between natural economy and monetary economy (Bloch 1939). When we think about the definition of money, Bloch certainly have a point. Generally speaking, people cannot directly exchange commodities with each other, and when two people exchange goods, each of them should have a good that the other person lacks. It is a seldom occurrence, widely called a "double coincidence of wants." However, with money, we can buy the commodities we want. A commodity's value is measured by the money. Money is a means to purchase commodities and is desirable because of its convenience. As far as monetary circulation is maintained, people do not have to worry about whether money is accepted. In this sense, there is no need to confine money as metal money. Furthermore, such a distinction between money and a commodity, strictly speaking, is not very rigid. For instance, the currency in each country cannot be used outside its own country even today. Conversely, regional currency whose interchangeability is restricted is also recognized as money. Thus, the range of the interchangeability of money as an index between money and commodity is actually ambiguous. Hayek (1976) pointed out this ambiguity a long time ago, and stated that it has been rather a misfortune that we describe money by a noun, and that it would be more helpful for the explanation of monetary phenomena if "money" were an adjective describing a property which different things could possess to varying *degrees*, i.e., whether it circulates a lot, at the most. In this sense, there is no need to confine money to metal money. Based on these discussions, besides metal, various other items could have functioned as money. Once nonmetal money is understood as money, the dichotomy between the barter economy and the nonmetal monetary economy soon also becomes ambiguous. For instance, Wǔ (1937) states that metal money did not circulate well in the northern dynasties and bolts of textiles circulated as money instead. Péng (1958) believed that grain sometimes functioned as money and labeled the pre-Shang period as a premonetary economy, the period between the Shang to the Warring States as the beginning of the monetary economy, and the period after the Warring States as a time of prosperity of the monetary economy. Here, Wǔ (1937) did not directly intend to criticize the dichotomy between the natural and monetary economies and Péng (1958) supported the "Rise and Fall Theory" although he recognized grain as money. However, these theories can also be used as a tool for criticizing the "Rise and Fall Theory" and the "Pre-modern natural economy theory."

Fifth, the "Rise and Fall Theory" includes an idea that the decrease of the quantity of money resulted in a decline of the monetary economy; however, it has a problem from an economic perspective. Kageyama (1984) has stated that if there is the shortage of gold and coin, their values will rise, gold and coin stored personally will flow back, and the outflow of gold and coin will be automatically restricted. Based on this criticism, Kageyama (1984) suggested to focus on not monetary quantity, but players of monetary economy, such as businessmen, landed gentry, peasant farmers, etc. While discussing the changes of the monetary economy through history, this theory also focused on the main players of the monetary economy, which had never been discussed before in depth.

These previous five criticisms against the "Rise and Fall Theory" are durable until now and are quite important. Thus, how can we do our research on the basis of previous criticisms?

First, according to the previous theoretical discussions, a new perspective or frame of reference is needed. Since 1990, some scholars have focused on the "economic anthropological multi-money theory," which absorbs the economic anthropology achievements of the Karl Polanyi School. This theory criticizes the dichotomy of the "Rise and Fall Theory," in other words, the "monetary economy or natural economy." It is insightful to consider temporal and regional qualitative changes of the history of the ancient Chinese monetary economy in perspective. This theory also exaggerated that the economy was embedded in society. However, this is just a simple proposal. It is necessary for us to deeply consider the meaning of embeddedness. Besides, paradoxes can be found, because this theory combines economic anthropology and history, two very different subjects, in an improper way.

Second, under the circumstances that the quantity of new excavated texts (e.g., bamboo and wooden strips) has sharply increased, these excavated texts should be referred to. Recent anthropological discussion about monetary economy of ancient China has not fully reflected the achievements of "micro-historical research" based on the rapidly developing unearthed text data research in recent years. Therefore, it should be indicated that under the present study conditions, the history of the ancient Chinese monetary economy requires new "macro-history research."

Origins of Concurrent Currencies in Ancient China

On the basis of previous criticisms, Kakinuma recently published two books in Japanese on the monetary economy of ancient China (Kakinuma 2011, 2018). Let me introduce a part of their contents.

Kakinuma (2011) firstly analyzes the condition of the special Cowry used in the Shang and Zhou periods which is formerly acknowledged as the oldest Chinese money. He points out that the misinformation of the actual situation results in the mistaken view of "Cowry = money" emerged in the Warring States period. The royalty of Shang and Zhou periods collected valuable Cowry along the route of "(South China Sea→) Southeast Coast → Tribes along Huai River → Central Plains," and assigned them in the form of "Juan (the Cowry string hanging round

the neck)" to each submissive and adjunctive clan to maintain their relationship. The so called "inscriptions recording cowry reward" reveals this historical fact. According to the inscriptions record, "Shang clan people" especially believed in the gift custom, i.e. through awarding the Cowry that used to symbolize "life and rebirth" to show the prosperity of the bestowed clan. After the middle and later periods of Western Zhou, they finally got rid of the theocracy beginning from Shang period, began to offer official posts by "title-conferring ritual," and cast "inscriptions recording title conferring form" in the hope of realizing the polity of Zhou people on this basis. Resulting from that, "Shang clan people" who had been bearing theocracy were crowded out, and they did not award Cowry any longer to build relationship with the Western Zhou royalty. Instead of Cowry, they awarded other properties according to the title-conferring ritual. This led to sharp decrease in the inscriptions recording Cowry award. However, the "Cowry culture" still remained in various regions after the disappearance of these inscriptions. Moreover, along with the gradual spreading of "Shang clan people," Cowry culture was spreaded all over the country. Especially in Chu State (a state in the Zhou period) and other regions, a new and unique Cowry culture appeared there. Thus, the memory of "Cowry = valuables" was passed down to later generations in an abnormal form, resulting in the generation of the cognition of "Cowry of Shang and Zhou periods = money" during the Warring States, Qin, and Han periods. By revealing the actual situation of the Cowry culture in Shang and Zhou periods, Kakinuma (2011) contests the general view of "Cowry of Shang and Zhou periods = oldest Chinese money" and attempts to outline the characteristics of the economy of Shang and Zhou periods.

Then, how did the monetary economy develop in ancient China exactly? Kakinuma (2011) investigates this issue from a philology perspective, and explores the gradual development process of monetary economy after the Shang and Zhou periods. As far as the methodology is concerned, Kakinuma focuses on property-exchange behavior and the vocabulary describing property-exchange. By tracing the linguistics evolvement and the changes of communication methods in each period, Kakinuma reveals the slow transition of economy from Shang and Zhou periods to Warring States, Qin, and Han periods. The communication before Qin period gradually lost the features of gift communication economy during Spring and Autumn and the Warring States periods in that words indicating gift exchange, such as Chinese characters "賚" and "賃" reduced gradually. From the Warring States period to the Qin and Han periods, the characteristics of the monetary economy tended to be strengthened (Chinese characters of "賣(sell)" and "買(buy)" appeared). The Warring States period was a transitional period from a social formation based on gift exchange to the one dominated by monetary economy. Meanwhile, Kakinuma pays attention to such unearthed text data as Qin bamboo slips of Shuihudi and Longgang, and Han bamboo slips of Zhangjiashan, and it clarifies that terminological distinction between "sell" and "buy," "award" and "repay," and "loan" and "borrow" formed from the end of Qin period to the beginning of Han period. And based on this, Kakinuma points out that this presumably a phenomenon accompanying with the "unified character" advocated by First Emperor of Qin period. It further points out that in such a circumstance, the monetary economy of Warring States, Qin, and Han

periods based on coin, gold, cloth, and silk began to develop. So, how about the definite situation of the monetary economy then? Concerning the monetary economy in the pre-Qin period, Kakinuma (2014) has already explained it in detail in English. Thus, this chapter focuses on the monetary economy after the Qin period.

By the unified Qin times, these multiple currencies must have been widely circulating, even in rural areas. A crucial stimulus was the imposition of a poll-tax that had to be paid in coins, forcing the peasantry to sell their economic surplus for money. This new poll-tax system was a deliberate attempt to expand the use of its currency and dodge the administrative inconveniences and costs associated with in-kind tax payments, which ranged from the need to provide storage to the uneven quality of items tendered. Another crucial stimulus was the policy "men should plow, and women should weave" and "husband should plow, and wife should weave." These policies urged people to produce bolts of textiles with the surplus of production circulated as money in the market.

The shapes and forms of these currencies were standardized. The basic unit of weight of gold in the Qin state during the Warring States period was "yi" (c.250 g) and it was renamed "jin" at the beginning of the Western Han period. There was an equation of weights: 1 jin = 16 liang = 384 zhu. Furthermore, according to the Qin statutes in the Warring States period, bolts of textile as money should have a good quality and uniform dimensions, e.g., length c.185 cm and width c.58 cm. All officials and merchants in shops in market should equally receive coins or bolts of hemp when paid by them. During the Han period, bolts of not only hemp but also silk were standardized. The size of bolts of textiles should be suitable for making clothes and people cannot sell non-standardized textiles. If these textiles were just commodities, they would not need to standardize them or even decide their size. It is, therefore, quite obvious that the government regarded these textiles as a means of payment or measurement. On the contrary, Von Glahn (Glahn 2013) regarded bolts of hemp textiles not as a routine means of exchange in commodity trade but as a means of both public and private payments. Certainly, the "moneyness" of coins was stronger than that of gold or textiles; however, according to Kakinuma (2011, 2018), it is not impossible to find examples of purchasing commodities by not only coins but also gold or textiles. There are examples to exchange clothes for bolts of textiles, textiles for grains, and textiles for gold. Moreover, the number of the examples of commodity exchanges by bolts of textiles had started to increase gradually after the Western Han period and the salaries of the governmental officials were sometimes paid by bolts of textiles. Thus, the officials needed to buy daily goods by bolts of textiles. In this sense, bolts of textiles cannot be identified with other commodities and the moneyness of bolts of textiles periodically changed.

However, in reality, it seems that the versatility of gold and bolts of textile was worse than that of coins in the market because gold and textiles were high-value money. For instance, 1 jin (c.250 g) of gold was equal to several thousands or tens of thousands of coins. Bolts of textiles also had values higher than coins but lower than gold. In order to understand the relationship among money, let us consider the exchange rate of money and the price system a whole in ancient China.

Understanding the Multimonetary Economy of Ancient China

Multiple Money in Market

By studying the price system at that time, Kakinuma (2011) attempts to explain the relationship between the main money – coin, gold, and bolts of textile of the day (especially the rate of exchange). During the Warring States, Qin, and Han periods, in the three-tier price system consisted of "fixed official-price," "Pingjia (i.e., monthly changed official-price)" and "actual price," all properties were included in this price system which takes bronze-coin as a measure of value (Fig. 1).

Gold, bolts of textile, and so on were also included in this system. According to the law in the Qin during the Warring States period, 1 bu. of hemp textile [length c.185 cm (8 chi) and width c.58 cm (2.5 chi)] was equal to 11 coins in "Pingjia (i.e. monthly changed official-price)." Additionally, bolts of textile cannot be cut, i.e., they are not divisible. Thus, gold and textiles were inconvenient when people wanted to buy daily goods but were more convenient than coins when people bought high-value items. Generally speaking, luxuries gather in markets in capital cities; therefore, high-value money seems to have circulated satisfactorily there as well. On the contrary, low-value money circulated in out-of-the-way markets where cheap daily goods were exhibited. This means that the three different kinds of money coexisted and there were regional differences in the way in which each was employed in local markets. This system may be assumed to have been flexible enough to absorb fluctuations in the supply of any of its component currencies, making demonetization improbable.

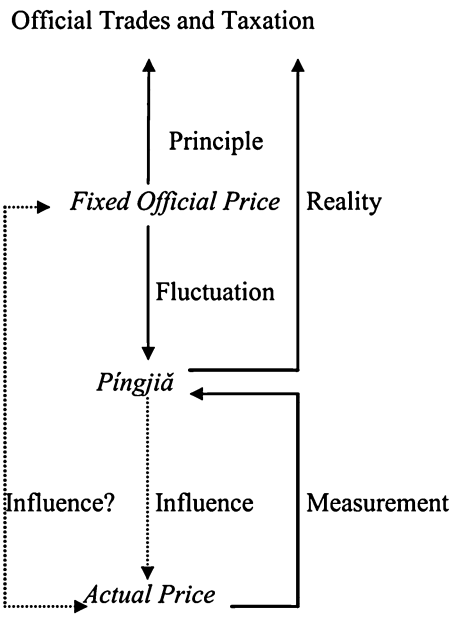


Fig. 1 Price systems during the Qin and Han Periods

Under the circumstances, one of the money of Warring States, Qin, and Han periods, i.e. bronze-coin did not maintain its value through gold standard, cloth standard, or similar standards, but through other elements to maintain its core status in the value system. Thus, when we review the history of recasting the bronze-coin during the first half of the Western Han period, we will find that the state's purpose of just allowing the coins with the same characters (showing its nominal weight) on them to circulate was to maintain the system which measured the value of the commodities by counting coin's quantity (called "counting principle"). On the other hand, the state attempted to further reduce the weight of coins so as to make the legal weight and actual weight of the coins lighter than the nominal weight indicated by the character on coins. However, people tended to divide coins into light ones and heavy ones in folk, resulting in that the light coins were used as money with low value and the populace no longer believed in the characters on the coins. As a result, the feudal officials had to follow public opinion, and then made the character which shows nominal weight on the coins closer to the actual weight. This is the main reason why the coins were often trial casted and recasted during the first half of the Western Han period.

The Social Meaning of Money Outside Market

Coins, gold, and textiles were also used outside of the markets. In the villages and imperial court, gifts, donations, requitals, and bribes were systematic and customary. In addition to coins, gold, and textiles, things like peerages, alcohol, beef, and clothing were also exchanged depending on the time and place. There were set rules for what a person would give and the circumstances in which the gift was made. Table 1 summarizes the use of coins, gold, and textiles in the Western Han, Eastern Han, Jin, and Northern and Southern dynasties periods. The use of these forms of currency was determined by the economic, social, and systemic conditions of each era and each region (Kakinuma 2011, 2018).

Accordingly, coins, gold, and textiles all had the common function of being means for economic distribution and each of them had their unique social functions. The movement of these things was impacted by not only economic rationality but also the institutions and customs of the time. The way in which the money was used changed forms; a similar phenomenon can be observed in other ancient societies. Karl Polanyi focused on this phenomenon and called currencies that were used for different things "special-purpose money" or "limited-purpose money." He also stated that we must be satisfied to simply list the sort of purpose for which "money" is actually used (Polanyi 1957). However, it can be argued that neither economic formalists, who directly apply a neo-classical economic approach to the ancient world, nor the so-called Polanyi school in economic anthropology, who recognize that ancient economy was deeply embedded into society, can fully explain the ancient Chinese world. Aside from the real intentions of Polanyi's ambiguous and sometimes arguable remarks (Gemici 2008), a recent study demonstrated that modern economies are more or less embedded in society (Granovetter 1985) and that

Table 1 Earmarking currencies in ancient China

Situation/period	Western Han	Eastern Han	Jin	Southern dynasties
Gifts from the State to the people on auspicious national occasions	Silk or gold	Silk or gold	Silk	Silk
Condolence gifts	Coins or silk	Coin or hemp	Coins or hemp	Coin and hemp
Retirement bonus	Gold	Coin or silk	Coin or silk	
Get-well gifts	Silk	Coin	Coin or silk	Coin
Prizes for meritorious services in wars	Coin or gold	Coin	Silk	Coin
Bounties for criminals	Coin or gold	Coin or gold	Hemp, silk, gold	
Compensation for forced emigration	Coin	Coin	Coin	
Gifts for foreigners	Gold or silk	Gold or silk	Gold or silk	Hemp and silk
Redemptions of punishment	Gold	Silk	Hemp	
Gifts to invited scholars	Silk?	Silk	Silk	Silk
Presents for wedding celebrations	Coin or gold	Silk	Coin or silk	
Presents for people who worked hard	Silk	Silk	Silk	Silk
Presents for the socially vulnerable	Silk	Silk	Silk	Silk
Farewells	Coin	Coin	Coin	

we can only discuss the different degrees of embeddedness in each societies (Braudel 1979). Modern currencies called "all-purpose money" are also actually used differently depending on the time and place (Zelizer 1994). In this sense, there is room for further examination of the view that money is a dichotomy where primitive coins and contemporary tender are treated as opposites. As far as the ancient Chinese world is concerned, my opinion on this topic is relatively simple: large marketplaces were walled in, which meant that all information concerning economic transactions were concentrated, and concurrent currencies in these markets therefore seemed relatively disembedded from society. In contrast, when people used currencies outside the market, i.e., as a gift, they needed to consider the institutions and customs of the period. These phenomena influenced the synchronic relationship among currencies.

These kinds of social functions cannot be classified as monetary functions in a limited sense (i.e., a means of economic distribution). However, both social and monetary functions serve as monetary functions in a broader sense (coins, gold, and textiles) and have an impact on each other. For instance, shortage of coins due to economic causes could also have an impact on the income and expenditure of coins at family ceremonies. Conversely, no matter how severe a silk shortage is, people are obliged to bring silk to the funeral of a family friend in the Western Han period. These phenomena add up to make it very difficult to calculate the distribution channels and general distribution of currency. In other words, the movement of

various currencies is related to the economic, social, and cultural history of an area, and it is not the case that "if a person simply grasps the economic trends, they will be able to understand the flow of currency during a time period." In this way, ancient China was not very homogeneous. Hence, there were places where market principles functioned relatively well and places where they did not. There were also places where the central government system had influence and places where it did not; additionally, there were places the customs overestimated specific currencies and places where they did not. This gave rise to uneven distribution of currency across different regions. The Chinese economy since the Qin and Han empires has by no means been able to completely transform into something simple, unified, singular, rational, and economical.

According to Kakinuma (2018), the multimonetary economy, which mainly comprised coins and bolts of textiles, developed even during and after the Eastern Han period. It has a feature that the Western Han did not have. Coins and bolts of textile do not only have a common function as a means of economic liquidity but also have different functions on the basis of institutionalized and political background (Table 1). From the Eastern Han period to the Three Kingdoms period, another big change occurred – as per the increased production of textiles, the poll-tax system by coins changed to a new tax system per house by bolts of textiles. Consequently, coins lost their main position as a means of governmental settlement. Later on, Emperor Cao Rui in Wei decided to revive coins neither because coins are precious nor because it is necessary and convenient for the government but because the common people preferred coins as a means of economic liquidity more than bolts of textile and grain. As a result, the monetary economy in the Warring States, Qin, and Han periods, which comprised "coin mainly minted by the government as a means of the governmental settlement and a means of economic liquidity" and "bolts of textile mainly produced by private sector as a supplement money" changed to the monetary economy in the Wei and Jin periods, which comprised "coin mainly minted by the government as a means of economic liquidity" and "bolts of textile mainly produced by private sector as a governmental settlement."

After the Three Kingdoms period, bronze coins and bolts of textiles functioned continuingly as currencies in the Jin Dynasty. After the War of the Eight Princes, the Jin distribution, led by the central government, did not account for a large proportion, and the local administrations and local armed groups stored their currencies and resources, respectively. Even though people paid taxes, most of them did not concentrate on the central government. Under these circumstances, the central and local governments used bolts of textiles as a means of governmental payment, e.g., as a means of taxation, penalty, tributes, and officials' salaries. In contrast, bronze coins were purified as a means of economic circulation in market transactions. This is totally different from the Han period, when bronze coins mainly functioned as a means of governmental payment and a means of economic circulation in market transactions (Kakinuma 2018).

According to the previous discussion, there was not only a continuity but also a great transformation from the Eastern Han period to the Three Kingdoms period. In that case, what influence did it have on the whole society? This can be the new research task from now on. Let us pick up several tasks as below.

New Research Tasks

First, coins and bolts of textile were used for daily commercial transactions and broadly functioned as a means of economic liquidity in the Eastern Han, the Three Kingdoms, and the Jin periods. In this sense, the author calls these "money." They also have different uses depending on the time and place. These are the nonmonetary functions of coins and bolts of textile; however, these expenses and income, as a result, will affect the monetary circulation of coins and bolts of textile. Something called "money," such as coins and bolt of textile, has various usages depending on the time and place just like facial expressions, and these facial expressions should be identified and analyzed respectively. However, the face per se cannot be divided into several pieces. Here, the face means a certain object, such as coin, bolts of textile, or gold. In this sense, a traditional theory by Karl Marx and George Simmel, which states that money badly affects intimate relationships among people, should be re-examined. For Simmel and Weber, money was a key instrument in the rationalization of social life. On purely technical grounds, the possibility of money accounting was essential for the development of rational economic markets. As the most abstract and "impersonal" element that exists in human life, as Weber defined it, money became the most "perfect" means of economic calculation. According to Simmel, it transformed the world into an arithmetic problem. With money, all qualitative distinctions between objects were equally convertible into an arithmetically calculable system of numbers. Similar to this traditional theory, many criticized the situation according to which a priceless item is priced; in other words, the commoditization of the world widens. It is still fresh in our mind that Karl Polanyi (1944) recognized the situation in which labor, land, and money become commodities as the "satanic mill." Michael Sandel (2012) recently also raised an alarm against fierce commoditization. However, according to the previous discussion, money in a broad sense also functions as a bond among people. Although it does serve as a key rational tool of the economic market, it operates outside the sphere of the market and is profoundly shaped by cultural and social structural factors. Moreover, there is a plurality of different kinds of money; for example, each special-purpose money is shaped by a particular set of cultural and social factors and, thus, is qualitatively distinct. It is also possible to find such a plurality of money shaped by cultural and social structural factors even in the modern society (Zelizer 1994). Here, the whole Chinese monetary economy needs to be reconsidered from this perspective so that we can understand the relevant features of the Chinese economy and society.

Second, a money-driven society gradually appeared for the first time in history after the Eastern Han period; however, it also caused repulsion. As mentioned above, money does not always function destructively against personal intimate relationships among people; however, people in the Wei and Jin periods realized money to be just that. This is the first time in Asia to sharply increase the tone that money was "bad." A critical review named *Qian shen lun* (*A Thesis on Money as God*), written during the Wei and Jin periods, was the realization of such a tone. Such a tone emerged because a money-driven society encouraged the gap between rich and poor, and poor

people were prejudiced against rich people or even money itself. During and after the Eastern Han period, the following groups of people rose in prominence: governmental officials evaluated as "qing" (a person who gives his property as a charity, does not earn money for himself, and does not have surplus money in store); literati or landowner, called "ren" (a person who gives his property to his village as a charity); people who called themselves poor exaggerated their integrity; people who rejected to have relationships with others in order to keep a distance from the dirty world; etc. Such people had their own ideas and resisted against something. Nonetheless, the most important thing is that all sought value, which could not be measured by money. There are such people in the modern world as well and thus some might believe that money is not special for the entire human society. However, apart from whether such people really existed in ancient times, the point is that it is the first time in Chinese history that the quantity of texts concerning people in the Eastern Han period has dramatically increased. In my opinion, the money-driven society appeared in those days and, as a result, resistance against such a society came to the fore. Such a resistance is not always a self-less activity, but in reality, some of them functioned as a way in which people made his dream come true, e.g., getting a social reputation. In that case, what is the feature of such a dichotomy between a money-driven society and repulsion against it throughout Chinese history?

Third, as mentioned above, multiple kinds of money circulated during the Han, Three Kingdoms, and Jin periods. Apart from the regional differences between the light and shade of quantity of monetary circulation, coins had spread throughout the whole dynasty. However, when we read historical texts from those days, we could also find landowners who had a self-sufficient life. In that case, how can we consider the relation between them? Let us first look at the research by Nishijima (1964), Tada (1965), Yoneda (1977), and Kamiya (1993). According to these authors, landowners during those days appeared to be self-sufficient, but in reality, they were not always so. They ran a manor house in which there were many tenant-farmers and there were commodity exchanges between a lord and the tenant-farmers. After the latter half of the Western Han period, the number of famous merchants decreased but commercial activity with day laborers and tenant-farmers flourished instead of the previous style. Thus, the flourish of manors does not always correspond to the decline of the monetary economy. Additionally, an idea about the economy in northern China in the first half of the twentieth century submitted by Sukekata Kashiwa should be acknowledged (Kashiwa 1944); in other words, it is widely reported that rich landowners were important players in the monetary economy more than poor people, because landowners had more property than the poor. In reality, however, poor farmers were surprisingly connected to the monetary economy. Farmers did not have enough land for running a self-sufficient life, and thus, they could not help being wage earners for money. On the contrary, landowners did not have to get money, because they had their own property for supporting a self-sufficient life (Kashiwa 1944). This idea seems applicable to the Eastern Han, Three Kingdoms, and the Jin periods, because it reflects similar features among them: (1) big landowners; (2) a huge gap between the rich and the poor; (3) many farmers who did not have enough land to maintain a self-sufficient lifestyle; (4) the

monetary economy infiltrated into the whole society, even into the mountain regions; (5) many wage earners whose number sharply increased after the Western Han period; and (6) the official labor and military service could be substituted by paying coins after the Western Han period. Consequently, the rich were not engaged in and the poor did them instead. This hypothesis should be considered.

Fourth, a money-oriented society emerged in the Warring States period and continued even after the Western Han period; however, this does not mean the appearance of so-called "capitalism." As is widely known, Max Weber proclaimed that Chinese officials were greedy and that Asian people in history had a stronger monetary fetishism than people in England on the one hand while the Chinese society lacked the spirit of capitalism on the other hand. Apart from the lack of the spirit of capitalism, there are many discussions about why capitalism did not develop in Chinese history. This is problematic per se because the premise is that capitalism is the goal of the development of the world economic history and it is a kind of Eurocentrism. Chinese economic history might be able to choose a different path and does not have to choose capitalism. Moreover, this problem does not attempt to capture the unique features of Chinese economic history. However, even so, the question still remains: why did a monetary economy simply not develop in China? Here, we should look back at a theory regarding China's economy that was submitted in the first half of the twentieth century by Yuji Muramatsu. According to Muramatsu, the economy was mainly composed of farming and people simply produced daily farming products. Under these circumstances, most people were annoyed by high the interest rates and land rent and were engaged in domestic works. On the basis of the small government and weak guilds, people worked hard according to laissez-faire. Thus, most of people could not have enough capital for innovations and were caught by unstable stagnation (Muramatsu 1944). In fact, there was a quite similar situation in the Eastern Han, Wei, and Jin periods. This hypothesis should be considered.

Fifth, the transformation from the poll-tax system by coins in the Han period to the tax system per a family by bolts of textile in the Wei period caused an increase in the stores of bolts of textile in the national treasury. This seems to have been influential in the so-called "Silk Road" trade. According to Hisao Matsuda, trade routes from Chang'an to Tarim Basin were opened to traffic before the Han period. Certainly, they did not reach the Roman Empire, but people mainly exchanged horses for silk by these multiple routes (Matsuda 1971). Moriyasu (2007) refer to these multiple routes as the "Silk Road network." It has recently been studied that the Sogdian merchants mainly worked for these transactions before and during the Tang period. Under these circumstances, it was legally forbidden to carry bronze coins abroad at least in the Tang and Song periods (Kuwabara 1924) and the Han period (Kakinuma 2011). China proper can be separated into several economic zones in accordance with their climates and principal products, and some types of bronze coins were originally created as a means of economic transactions among these economic zones in the pre-Qin era (Kakinuma 2014). However, the shape of bronze coins was unified in the Qin period, and their usage was only permitted within the Dynasty. Subsequently, in the Han period, silk textiles became the main export good,

and they were eagerly accepted by foreigners who mainly brought horses. Matsuda (1936) coined it silk-horse trade. In this sense, it can be said that bolts of hemp textiles functioned as an inter-regional currency. Gold also had an important function as an inter-regional currency, and the Han Dynasty sometimes presented it to the northern nomads. In fact, the Sogdian merchants from the Western regions came to the fore from the third century onward, some of whom stayed in the city of Dunhuang, and their letters show how they used gold as currency on the Silk Road at the beginning of the third century (Hennin 1948; Sims-Williams 2001). In short, bronze coins were just one of many regional currencies from the perspective of the world history, and they legally circulated only within China proper. In contrast, bolts of hemp and silk textiles were inter-regional currencies throughout Asia. In this sense, the Wei period can be considered a turning point when the increase of stores of bolts of textile in the national treasury occurred that could then be used as a means of transaction by the Silk Road. Although a recent study presented a negative idea about the existence of the Silk Road transactions in private sectors before the Jin period (Hansen 2012), it is at the least true that the quantity of bolts of textile in the national treasury dramatically increased during the Wei and Jin periods. Is it really possible that the Wei dynasty just collected bolts of textiles and did not use them as a transactional tool with Central Asia? or there were Silk Road transactions in those days but historical texts as their evidence has been lost? This can be one of the future tasks.

References

Adachi K (1990) Sensei Kokka to Zaisei & Kahei. In 2012. Minshin Chūgoku no Keizai Kōzō. Kyūko Shoin, Tokyo, pp 345–373

Bloch M (1939) Èconomie-nature ou Èconomie-argent: un Peudo-dilemme. Annales d'histoire sociale 5:7–16

Braudel F (1979) Civilisation Matérielle, Èconomie et Capitalisme, XVe-XVIIIe siècle. A.Colin, Paris

Cài YZ (1996) Jiǎgǔ Jīnwén yǔ Gǔshǐ Xīntàn. Zhōngguó Shèhuì Kēxué Chūbǎnshè, Beijing

Chén L (2006) Chūnqiū Zhànguó Huòbì Dìlǐ Yánjiū. Rénmín Chūbǎnshè, Beijing

Chén L (2008) Xiānqín Huòbì dìlǐ Yánjiū. Kēxué Chūbǎnshè, Beijing

Dài Z (1981) ānyáng Yīnxū Chūtǔ Bèihuà Chūtàn. In: 2006. Dài Zhìqiǎng Qiánbìxué Wénjí. Zhōnghuá Shūjú, Beijing

Ebrey P (1986) In: Twichett D, Loewe M (eds) The economic and social history of later Han. The Cambridge history of China, vol 1. Cambridge University Press, London

Egami H (1967) Higashi Asia ni okeru Koyasugai no Ruden. In. 1999. Egami Namio bunkashi ronshū. No. 2. Yamakawa Shuppansha, Tokyo

Emura H (2000) Syunju Sengoku Shinkan Jidai Shutudo Moji Shiryo no KenKyū. Kyūko-Shoin, Tokyo

Emura H (2011) Syunju Sengoku Jidai Seido Kahei no Seisei to Tenkai. Kyūko-Shoin, Tokyo

Fāng X (2004) Zhōngguó Fēngjiàn Jìngjì Lùngǎo. Shāngwù Yìnshūguǎn, Beijing

Fù Z (1980) Zhōngguó Jìngjìshǐ Lùncóng. Joint Publishing, Hong Kong

Gemici K (2008) Karl Polanyi and the antinomies of embeddedness. Soc Econ Rev 6:5–33

Glahn V (2013) Book review Chūgoku kodai kahei keizai shi kenkyū [historical studies of the monetary economy in ancient China]. Kakinuma Yōhei The Journal of Asian Studies 72:443–445

Granovetter M (1985) Economic action and social structure: the problem of embeddedness. Am J Sociol 91(3):481–580
Guō M (1933) Bǔcí Tōngcuàn. Wénqiútáng, Tokyo
Guō X (1998) Zhōngguó Fēngjiàn Shèhuì Jīngjì Yánjiū. Shànghǎi Cáijìng Dàxué Chūbǎnshè, Shanghai
Hamada K (1912) Shina Kodai no Baika ni tsuite. In. 1989. The collection of HAMADA KOSAKU, vol 3. Dōhōsha, Kyoto
Hamaguchi F (1994) Shindai Kokyogaku no Shisoshiteki Kenkyu. Kokusyo Kankokai, Tokyo
Hansen V (2012) The silk road: a new history. Oxford University Press, Oxford
Hayek F (1976) Denationalization of money, the argument refined: an analysis of the theory and practice of concurrent currencies, 3rd edn (1980). The Institute of Economic Affairs, London
Hé Z (1948) Dōngjìn Nánchάo de Qiánbì Shǐyòng yǔ Qiánbì Wèntí. In 2006. The collection of HÉZĪQUÁN. No. 1. Chuka Shokyoku, Beijing
Hennin WB (1948) The date of the Sogdian ancient letters. Bull Sch Orient Afr Stud 12:601–615
Hildebrand B (1864) Naturalwirtschaft, Geldwirtschaft und Kreditwirtschaft, Jahrbücher für Nationalökonomie und Statistik, no. 2. F. Mauke, Jena
Huáng X (1999) Gǔwénzì yǔ Gǔhuòbì Wénjí. Wénwù Chūbǎnshè, Beijing
Huáng X (2001) Xiānqín Huòbì Tōnglùn. Zǐjīnchéng, Běijīng
Huáng J (2005) Qínhàn Shāngpǐn Jīngjì. People's Publishing House, Beijing
Iggers G (1993) Geschichtswissenschaft im 20. Jahrhundert. Vandenhoeck & Ruprecht, Göttingen
Inaba I (1984) Kandai Kazoku Keitai to Keizai Hendo. Toyoshi Kenkyu 43(1):88–117
Inaba I (1985) Sengokushin no Kazoku to Kahei Keizai. In: Hayashi M (ed) Sengoku Jidai Shutsudo Bunbutsu no Kenkyu. Kyoto Daigaku Jinbun Kagaku Kenkyujo, Kyoto
Jiǎ G (1976) Shāngpǐn Huòbì yǔ Yīnshāng Núlìzhì. Kǎogǔ. no. 1
Jìng J (1987) Shìlùn Dìzhǔzhì Jīngjì yǔ Shāngpǐn Jīngjì de Běnzhì Lián xì. Zhōngguó Jīngjìshǐ Yánjiū. no. 2
Kageyama T (1984) Chūgoku Kodai no Shōkōgyō to Senbaisei. Tokyo Daigaku Shuppankai, Tokyo
Kakinuma Y (2011) Chūgoku Kodai Kahei Keizaishi Kenkyū. Kyūko Shoin, Tokyo
Kakinuma Y (2014) The Emergence and Spread of Coins in China from the Spring and Autumn Period to the Warring States Period. In: Bernholz P, Vaubel R (eds) Explaining Monetary and Financial Innovation. Springer, Cham
Kakinuma Y (2018) Continuity and Transformation in the Monetary Economy of Ancient China. Kyūko Shoin, Tokyo
Kamiya M (1993) Zenkan Kohanki Iko no Kahei Keizai ni tsuite. Higashi Asia ni okeru Seisan to Ryutu no Rekishi Syakaigakuteki Kenkyu. Chugoku Shoten, Fukuoka, pp 72–96
Kashiwa S (1944) Hokushi no Noson Keizai Shakai. Kobundo Shobo, Tokyo
Kato S (1991) Chūgoku Kaheishi Kenkyū. Tōyō Bunko, Tokyo
Kawakatsu Y (1982) Kahei Keizai no Shinten to Kōkei no ran [original version was printed in 1962]. In: Rikuchō Kizokusei Shakai no Kenkyū. Iwanami Shoten, Tokyo
Kimura M (1955) Chūgoku Kodai Kaheisei. In: Tōyō Shigaku KenKyū, vol 4. San'ichi Shobō, Tokyo
Kimura M (1960) Chugoku Kodai Teikoku no Keisei. Hikaku Bunka Kenkyujo, Tokyo
Kondo K (1998) Seishu Jidai Takaragai no Kenkyu. Asia no Rekishi to Bunka. no. 3
Kuwabara J (1924) Toso Jidai no Dosen. In 1968. Collections of KUWABARA JITSUZO. Iwanami Shoten, Tokyo, pp 270–360
Kuwabara J (1925) Rekishijo yori mitaru Nanboku Shina. In 1968. Collections of KUWABARA JITSUZO. Iwanami Shoten, Tokyo, pp 270–360
Lao G (1976) Hàndài Huángjīn jí Tóngqián de Shǐyòng Wèntí. Láogàn Xuéshù Lùnwénjí. Jiǎbiānxià. Yìwényìn Shūguǎn, Taipei
Lín G (1997) Qínhàn de Zìrán Jingjì hé Shāngpǐn Jìng jì. In. 2005. Lín, Gānquán Wénjí. Shànghǎi Císhū Chūbǎnshè, Shanghai
Liú J, Lǐ Z (1995) Zhōngguó Huòbìshǐ. Wénjīn Chūbǎnshè, Taipei
Mǎ F (ed) (1988) Zhōngguó Lìdài Huòbì Dàxì: Xiānqín Huòbì. Shànghǎi Rénmín Chūbǎn, Shànghǎi

Mǎ F (ed) (2002) Zhōngguó Lìdài Huòbì Dàxì 2: Qínhàn Sānguó Liǎngjìn Nánběicháo
Makino T (1950) Chūgoku Kodai no Kazoku ha Keizaiteki Jikyu Jisokutai ni arazu: Chūgoku Kodai Kahei Keizai no Hatten. In 1985. The Collection of MAKINO, TATSUMI. No 6.: Chūgoku Shakaishi no Shomondai. Ochanomizu-Shobō, Tokyo
Makino T (1953) Chugoku Kodai Kahei Keizai no Suitai Katei. In: Uehara S (ed) Shakai to Bunka no Shoso. Josui Shobo, Tokyo
Matsuda H (1936) Kenba Koeki Oboegaki. In 1986. The collection of MATSUDA HISAO. No. 2. Rokko Shuppan, Tokyo, pp 140–153
Matsuda H (1971) Asia no Rekishi. Tokyo. In 1987. The collection of MATSUDA HISAO. No. 5. Rokko Shuppan, Tokyo, pp 1–178
Matsumaru M (1992) Seishū Jidai no Jūryō Tan'i. Asia no Bunka to Shakai. no. 2. Kyūko Shoin, Tokyo
Miyazaki I (1964) Rikuchō Zuitō no Shakai. In. 1992. The collection of MIYAZAKI, ICHISADA. No. 7. Iwanami Shoten, Tokyo, pp 119–132
Moriyasu T (2007) Silk Road to To Teikoku. Kodansha, Tokyo
Muramatsu Y (1944) Chugoku Keizai no Shakai Taisei. Toyo Keizai Shinposha, Tokyo
Nishijima S (1964) Shinkan no Nogaku. In 1966. Chugoku Keizaishi Kenkyu. Tokyo Daigaku Shuppansya, Tokyo
Odake F (1935) Chusei Shina Shakai Keizai Hattatsushi Gairon. Shina Kenkyu. no. 36
Okazaki F (1932) Gishin Nanbokucho Tsushi. Kobundo Shobo, Tokyo
Okazaki F (1935) Nanbokucho ni okeru Shakai Keizai Seido. Kobundo Shobo, Tokyo
Péng X (1958) Zhōngguó Huòbìshǐ. Shànghǎi Rénmín Chūbǎnshè, Shànghǎi
Peng K (2000) Coinage and commercial development in eastern Zhou China. Ph.D. Dissertation, University of Chicago, Illinois
Polanyi K (1944) The great transformation: the political and economic origins of our time. Beacon Press, Boston
Polanyi K (1968 [1957]) The semantics of money-uses. In: Dalton, G. (ed) Primitive, archaic and modern economies: essays of Karl Polanyi. Beacon Press, Boston
Quán H (1941) Zhōngguó Zìrán Jīngjì. In QUÁN, HÀNSHĒNG. 1980. Zhōngguó Jīngjìshǐ Yánjiū. no. 1. Dàoxiāng, Taipei
Sandel M (2012) What money Can't buy: the moral limits of markets. Penguin, London
Sekino T (1956) Chugoku Koukogaku Kenkyu. Toyo Bunka Kenkyujo, Tokyo
Sekino T (2005) Chūgoku Kokogaku Ronko. Doseisha, Tokyo
Sheidel W (2009) The monetary Systems of the Han and Roman Empires. In: Sheidel W (ed) Rome and China: comparative perspectives on ancient world empires. Oxford University Press, Oxford
Sims-Williams N (2001) The Sogdian ancient letter II. In: Schmidt MG, Bisang W (eds) Philologica et Linguistica. Historia, Pluralitas, Universitas. Festschrift für Helmut Humbach zum 80. Geburtstag am 4. Dezember 2001. Trier, Wissenschaftlicher Verlag, pp 267–280
Tada K (1965) Kandai no Chiho Shogyo ni tsuite. Kangishin shi no Kenkyu. Kyuko Shoin, Tokyo, pp 25–48
Táo X (1934) Biānjí de Huà. Shíhuò Bànyuèkān no 1-1
Tierry F (1997) Monnaies Chinoises, I: L'antiquite preimperiale. Bibliothèque Nationale de France, Paris
Tierry F (2017) Les Monnaies de la Chine Ancienne. Les Belles Lettres, Paris
Wáng G (1921) Shuō Juépéng. In 1959. Guāntáng Jílín. Zhōnghuá Shūjú, Běijīng
Wáng M (1947) Xiānqín Huòbìshǐ. Guólì Zhōngshān Dàxué Chūbǎnzǔ Chūbǎn, Taipei
Wáng Y (1957) Zhōngguó Gǔdài huòbì de Qǐyuán hé Fāzhǎn. In. 2005. Wáng Yùquán Shǐlùnjí. Zhōnghuá Shūjú, Beijing
Wáng Y (2007) Wèijìn Nánběicháo Huòbì Jiāoyì he Fāxíng. Wénjīn Chūbǎn, Taipei
Weber M (1924) Wirtschaftsgeschichte: Abriss der universalen Sozial- und Wirtschafts-Geschichte, 2nd edn. Duncker & Humblot, München und Leipzig
Wǔ X (1937) Wèijìn Nánběicháo Jìngjìshǐ. Shāngwù Yìnshūguǎn, Shanghai

Xiao Q (1984) Zhōngguó Gǔdài Huòbìshǐ. Shànghǎi rénmín chūbǎnshè, Shànghǎi
Yamada K (2000) Kahei no Chūgoku Kodaishi. Asahi Shinbunsha, Tokyo
Yáng S (2003) Bèi shì Shāngdài de Huòbì. Zhōngguóshǐ Yánjiū 1:29–40
Ye M (1994) Chuántǒng Shìchǎng yǔ Shìchǎng Jīngjì Yánjiū Shùpíng Fēngjiàn Dìzhǔzhì Qiánqī. Zhōngguó Jīngjìshǐ Yánjiū. no. 4
Yoneda K (1977) Chugoku Kodai Nomin Gijutsushi Kenkyu. Dohosha, Kyoto
Yoshida T (1933) Shina Kahei Kenkyu. Toya Keizai Kenkyu Kai, Tokyo
Zelizer V (1994) The social meaning of money. Basic Books, New York
Zhū H (1984) Gǔqián Xīntàn. Qílǔ Shūshè, Jǐnán

The Monetary System of China Under the Qing Dynasty

22

Niv Horesh

Contents

Abstract

This chapter delineates the main trends that characterized Qing monetary history. The early Qing (1644–1722) was generally a period when military expenditure relative to fiscal revenue was high and coin casting sites in bad repair. The mid-Qing (1723–1842) was, on the other hand, a period of great territorial expansion, strong central government, and vigorous coin output, mainly under the Qianlong Emperor (r. 1735–1796). The late Qing (1842–1911) was, in turn, a period in which China came under intense foreign pressure and was rattled by the catastrophic Taiping rebellion. It was also a period in which China began suffering from an acute outflow of silver due to an adverse balance of payments. These circumstances eventually led the Qing government to reintroduce paper money. And while Japanese silver and copper imports had helped replenish the Chinese monetary system in the early-Qing period, late-Qing China saw a decline in domestic mining output as well as a drop in Japanese metal imports. These

N. Horesh (✉)
Durham University, Durham, UK

Western Sydney University, Sydney, NSW, Australia
e-mail: niv.horesh@westernsydney.edu.au; niv.horesh@gmail.com

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_54

shortfalls were tempered in the mid-Qing era by the influx of Latin American silver, which lasted until the 1800s, as well as by the development of copper mines in Yunnan.

Keywords
Copper · Silver · Bimetallism · Debasement · Paper money

Introduction: Sources, Periodization, Themes

The most fundamental term with which to describe the monetary system during the Qing dynasty (1644–1911), as opposed to some other periods, would be bimetallism. That is to say those two types of monetary metals circulated concurrently in the economy of China proper. The first one being copper in the form of cast coinage (*zhiqian*), with minimal inscription in Chinese and Manchu, and a square hole in the middle. Copper coinage broadly served the peasant economy. The other type consisted of silver bars (*yinding*) broadly serving the government and mercantile community.

Customarily, batches of 1000 Chinese copper coins would be threaded together through the round hole in the middle for ease of transfer. Such "strings of cash" were known as *diao*, each being conceptually on par with one tael or *liang* of silver (approximately 37.5 g). Higher denomination money in the form of silver principally meant bars resembling traditional Chinese shoes – known to Westerners as *sycee*. These shoe-shaped bars weighed about 50 tael, but there could be considerable variations of ingot weight and fineness even within the same province. Such variations gave rise to assaying establishments, mostly of private-order nature, which guaranteed the quality of sycee used in market transaction (Wagel 1915; Kann 1927; Cribb 1992; Rang 1988). Tax receipts in loose silver ingots would often incur "meltage fee" (*huohao* or *haoxian*) whereby local officials would demand a surcharge from taxpayers for the silver lost during the process of assaying and recasting the ingots into standard bars.

The term *sycee* was an English derivative of the Cantonese pronunciation for "pure silk." In South China, a good-quality silver bar was thought to possess a shiny veneer reminiscent of silk. From mid-seventeenth century, foreign-minted silver dollar coins (*yinyuan*, meaning round silver) otherwise constituted a lighter and therefore more convenient medium of exchange. The latter were virtually all of Latin American provenance until the Qing started minting silver dollars of its own in the late nineteenth century. Tax was remitted to imperial coffers in the form of silver ingots, while official and military wages were often paid out in copper coinage. Foreign silver dollars, however, became an increasingly common means of payment by 1700, particularly in coastal areas (Man-houng 2006, p. 43; Shi 2016). In addition to copper coins, the Qing authorities produced distinct silver coins and relatively small amounts of gold coins exclusively for use in Xinjiang and Tibet (*Zhongguo lidai huobishi* 1999; Boulnois 1983; Mu Yuan 1994).

One copper "cash" coin would conceptually weigh approximately one tenth of a silver tael, thereby implying a 1:100 exchange rate between silver and copper. However, the actual exchange rate was affected by coin weight fluctuation over time of between 0.1 and 0.14 of one tael, as well as variations in alloy composition (i.e., zinc and tin infusion). In addition, that bimetallic exchange rate fluctuated according to the availability of either metal at any point in time and respective demand. Namely, if the exchange rate between silver tael and copper cash was 1:1000 in principle, then shortages of either metal or coin supply shifts could bring about fluctuations of anywhere between 500 and 1500 coins per one silver tael (Shi 2016, xxiii).

Within the cash economy, a distinction was made between "smaller" coins (*xiaoqian*) and "bigger" ones (*daqian*) whereby "big" would usually denote standard-weight or full-bodied coins and "small" would denote debased or forged coins. One thousand standard coins would count as 1000 legal tender units (*wen*), thus qualifying as one string of cash. On the other hand, *xiaoqian* would circulate below their face value, namely, 1000 thereof would fall short of one standard string. However, it is important to note that in other settings, *daqian* could mean quite the opposite. Historically, one inflationary means for Chinese rulers was to attempt disbursing bulky coins with a face value of anywhere between 5 and 1000 *wen* but weighing much less that what their face value might convey (see, e.g., the "Xianfeng" era section below).

There is a large body of excellent scholarly literature focusing specifically on the Qing monetary system within the larger sweep of East Asian monetary history. The most comprehensive and influential studies arguably include Morse (1906), Kann (1927) and King (1965) in English, Yang Duanliu (1962) and Chen Feng (2008) in Chinese, and Ichiko Shōzō (2004) in Japanese. The monetary system is also covered in general surveys of the Qing economy – most notably Wang Yeh-chien (1981) and Quan Hansheng (1987). Basic Qing monetary, financial, and fiscal thought is covered, in turn, in studies of Chinese premodern economic thought – most notably Ye Shichang (2003) and Zhang Jiaxiang (2001) in Chinese.

Vital, as well, to gaining a better understanding of the system are comparative studies that underscore the Qing era within the larger sweep of Chinese and world monetary progression, including the various iterations of Peng Xinwei's pioneering studies (1958–2007, 1994) in Chinese and in English translation (1994); Akinobu Kuroda (2003) and Kishimoto Mio (1997) in Japanese (2003); and Elvin (1973), Davies (2002), Horesh (2014), and Von Glahn (2016) in English. Clearly, the Qing monetary system diverged from Western forms of bimetallism before the British adoption of the gold standard in 1821. Western bimetallism comprised silver alongside higher-value gold, both in the form of minted coinage with elaborate inscription, as opposed to China's traditionally cast "cash," which was easier to forge. Western bimetallism also typically entailed a fixed exchange rate between gold and silver which was determined by the government, whereas the Chinese silver-copper exchange rate was largely determined by the market.

This divergent Chinese system did not usually appear discordant to Western observers until well into the nineteenth century. From then on, however, the peculiarities of the Qing monetary system would be increasingly criticized as eccentric and anachronistic (Horesh 2014; Burger 2008). That is not to say that gold had never served as important medium of exchange in China proper. To the contrary, as Kakinuma's chapter in this Handbook clearly suggests, gold had in fact been a key component of the Chinese monetary system in antiquity. However, by the mid-Qing era, decorative tastes and monetary patterns alike made the gold to silver exchange rate (roughly 1:10) lower than in India and Europe (1:12–15), where gold *was* coined. In other words, silver was comparatively dearer in China, and consequently much imported silver was exchanged there in return for gold, which was later on sold in India and Europe (Giráldez 2015, 174–176).

In fact, prior to the Qing coming to power, iron base coinage, bolts of fabric, and paper money had also been important components of the Chinese monetary system at various times. However, Qing emperors preferred to stick with the bimetallic monetary principles of the late-Ming dynasty (1368–1644). Unlike the Ming emperors, however, the Qing emperors were ethnic Manchus who were overseeing the Chinese majority. To consolidate power, the early-Qing emperors had to defeat or placate many ethnic-Chinese (i.e., Han) Ming loyalists. They therefore felt they had to demonstrate their adherence to Chinese norms of governance as practiced in the late Ming. As such, conservatism in the monetary realm implied rigid continuation of the bimetallic standard.

On the other hand, the *early* Ming emperors inherited from the Song and the Mongol (Yuan) dynasties a predilection for paper money issuance. That China was the birthplace of paper money long before Europe had been attested to by Marco Polo and other contemporaries (Vogel 2012). By the 1500s, however, government paper money had fallen from grace because it became associated with inflation. As will be explained below, when consolidating power, the first Qing emperor briefly resorted to convertible paper money issuance, but phased those notes out as soon as the fighting subsided.

Qing economic conservatism endured in the face of the great monetary upheaval sweeping other parts of the world in the early modern period. It ought to be understood in the context of China's political economy where ideals revolved around low taxation and the geostrategic orientation was toward defense of the northern steppe lands whence the Manchus themselves had hailed. In other words, the Qing political economy was less geared toward maritime expansion of the early-modern mercantilist streak (Goldstone 1991).

Although there had been relatively little systemic change in monetary thought over the course of Qing rule, it is useful to divide it into three periods. Generally speaking, the central authorities' ability to regulate coin output across the length and breadth of this vast empire was seen primarily as means of stabilizing grain prices, i.e., a marker of upright governance in Confucian thought. Thus, ample coin output reflected the strength of the dynasty in the eyes of Han officials and commoners alike. The early Qing (1644–1722) was generally a period when residual armed resistance to the Manchus among the Han populace was still rife. Military expenses

were therefore high, and foundries (i.e., "mints" where coins were *cast*) were in bad repair. The mid-Qing (1723–1842) was, on the other hand, a period of great territorial expansion, strong central government, and vigorous coin output, mainly under the Qianlong Emperor (r. 1735–1796).

The late Qing (1842–1911) was, in turn, a period in which China came under intense Western and Japanese pressure, beginning with the first Opium War (1839–1842). It was also a period in which China began suffering from an acute outflow of silver due to an adverse balance of payments. While Japanese silver and copper imports had helped replenish the Chinese monetary system in the early-Qing period, late-Qing China saw a decline in domestic mining output as well as a drop in Japanese metal imports. These factors were tempered in the mid-Qing era by the influx of Latin American silver, which lasted until the 1800s. To be sure, the Portuguese had been trading in silver coins and in South China as early as the sixteenth century. But Latin American silver played a more crucial role in the mid- and late-Qing economy. That role is explored in depth in studies by Flynn and Giráldez (2002), Lin Man-houng (2006), and Irigoin (2009).

In the early and mid-Qing period, trade with the West had been largely confined to the city of Guangzhou (Canton) and was broadly in China's favor. Namely, Chinese silk, tea, and porcelain were highly sought after in Europe, but only few Western commodities proved attractive to the Chinese. For that reason, Latin American-sourced silver was the only means of plugging the trade gap. Millions of Spanish colonial silver coins ("pieces-of-eight") and, later on, Mexican Republican dollars found their way east annually. Even though silver dollars from the New World became popularly entrenched in the monetary system to the exclusion of most other foreign coinage, the Qing authorities had little clue before the twentieth century as to the condition of coin production in Latin America (Von Glahn 2007).

Apart from the expansion of maritime trade after 1842, Chinese tea had traditionally also arrived in Europe by way of Russia, as did porcelain, rhubarb, nankeens, and silk. In fact, the Romanov Empire was one of the first European polities to come into direct contact with the Qing authorities in Beijing as a result of its expansion into Siberia in the seventeenth century. The two empires agreed to form a trading hub in Outer Mongolia. However, that trade was largely in the form of barter – the Russians usually paying for Chinese commodities in furs and cattle (Foust 1969). Thus, although Russia became an important global supplier of gold in its own right by the 1830s, its monetary impact on the Chinese system remained peripheral.

Eventually, over the course of the eighteenth century, the secrets of porcelain and silk processing found their way to the West. Then, in the late nineteenth century, tea seeds were transplanted by the British in India. That import substitution, coupled with the British ability to popularize the consumption of Indian-grown opium in China after 1842, turned around the balance of payments. Much of the late-Qing period was thus characterized by a flight of silver from China. As a result, the local exchange rate between silver and copper was shooting up. Concomitantly, the land tax liability of the peasantry, which was calculated in silver tael but paid in copper cash, became more onerous (Perkins 2017).

Following their defeat in the Opium Wars, the Qing authorities were compelled to open up more cities to foreign trade. At the same time, more and more of China's customs revenue was surrendered to the foreign-run Imperial Maritime Customs in order for the faltering Qing authorities to raise loans in London. These loans were necessary for the payment of war indemnities to the foreign powers that defeated China and for quelling the rebellions that had broken out against Manchu rule in the 1850s. That these loans were denominated in gold, while the world price of silver was falling, made it all the more difficult for the Qing authorities to repay their debts (Van de Ven 2014). Consequently, the foreign powers applied pressure on the Qing authorities to streamline their finances and taxation structure (Vissering 1914). But China remained loath to joining the gold standard. And as it turned out, it never did even after the downfall of the Qing in 1911.

On the upside, the cities opened for trade after 1842 ("treaty ports") lured in Western financial institutions, which Chinese officialdom and private entrepreneurs began emulating. The treaty ports would therefore become the cradle of modern Chinese finance in the latter half of the nineteenth century (McElderry 1976; Ji Zhaojin 2002; Horesh 2009, 2015; Cheng Linsun 2003).

Thematically, Qing monetary history feeds into two of the most consequential debates in Chinese modern historiography. First, following Pomeranz's (2001) pathbreaking study, there has emerged a revisionist school of thought suggesting that as late as 1850 Chinese per capita income was on par with Northwestern European norms. Quite naturally, in this complex debate straddling such vastly different polities on either extreme of Eurasia, whatever diffuse data on salaries are available, they can and should [data] be augmented by other indicators of wealth. Reliable money supply estimates could be crucial here, as are international currency exchange rates. More recent work by Allen et al. (2011) and Broadberry and Gupta (2006) has compensated for these data shortfalls by incorporating price and nutrition data with results that have tended to reestablish conventional wisdom. Horesh (2014) has argued, in turn, that the level of monetization in late-Qing China was far lower that the norm across much of Eurasia, so high income is counterintuitive.

An addendum to this debate is the conventional wisdom postulating that interest rates in Qing China were much higher than in the West. This factor, it is often argued, held back the industrialization of China right until the twentieth century. The variance here is attributed to the unruly nature of Chinese money markets before the nineteenth century. In his earlier work, Kenneth Pomeranz has suggested, for example, that speculators often transported copper coins from one province to another in order to profit from exchange rate differentials (Pomeranz 1993, pp. 43–51). However, Rosenthal and Wong (2011, pp. 134–140) have challenged this perception, arguing that Chinese interest rate data in previous studies were erroneously inferred from pawnshop data, while European interest rates were derived from annuities. In any event, the debate about historical Chinese interest rate and its correlation with agricultural productivity or living standards is still underway, calling for more research (see, e.g., Liu Qiugen 2000).

The third thematic thread linking monetary history with broader debates in Chinese historiography has to do with the so-called New Qing History (NQH).

First emerging in the 1980s, it is a revisionist school of thought, portraying the Qing era as one of eclectic policy dynamism rather than conservatism. The historiographic debate here has so far largely revolved around the degree to which Manchu emperors honed their steppe acumen to form a multiethnic empire stretching far beyond China proper (e.g., Crossley 2002; Elliott 2002). The upshot was that, in order to hang on to power in the steppe frontier, Qing administrators creatively used very different tactics than those applied to China proper. Conventionally, the Qing emperors are seen as having progressively become sinicized, but NQH historians show the enduring significance of Manchu language skills retention at court. And while they accept that Qing emperors were much less geared toward maritime trade as compared with Western rulers, they do not draw parallels between the Qing patterns of governance and those of the Ming but with those of the Romanov, Ottoman, and Mughal empires (Waley-Cohen 2000).

The NQH revision of conventional wisdom has a distinct monetary dimension, calling for more research in the future. At heart here is the question of whether the funding of mid-Qing territorial expansion westward resembled the Ming, Ottoman, Russian, or Western military outlay pattern. Clearly, the mid-Qing polity did not have at its disposal the kind of national debt instruments pioneered in Sweden, the Netherlands, England, and France at the time. In much of the premodern world, coin debasement had instead constituted an important source of sovereign funding along with land tax receipts. Recoinage, on the other hand, could be a form of monetary stabilizing. In early modern Europe, debasement was then eclipsed by the issuance of convertible paper money and government bonds, quite often in order to raise funds with which to fight wars.

Qing China was no exception as territorial expansion under the Qianlong Emperor required new sources of funding. Yet it appears that, ultimately, most of the additional revenue derived from the taxation of commerce rather than from debasement. Because of the nature of the Qing base metal coinage and the expenses associated with copper casting, it appears debasement funded only a small part of the Qing military campaign in Xinjiang (see, e.g., Millward 1998, pp. 64–75; Perdue 2009, pp. 379–393). Arguably, since Western precious metal coinage was minted, it was at once harder to forge and easier to extract seignorage from (Horesh 2014).

Qing memorials are generally replete with debasement measures, but there are also notable exceptions. Copper cash could be called in to government foundries and recast (*gaizhu*) in inflationary or deflationary manner. As already indicated, inflation could, for example, entail the production of heavier coins (*daqian*) bearing face value above their intrinsic copper weight; producing standard-size, finer copper coins could mean replenishing markets with currency bearing a face value below their worth by tale. Depending on enforcement, in the long run, better quality coinage could either be hoarded or become trusted to the extent it would eventually be traded at a premium.

Given its base qualities, there were periods and localities where copper cash output could drop precipitously because it proved a net drain on government resources. In the European historical setting, such episodes are known as “the big problem of small cash” (Sargent and Velde 2014). At other times, officials prescribed

a boost to casting (*guzhu*) because the dearth of cash was unsettling the peasantry and considered a potential cause for rebellion.

It would appear that the Qing polity – physiocratic as it was – ordinarily prized the ready availability of cash over quick seignorage gains, at least in its grain surplus heartland (cf. Dunstan 2006). So much so that the authorities at times forbade the casting of copper utensils in order to free up raw material for coinage (*tongjin*). Such measure could be implemented because virtually all mining output was subject to an imperial monopsony of sorts. However, far from seizing up new seignorage avenues, the authorities remained relatively tolerant of fiduciary private-order means of payment, including copper-denominated notes, until the 1850s (He 2013, Chap 5). There were even times when the Qing authorities acquiesced in the coinage of previous dynasties (*guqian*) continuing to circulate as legal tender (Horesh 2014).

Copper Cash and Silver Sycee

As noted earlier, the only money which circulated universally throughout the length and breadth of the Qing empire was copper "cash." The name derived from the Portuguese word *caixa*, but the term used in Chinese was *zhiqian* or *tongqian*. During the late-Ming era, the Portuguese were of course the first Europeans to have come to China by sea and were thus the first to observe the intricacies of the local monetary system.

The origin of the English term *cash* as distinctly denoting Chinese-style copper coinage is the Sanskrit silver and gold weight unit "karsa." The latter evolved into the Tamil word *kaasu*, denoting low-value coinage. It then entered the English language via the Portuguese variants for *kaasu*, *caixa*. This is not to be confused with the primary (and older) instruction of *cash* in the English language, which derives from the medieval Italian *cassa.* The tael unit, on the other hand, originated in the Malay word for weight, which similarly entered English usage by way of Portuguese.

The influx of millions of silver coins minted in present-day Bolivia and Mexico gradually changed the Chinese monetary landscape. Kishimoto (1997, pp. 359–363) argues that by the mid-eighteenth century most property contracts had been denominated in smaller silver dollars units rather than in tael units (a silver dollar usually contained around 0.7 the amount of silver in one tael). On the mainland, however, the silver dollar would become a more popular unit of account than the tael or copper coins only around the mid-nineteenth century.

Across East Asia, one could detect difficulties in sourcing sufficient metal to meet the popular demand for coinage during the early eighteenth century. In other settings, this might have catalyzed the emergence of fiduciary currency. Yet, by then, government-issued banknotes had fallen out of grace as a credible means of payment in both Korea and China.

In the early-Qing era, the exchange rate between copper coin and silver had been very volatile. Shortage of copper had in fact forced the closure of provincial foundries during the Kangxi reign (1661–1722). Nationwide casting of coin would

only resume in the late 1740s. By contrast, throughout the Qianlong era, copper coin prices were relatively high, namely, above 1000 coins per 1 silver tael (Vogel 1987; Dunstan 1996, p. 70).

As discussed in great detail by Vogel and Theisen-Vogel (1989) and Golas (1999, pp. 389–392), much of the copper needed to produce Chinese "cash" had been imported from Japan between 1685 and 1715 (Shimada 2006). Nevertheless, in 1726 the *bakufu* authorities prohibited further exports to China, in part, so as to increase its own domestic coin output. In effect, Japanese copper continued to trickle into China thereafter, but in much smaller quantities. By the end of the eighteenth century, it accounted for only one tenth of Qing coin raw material. Moreover, between 1795 and 1851, it dropped by as much as 50% (Lin Man-houng 2015, p. 182).

Faced with the main source of copper supply drying up, the Qianlong Emperor was forced to develop alternative copper sources in remote Yunnan early in his reign. Since Yunnan was further away from the capital foundries than Japan, transport costs made coin production less profitable. To make matters worse, Yunnan was capital scarce. So in order to facilitate private mining ventures there and elsewhere in China, the Qianlong Emperor reduced output taxes to 10 – 20% and at times funded mining ventures directly (Dunstan 1996; Rowe 2005; He 2013, p. 32).

Despite these efforts to offset the loss of Japanese copper supply, a spike in the relative price of "cash" and subsequent coin dearth persisted until the late eighteenth century, often leading to lighter coin forgeries (*sizhu*). At the same time, China attracted much Latin American silver from foreign traders, thus enhancing the relative dearness of copper (Rowe 2009, pp. 156–158).

The Qianlong Emperor recognized that, in the short run, heavy-handed measures like impounding lighter coins or banning private copper casting would disrupt rural markets and push the price of copper even higher. The policy adopted instead was long-tern in nature: it was decided to recast better-quality coins from Yunnan copper with a view toward both seignorage and market stability. On the other hand, there is little evidence to suggest progress in the technology used for copper mining during that era; Yunnan copper mining does not seem to have incorporated water power devices, horsepower, or mechanized intervention (Golas 1999; He 2013, pp. 134–136).

In sum, under the young Qianlong Emperor, the post-Song secular decline in Chinese per capita coin output levels was temporarily halted by virtue of the newly sourced Yunnan copper and a vigorous expansionist policy. Chinese coin output level started dropping again after the 1770s, and throughout the late-Imperial era, coins genuinely or falsely pertaining to previous dynasties (*guqian*) widely circulated in the Chinese marketplace (see Zelin 1984, pp. 264–301; Lin Man-houng 2006, pp. 29–30).

The price of copper coin relative to silver dropped through much the late-Qing era. Conventional wisdom suggests the two key reasons for that were an adverse balance of payment leading to silver depletion and coin debasement. However, Lin Man-houng (2006) has argued that world silver supply shocks resulting from Latin American emancipation or the Napoleonic wars were also at play. Rather than making "cash" more precious, the drop in coin output and quality after the Qianlong

heyday appears on balance to have dampened popular demand. If 1.6 billion *wen* had been cast in Beijing alone in 1800, then the output annual figure for 1841–1850 was just 1.2 billion on average (Lin Man-houng 2015).

Crucially, Yunnan copper veins had been depleted by the beginning of the nineteenth century. Nevertheless, it was only in 1867 that officials started memorializing the throne about the need to employ modern (i.e., steam-powered at the time) mints like the one then operating in British-ruled Hong Kong so as to economize on raw material, improve coin consistency, and discourage forgeries. The intricacies and persistence of Qing-era coin forgery prior to the adoption of steam-powered mints are discussed by Cao and Vogel (2015) and Greatrex (2008). For those interested in the topic, there is available in addition a historical overview of coin forgeries in China (Zheng Jin 2007).

In the event, steam-powered minting machinery was purchased in Birmingham, England, in 1887, and used only in Guangdong at first. The Qing government subsequently minted its first silver coin for usage all over China ("the dragon dollar") in 1889 with a view toward chipping into the premium that Mexican dollars could fetch.

In addition, by the 1900s modern copper coin minting under imperial and provincial auspices was so extensive and debased that the value of these so-called *tongyuan* coins dropped precipitously. Although notional bimetallism survived in the Chinese hinterland until 1935, "cash" production and the customary stringing of "cash" had in effect died out in the early twentieth century. This was because *tongyuan* coins were unholed and often denominated in multiple units well above their actual metallic value.

Being full-bodied, traditional holed "cash" was gradually driven out from circulation in favor of ever-lighter types of modern-minted fractional coinage of the *tongyuan* variety. Traditional coins were otherwise smelted or exported to Japan, where the demand for raw copper in the ammunition industry was on the increase (Yang Duanliu 1962, pp. 22–23, 44–45; Ho Hon-wai 1993; Zheng Jin 2007, pp. 218–220). Ironically, some three decades earlier, it was actually traditional Japanese copper coins (*kan'ei*) that had started being shipped to coastal China, following the establishment of a modern mint in Osaka. Burger (2015, p. 221) estimates, for example, that around 15% of all coins circulating in Tianjin around 1900 were Japanese.

Qing Ambivalence to Paper Money

The nineteenth-century Western sources did not by and large provide an incisive answer as to the question of precisely why government and private-issued paper money went into decline in China in previous eras, how and precisely when privately issued notes reemerged in the Qing era, and how important a component they made for within the late-Qing monetary system. Neither did they explain in great detail whether or not the gradual reemergence of privately issued banknotes in China –

which could be traced back to the latter part of the eighteenth century at the earliest – had anything to do with global financial stimuli.

Lin Man-houng (2006, pp. 36–37) suggested, for example, that premodern Chinese money shops did not start issuing notes until very late in the Qianlong reign; these shops were probably more developed in the north of China, where they mostly issued copper coin-denominated notes. Notes issued by southern money shops were denominated in silver *taels*. Hou and Wu (1982, vol. 1, p. 11), on the other hand, suggest private-order notes might have emerged earlier in the eighteenth century and that they were very common by the 1820s. Horesh (2014) and Cheng Linsun (2003) have more recently argued that foreign banks operating in China as of the late 1840s greatly bolstered the reformist monetary discourse there, both within and outside the imperial bureaucracy. These foreign banks were later subject to emulation by China's homegrown modern banks.

Between 1848 and 1945, European, American, and Japanese colonial banks all issued locally denominated notes on Chinese soil, relying on metallic reserve policies and regulatory prescripts derived from the Western and Japanese monetary experience (Xian Ke 1958; Rawski 1989). As noted, these foreign financial institutions both challenged and set an example for homegrown financial institutions, laying the foundation for the rise of modern Chinese banks and the popularization of banknotes in China in the early twentieth century.

British banknotes in China were by far the most common and esteemed of all foreign banknotes. How early British banknote issuance on Chinese soil started informing Chinese monetary thinking later in the nineteenth century and early twentieth century is explained in detail by Feuerwerker (1958) and Horesh (2009). Suffice it to note here that from the 1890s onward the invocation of Western and Japanese monetary history would all but override, in argumentative terms, Chinse historic precedents in court debates on the viability of paper money. Prominent monarchist, reformer-turned-dissident, and intellectual luminary, Liang Qichao (1873–1929) is perhaps the best example: he tirelessly campaigned for China to emulate the West by embracing the gold standard, unify refractory currencies, and issue state-backed banknotes with a 1/3 metallic reserve (Hou and Wu 1982, vol. 3, pp. 322–339).

Other scholars, however, argue that *privately* issued banknotes had become a pillar of the late-Imperial Chinese monetary system well before the entry of Western banking in the late 1840s. In the alternative narrative, these scholars seem to offer what the Chinese at the time did not trust were only *government*-issued notes. Notes issued by established private financiers were well accepted, it is argued (Wang Ye-chien 1981; Hao Yen-p'ing 1986, pp. 47–50).

This argument underscores what may look at first like academic nit-picking. Yet beyond the small detail lies an important scholarly debate about the nature of the foreign contribution to China's economic modernization and the circumstances in which *private* banknote issuance spread as of the mid-Qing era. The mainsprings of private banknote issuance entail, in turn, at least five foci of discussion: why did the Qing court discontinue its first issue not long after it had securely asserted its rule over China (1650s)? Why did the Qing court refrained from engaging in further

issues right up until it was hit by financial and political instability in the 1850s? Why was that 1850s issue also discontinued? Finally, and most importantly, were Chinese privately issued notes considered reliable, and did they fill the void left by Qing reluctance to engage in note issuance before the 1850s?

Shunzhi Issue: Background and Aftermath

Over the course of its 268 years, the Qing dynasty (1644–1912) issued traditional (read: vertically printed) paper money only twice. The first issue of Qing notes (*chaoguan*) was ordered by the Shunzhi Emperor (r. 1643–1661) not long after the Qing had come to power and amid residual ethnic Han resistance to the Manchu invaders. It was on a small scale, amounting to 120,000 "strings" per annum, and only lasted between 1651 and 1661. After Shunzhi's demise in 1661, calls for resumption of note issuance weakened, although they never completely disappeared (Xiao Qing 1984, pp. 292–294; Lin Man-houng 2006, p. 40).

Peng Xinwei (rep.1988, pp. 807–808) and others (Yang Lien-sheng 1954, p. 68; Shi Yufu 1984, pp. 109–11) suggested that the Manchu emperors were atavistically mindful of the inflationary pressure, which had broken out because of the abuse of paper money during the reign of their ethnic forebears, the Jin dynasty (1115–1234), over North China. While Peng deemed the extant historic materials insufficient to assess the volume, metallic reserves, or precise circulation dynamics of the limited Shunzhi issue, he did state that it had been initiated as response to treasury funding shortfalls occasioned by the Qing's expedition to occupy the island of Zhoushan near modern Shanghai.

In all likelihood, the reemergence of Qing issues in the Shunzhi reign embodied a short-lived monetary policy aberration that defied the overarching pattern of Qing statecraft and was borne out of military exigency. It was in view of this overarching pattern that scholars have recently made the argument that the Shunzhi issue, while short-lived, probably entrenched the Qing reluctance to issue notes because it also proved inflationary (Horesh 2014; Wei Jianyou 1986, p. 83; Peng Xinwei 1958, pp. 556–559). Either way, the Qing refrained from issuing notes for nearly 200 years thereafter.

The Xianfeng Resumption of Paper Money Issues: Background and Consequences

Following the Shunzhi Emperor's demise, it took another 200 years before supporters of fiduciary currency could muster sufficient clout at court to persuade the Xianfeng Emperor (r. 1850–1861) to embrace this course of action on a large scale. Unusual circumstances made the notion of issuing government notes less objectionable: a grave fiscal crisis engulfed the new emperor right from the moment he acceded to power, a crisis brought about by Opium War indemnities and the onset of the Taiping rebellion. Therefore, Xianfeng-era notes (*Hubu guanpiao*) were

issued on a much larger scale than under the Shunzhi Emperor 200 years earlier. At the same time, government foundries released a substantial amount of new *daqian* coins, whose face value exceeded their intrinsic copper worth. Both media of payment proved exceedingly inflationary, disrupting the market exchange rate of copper to silver and fueling popular unrest.

In 1853, the Qing Treasury (*Hubu*) authorized provincial authorities to *directly* disburse silver and copper-denominated notes bearing its seal of approval on a much larger scale. These *Hubu* notes were, however, "dumped" (*lanfa*) on purveyors of services and goods to the imperial bureaucracy and not as readily accepted back in tax payments. Their value depreciated so much that such local issues were curtailed in 1860. At the time, local authorities in Fuzhou also issued and readily converted notes, partly in order to alleviate cash scarcity. However, debased coins would often be paid in return, to the extent that the metallic reserve backing this issue up lost much of its value, thereby compounding inflationary pressures and the longstanding popular mistrust of paper money (Horesh 2014).

The accession of the Tongzhi Emperor to the throne brought this monetary episode to an end. Nonetheless, during the last year of issue, notes with a total face value of no less than 60.2 million silver *taels* had been disbursed, while imperial tax revenue during that time amounted to meagre 86.6 million taels (Ji Zhaojin 2002, p. 20–32).

In a notable study, Yang Duanliu (1962, pp. 112–113) cited memorials to the Xianfeng Emperor himself in order to show how – despite his instruction to back imperial and local *yinqianhao* banknote issues with up to 50% in silver reserves – the actual reserve ratio varied considerably from province to province. Some provinces did not keep a silver reserve at all. Worse still, several provinces including Henan – being situated not very far from the capital – did not accept their own notes when collecting tax. In Yang's view, these inadequate ratio and insubordination to central government edicts were the main reasons why by 1862 all notes had disappeared from circulation. On his part, Ch'ên (1958) suggested that the imperial treasury made the situation much worse by insisting that no more than half of individual tax remittance could be made with Xianfeng notes, the remainder being demanded in hard currency.

As indicated, the Xianfeng-era hyperinflation was not only driven by inadequately backed paper money but was aggravated by the fact that – after initial hesitation – the Xianfeng approved the issue of debased *daqian*. To be sure, because of the Opium War outlay and the subsequent indemnity imposed on the court, Chinese imperial fiscals had already suffered a 9.2 million tael deficit on the eve of the Xianfeng Emperor's accession in March 1850. To understand why he eventually changed his mind and resorted to the historically tainted *daqian*, it is useful to recall that by December that year a much more ominous threat to Qing suzerainty appeared: the Taiping rebellion. In that sense, the late-Qing undoing of historical constraints concerning *daqian* and paper money issues was not so much a policy choice but an emergency measure. Indeed, many of the supporters of note issuance at the Xianfeng court had also been in favor of issuing *daqian* and iron coinage, as a means of alleviating the fiscal

crisis and the shortage of copper cash caused by the Taiping inroads in the southeast (Ch'ên 1958).

The last phase of Qing banknote issuance lasted from 1897 right until the dynasty's collapse in 1912. It was substantively different to the two previous phases because, by then, notes had been imported from Western printers; they were novel in design in Chinese terms (i.e., horizontal-shaped). They were therefore harder to forge. But more importantly, the new notes were not secured by the imperial treasury (*Hubu*) but disbursed in the main by Chinese semiofficial modern banks, explicitly modeled on foreign ones (Cheng Linsun 2003; Ji Zhaojin 2002; Horesh 2009).

Late–Qing Foreign and Indigenous Financial Institution

According to Peng Xinwei (revised edition, 1988, pp. 886–889), as late as 1900, the volume of privately issued notes accounted for no more than 3% of the Chinese currency stock. Even if we add to the tally silver-denominated notes which, according to Peng, had by then been mostly issued by state or modern banks, it is highly unlikely that the volume of Chinese privately issued notes surpassed 6% of the total. Apart from that, the lack of disclosure by foreign banks as to the volume of their notes outstanding in China proper resulted over the years in inflated estimates thereof (Horesh 2014). It would thus be implausible to conclude that the ratio of all notes as part of the entire Chinese money stock was – even as late as 1900 – much higher than 10%.

That said, by the early 1900s, traditionally cast copper "cash" made up only 17.78% of the Chinese currency stock and gradually receded from the marketplace thereafter. The vitality of copper cash to the Chinese hinterland economy hitherto needs to be placed in the context of Western Europe's tiered monetary systems between the thirteenth and eighteenth century, namely, before the spread of steam-powered mints and nation-state territorial currencies (Helleiner 2003; Sargent and Velde 2014). In that context, it could be argued that the Chinese late-Imperial polity was preoccupied to a much greater degree with maintaining grain reserves and making grain affordable, and its attention the copper-silver exchange rate should be seen in that light (Dunstan 2006).

Nevertheless, Peng (revised edition, 1988, pp. 886–889) estimated that foreign trade dollars – among which the Mexican silver dollar was by far the most prominent – had made up nearly a quarter of the Chinese currency stock by the 1900s, i.e., considerably above the share of traditional "cash." The trend in Japan and in the Western-dominated world reversed in the meantime. Namely, the worldwide introduction of government steam-minted coinage as of the mid-nineteenth century had led in fact to a rapid *diminution* in the spread of Mexican-mined silver dollars, particularly in the Americas and the Philippines. So, too, did the concomitant diffusion of the gold standard and the discovery of extensive silver deposits in Nevada, California, and West Africa (Helleiner 2003, p. 38).

The implications of Xianfeng-era central and provincial government intemperance were such that Shanghai's Bund district, because it headquartered large foreign

banks, ended up one of the few places in early twentieth-century China where the use of paper money was standardized. Thus, it was one of the few places where banknotes were readily convertible into silver dollars, or into bullion of verifiable quality, regardless of the political climate at large.

Before the establishment of the first British banks on the *Bund*, Shanghai's monetary predicament was not much different. Since about the mid-eighteenth century, the main financial intermediary in the lower Yangtze basin – by far the most commercialized part of the country – were the *qianzhuang*. Known in English as "native banks" or "money shops," these institutions were typically small proprietorships with unlimited liability, loosely aligned along family and dialect ties and rarely patronized by local authorities (McElderry 1976).

Large money shops issued scrip against individual deposits called *zhuangpiao*. ("shop coupon"), but proximate shops could cash this only after 10–15 days during which couriers would liaise with the issuing shop to rule out fraud. The main business of *qianzhuang* was, however, advancing loans in rural areas to support the exchange of commodities between the seaboard and hinterland.

By comparison, the financial landscape in much of North China remained closer to the central government orbit. Until the fall of the Qing in 1911, the dominant institutions there were what Western observers called Shansi [Shanxi] banks or *piaohao* (remittance houses) in Chinese. The *piaohao* were first set up in the 1820s by merchants from the city of Pingyao. They specialized in long-distance money remittances on behalf of government agencies or the dispatch of officials' emolument (Cheng Linsun 2003, pp. 11–15; Huang Jianhui 1992, 194).

The scope of business of these two types of financial institutions was patently dissimilar: the *qianzhuang* profited from relatively high interest rates charged on unsecured loans to medium-range merchants, while the *piaohao* subsisted on draft commission. For this reason, some *piaohao* were known to deposit idle funds in *qianzhuang* interest-bearing accounts, so that the general relationship between the two was one of cooperation and complementarity rather than of confrontation.

China's economic center of gravity began to shift during the 1850s from Guangzhou to Shanghai. As affluent landlords and officials were seeking refuge from the upheavals of the Taiping rebellion (1851–1864) in the city's foreign concessions along the Bund, *qianzhuang* activity diversified to include investments in real estate and later in stock exchange speculation. Thus, while the dynastic collapse dealt a fatal blow to the Shansi banks in 1911, Shanghai's money shops survived right up to the 1940s (McElderry 1976, pp. 68–69; Cheng Linsun 2003, p. 38).

As from the 1860s, the *qianzhuang* clearly gravitated toward a new source of capital in Shanghai: the foreign banks. By 1888, as many as 62 of Shanghai's largest money shops were borrowing capital from foreign banks to the tune of millions of taels annually. Buoyed by this new source, the money shops transformed themselves into an indispensable conduit for the finance of treaty port trade. Put simply, they on-lent foreign bank funds to Chinese wholesalers of imported goods until the latter sold their stock and paid off the debt. This form of on-call foreign bank credit came to be known as chop loans, or *chaipiao* in Chinese (McElderry 1976, p. 21; Pan Liangui 2004, p. 105).

British, European, and Japanese banks could readily lay down funds in the treaty ports because they had cultivated exclusive relationships with foreign trading houses, ever ready to exchange local money for sterling bills. The foreign banks were otherwise much better capitalized than the diffuse money shops; the credit they advanced reinvigorated treaty port trade, which had often lapsed into barter patterns (McElderry 1976, p. 18–19; Hamashita 1980).

At the same time, foreign banks identified the opportunity afforded by China's fragmented money markets and political instability. By 1911 they mobilized domestic resources on an order of magnitude that exceeded their initial paid-up capital several times over – largely through banknote issuance and deposit receipts. As security for the chop loans, British banks often accepted *zhuangpiao*. It is therefore plausible to assume that the origins of the chop loan contract lie in the fact that these "shop coupons" had been so widely used when the British banks set up their first branches – that the latter could not simply reject them, when presented for encashment by foreign purveyors. Moreover, foreign banks were also compelled to keep an account with at least one money shop, since only the money shop guild could clear the myriad forms of *zhuangpiao* circulating in Shanghai through an elaborate daily mechanism dubbed *huihua*, i.e. "draft exchange" (McElderry 1976, pp. 8–10).

Chinese historians have routinely depicted the foreign bank (*qianzhuang*) nexus as hugely exploitative, because the money shops were thought to have paved the way for foreign inroads into China's economy (e.g., Zhang Guohui 1989). But in a recently published article, Nishimura (2005) has debunked much of this contention by showing that chop loans did not take up the lion's share of money shop activity and were even more marginal on foreign bank balance sheets; at times, the money shops even lent money back to foreign banks on call.

The breakdown of the chop loan mechanism projected its effects not just on the standing of *zhuangpiao* in the money markets – all the other organic, private-order arrangements were badly hit. In the first instance, these arrangements concerned the individual intermediaries employed by foreign firms to guarantee Chinese liabilities like *zhuangpiao* – commonly known to Westerners as "compradors" or *maiban* in Chinese (Ji Zhaojin 2002, pp. 54–57).

The compradoric arrangements embodied organic, private-order mechanism that emerged in the mid-nineteenth-century treaty ports in response to language barriers and information deficits facing foreigners who wished to penetrate Chinese markets. Compradors personally guaranteed *zhuangpiao* and other Chinese liabilities before foreign institutions, but did not have the leverage to guarantee metallic money disbursed by foreign institutions in the Chinese marketplace. In the fractured monetary condition of the times, other highly specialized organic institutions emerged precisely to that end: the *gonguju* and *yinlu* metal assayers (Wagel 1915, pp. 114–118).

Japanese and continental banks, which were latecomers into the Chinese market, may well have been more ready to engage the domestic sector. In his portrayal of commerce in late-Imperial Tianjin, Kwan Man Bun has shown, for example, that the foreign banks, which helped the city's salt merchants tide over losses in the wake of the Boxer rebellion (1899–1901), included the Yokohama Specie Bank, the Banque

Russo-Chinoise, the Deutsch-Asiatische Bank, and the Saigon-based Banque de l'Indochine (Kwan Man Bun 2001, pp. 138–144).

There is another ample evidence to suggest that domestic financial institutions had lent profusely to expatriate-run ventures even when foreigners were not among their main shareholders. The most clear-cut example is the Imperial Bank of China [IBC], 73% of whose 1901 loan portfolio was advanced to expatriate firms (Huang Jianhui 1994, p. 87).

It should be pointed out that the good reputation that the IBC still enjoyed among foreigners at that stage was due to the fact that its patron, official Sheng Xuanghuai, had partly modeled it on HSBC in 1897. Since imperial coffers had been depleted, Sheng sold equity in the new institution to both within and outside the government bureaucracy. Along the way, he overcame concerns raised by prominent official Zhang Zhidong fearing "excessive" private ownership of imperial institutions (Feuerwerker 1958, pp. 228, 232; Chen Limao 2003; Hamashita 1980, p. 459).

The IBC was China's first limited liability bank, the first to employ foreign staff, and the first of many Chinese banks to order notes from overseas printers. In conformity with early HSBC guidelines, IBC's total banknote issue was at no time to exceed paid-up capital, and one third thereof was to be covered by a cash reserve. However, fragmentary balance sheet data suggest that this stipulation may not have been enforced after 1906 (Feuerwerker 1958, pp. 230–240).

Structural disparities between the indigenous and foreign banking sectors partly derived from geographical distribution: Chinese banks had, of course, much more leeway to popularize notes in the vast hinterland that stretched beyond the confines of the treaty ports. Within the treaty ports, foreign banks were protected from Chinese government intervention by virtue of their extraterritorial status. Foreign banks rarely opened branches elsewhere, with the exception of politically sensitive Beijing. But the disparities may have also derived from overemphasis late-Qing reformers had laid at banknote issuance as a definitive constituent of modern banking. The reformers were quick to note how China's institutional weakness invited foreign banks to recoup profits from fiduciary note issuance in the treaty ports. On the other hand, they failed to heed attendant reserve requirements, which set foreign banks apart from traditional financial institutions (Horesh 2009).

The reformers called on the throne to foster government-run banks to counter foreign economic inroads. But their frame of references and argumentation was not materially different from that of Qing bureaucrats who had previously tried to persuade the Manchus to overprint paper money as panacea for the dynastic decline. Nonetheless, the traditional monetary discourse did change to the extent that the early reformers had been castigated during the 1850s, whereas reformers in the 1890s could make more daring suggestions – like the adoption of a gold standard – with impunity (Ye Shichang 2003, pp. 36–43; Yang Duanliu 1962, pp.104–113; Huang Jianhui 1994, pp. 90–91).

In the interim, the dilemma for the fiscally strained Qing court remained the same: how to retain imperial revenue without unleashing inflation that would provoke popular resistance, and without surrendering more central powers. Late-Imperial body politic was often blighted by indecision and contradictory thrusts that

precluded lasting synergy between the state and private spheres in the realm of money. By the mid-nineteenth century, perhaps earlier, this shortfall had opened up a gaping loophole through which privately funded British trading houses and Anglo-Indian financial institutions could thrive on China's coastline (McLean 1976, pp. 292–293, 300–304).

Imperial attempts at regulating banks and currency in Qing China came too little, too late. When financial reforms were finally promulgated, the survival of the dynasty had already been largely beholden to the foreign powers, so that the issue of foreign banknotes and restrictions on foreign banking in general were hardly broached (Huang Jianhui 1994, pp. 100–104). The Mackay Treaty signed between Britain and China on 5 September 1902 included, for example, an imperial pledge to coin uniform currency ". . .which shall be legal tender in payment of all duties taxes and other obligations throughout the Empire by British and Chinese subjects" (Wagel 1915, p. 83; Hall 1921, pp. 20–23). The idea was to remove barriers to trade from coin variation to provincial turnpikes (likin). However, subsidiary coin minting was not so lucrative to begin with. By contrast, the Treaty made virtually no reference to banknote issuance.

The demand for a uniform coinage from both foreign and local merchants certainly made headlines at the turn of the twentieth century. But the freshly minted *tongyuan* coinage (see above) was hardly the solution that they had had in mind. The matter was to be resolved only in the Republican era. Any question of currency reform around 1900 had to determine whether the currency base should be silver or gold. It would have also required the monetary power to be concentrated in the hands of the Qing central government at a time when it was heavily indebted and losing its grip on finance. In that sense, the inflation that the *tongyuan* coinage set in train made up one of the major factors leading to the Republican Revolution of 1911, which eventually toppled the Qing dynasty (Faure 2000).

After the fall of Qing in 1912, the volatility of the European powers' informal *imperia* in East Asia became abundantly clear. To begin with, the Russo-Japanese War (1904–1905) had unanticipated outcomes reverberating throughout the region. One of the implications was a boost for Japanese banknote issuance in China. Then, after World War I, German financial institutions were ejected from the region. In the 1920s, the three principal British overseas banks in East Asia – HSBC, Chartered Bank, and the Mercantile Bank of India – were still unchallenged financially. So too was the British-run stock exchange in Shanghai, which attracted both foreign and Chinese investors (Horesh 2015). Yet with the rise of Chinese nationalism and the phenomenal leap in homegrown modern banking, their future prospects were becoming increasingly uncertain (Horesh 2014).

Cross-References

- ▶ Experiments with Paper Money
- ▶ Monetary System in Ancient China
- ▶ Rise and Demise of the Global Silver Standard

References

Allen RC, Bassino J-P, Ma D, Moll-Murata C, Van Zanden JL (2011) Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. Econom Hist Rev 64(s1):8–38

Broadberry S, Gupta B (2006) The early modern great divergence: wages, prices, and economic development in Europe and Asia, 1500-1800. Econom Hist Rev 51(1):2–31

Burger W (2008) Coin production during the Qianlong and Jiaqing reigns (1736-1820): issues in cash and silver supply. In: Kim N, Hirzel T (eds) Metals, moneys and markets in early modern societies: east asian and global perspectives. LIT Verlag, Münster, pp 171–189

Burger Werner (2015) Japanese and Vietnamese coins circulating in China: a numismatic approach. In: Leonard, Jane Kate, Ulrich Theobald (eds) Money in Asia (1200–1900): small currencies in social and political contexts. Brill, Leiden, pp 220–223

Boulnois L (1983) Poudre d'or et monnaies d'argent au Tibet. Editions du Centre national de la recherche scientifique, Paris

Chen Feng 陈锋 (2008) Qingdai caizheng zhengce yu huobi zhengce yanjiu 清代财政政策与货币政策研究. Wuhan daxuechubanshe, Wuhan

Ch'ên J (1958) The Hsien-Fêng inflation. Bull Sch Orient Afr Stud 21:578–586

Chen Limao 陳禮茂 (2003) Zhang Zhidongzai Zhongguo tongshang yinhang chunagban guochengzhong de yanlun shuping 張之洞在中國通商銀行創辦過程中的言論述評. Anhui shixue 5:29–35

Cheng Linsun (2003) Banking in modern China: entrepreneurs, professional managers and the development of Chinese banks. Cambridge University Press, New York, pp 1897–1937

Cribb J (1992) Catalogue of Sycee in the British museum. British Museum, London

Flynn DO, Giráldez A (2002) China and the birth of globalization in the 16th century. Ashgate, Farnham

Crossley PK (2002) A translucent mirror: history and identity in Qing imperial ideology. University of California Press, Berkeley

Davies, Glyn. [1994, rep. 2002]. A history of money: from ancient times to the present day. University of Wales, Cardiff

Dunstan H (1996) 'Orders go forth in the morning and are changed by nightfall': a monetary policy cycle in Qing China, November 1744–June 1745. T'oungpao 82:66–136

Dunstan H (2006) State or merchant? Political economy and political process in 1740s China. Harvard University Press, Cambridge, MA

Elliott MC (2002) The Manchu way: the eight banners and ethnic identity in late Imperial China. Stanford University Press, Stanford

Elvin M (1973) The pattern of the Chinese past. Eyre Methuen, London

Faure D (2000) The Mackay treaty of 1902 and its impact on Chinese business. Asia Pac Bus Rev 7(2):79–92

Feuerwerker A (1958) China's early industrialization: sheng Hsuan-huai (1844–1916) and the mandarin enterprise. Harvard University Press, Cambridge, MA

Flynn DO, Giráldez A (2002) Cycles of silver: global economic Unity through the mid-eighteenth century. J World Hist 13(2):391–427

Foust CM (1969) Muscovite and mandarin: Russia's trade with China and its setting, 1727–1805. University of North Carolina Press, Chapel Hill

Giráldez A (2015) The age of trade: the Manilla Galeons the Dawn of the global economy. Rowman & Littlefield, Lanham

Golas PA (1999) Mining. In: Needham J (ed) Science and civilisation in China vol. 5, Chemistry and chemical technology, Part 13. Cambridge University Press, Cambridge, UK

Goldstone JA (1991) Revolution and rebellion in the early modern world. University of California Press, Berkeley

Greatrex R (2008) Layers of deception: counterfeiting cases in the mid-Qing. In: Hirzel T, Kim N (eds) Metals, monies, and markets in early modern societies. Lit Verlag, Berlin, pp 309–331

Hall Ray Ovid (1921) Chinese national banks: from their founding to the moratorium. s.n, Berlin

Hamashita Takeshi 濱下武志 (1980) Chūgokutsūshōginkō no setsuritsu to Honkon Shanhaiginkō中国通商銀行の設立と香港上海銀行. Hitotsubashironsō 84.4:448–464
Hao Yen-p'ing (1986) The commercial revolution in nineteenth-century China. University of California Press, Berkeley
He W (2013) Paths toward the modern fiscal state. Harvard University Press, Cambridge, MA
Helleiner E (2003) The making of National Money: territorial currencies in historical perspective. Cornell University Press, Ithaca
Ho Hon-wai 何漢威 (1993) "Cong yinqianhuang dao tongyuan fanlan — Qingmo xinhuobi de faxing ji qi yinxiang" 從銀錢荒到銅元汎濫——清末新貨幣的發行及其影響. Zhongyang yanjiuyuan lishiyuyanyanjiusuo jikan 62.3:389–494
Horesh N (2009) Shanghai's Bund and Beyond: British Banks, Banknote Issuance, and Monetary Policy in China, 1842–1937. Yale University Press, New Haven Conn
Horesh N (2014) Chinese Money in Global Context: historic junctures between 600 BCE and 2012. Stanford University Press, Stanford
Horesh N (2015) Gerschenkron redux? Analysing new evidence on joint-stock enterprise in pre-war shanghai. Asian-Pacific Econom Liter 29(1):25–46
Hou Houji 侯厚吉, Wu Qijing 吴其敬 (1982) Zhongguo jindai jingji sixiang shigao 中國近代經濟思想史稿. Heilongjiang renminchubanshe, Haerbin
Huang Jianhui 黃鑒暉 (1992) Shanxi piaohaoshi. Shanxi jingjichubanshe, Taiyuan
Huang Jianhui 黃鑒暉 (1994) Zhongguo yinhangyeshi 中國銀行業史. Shanxi jingji chubanshe, Taiyuan
Ichiko Shōzō 市古尚三 (2004) Shindai kaheishikō 清代貨幣史考. Hōshobō, Tokyo
Irigoin A (2009) The end of a silver era: the consequences of the breakdown of the Spanish peso standard in China and the United States, 1780s–1850s. J World Hist 20(2):226–229
Cao Jin, Vogel HU (2015) Smoke on the mountain: the infamous counterfeiting case of Tongzi district, Guizhou province, 1794. In: Leonard JK, Theobald U (eds) Money in Asia (1200–1900): small currencies in social and political contexts. Brill, Leiden, pp 188–219
Ji Zhaojin (2002) A history of modern Shanghai banking. M.E. Sharpe, Armonk
Kann E (1927) The currencies of China. Kelly & Walsh, Shanghai
King FHH (1965) Money and monetary policy in China, 1845–1895. Harvard University Press, Cambridge
Kishimoto Mio 岸 (1997) Shindaich¯ugoku no bukka to keizaihend¯o清代中国の物価と経済変. Kenbun, Tokyo
Kwan Man Bun (2001) The salt merchants of Tianjin: state-making and civil society in late Imperial China. Hawai'i University Press, Honolulu
Kuroda, Akinobu 黒田明伸 (2003) Kaheishisutemu no sekaishi : "hitaishosei" o yomu 貨幣システムの世界史：〈非対称性〉をよむ. Iwanami Shoten, Tokyo
Lin Man-houng (2006) China upside down: currency, society and ideologies. Harvard University Press, Cambridge, MA, pp 1808–1856
Lin Man-houng (2015) The devastation of the Qing Mints, 1821–1850. In: Leonard, Jane Kate, Ulrich Theobald (eds) Money in Asia (1200–1900): small currencies in social and political contexts. Brill, Leiden, pp 155–187
Liu Qiugen 劉秋根 (2000) Ming Qing gaolidaiziben 明清高利貸資本. Shehui kexue wenxian chubanshe, Beijing
McElderry AL (1976) Shanghai old-style banks (Ch'ien-Chuang), 1800–1925: a traditional institution in a changing society. Center for Chinese Studies, University of Michigan, Ann Arbor
McLean D (1976) Finance and 'informal empire' before the first world war. Econom Hist Rev 29(2):291–305
Millward J (1998) Beyond the pass: economy, ethnicity, and empire in Qing Central Asia, 1759–1864. Stanford University Press, Stanford
Morse HB (1906) Currency in China. Kelly & Walsh, Shanghai
Mu Yuan 穆淵 (1994) Qingdai Xinjiang huobishi 清代新疆货币史. Xinjiang daxue chubanshe, Urumqi

Nishimura S (2005) The foreign and native banks in China: chop loans in Shanghai and Hankow before 1914. Mod Asian Stud 39(1):109–132
Pan Liangui 潘連貴 (2004) Shanghai huobishi 上海貨幣史. Shanghai renminchubanshe, Shanghai
Peng Xinwei 彭信威 (1958) [rev. 1970, 1988, 2007] Zhongguo huobishi 中國貨幣史. Shanghai renmin chubanshe, Shanghai
Peng Xinwei 彭信威 (1994) A monetary history of China (trans: Edward H. Kaplan). Western Washington, Bellingham
Perkins D (2017) Agricultural development in China, 1368–1968. Routledge, London
Perdue PC (2009) China marches west: the Qing conquest of Central Eurasia. Harvard University Press, Cambridge
Pomeranz K (1993) The making of a hinterland: state, society and economy in inland North China, 1853–1937. University of California Press, Berkeley
Pomeranz K (2001) The great divergence. Princeton University Press, Princeton
Quan Hansheng 全漢升 (1987) Ming Qing jingjishiyanjiu 明清經濟史研究. Lianjing chuban shiye gongsi, Taipei
Rang Huey-shin [Zhang Huixin] 張惠信 (1988) Zhongguoyinding 中國銀錠. Liu Qiuyan, Taipei
Rawski TG (1989) Economic growth in prewar China. University of California Press, Berkeley
Rosenthal J-L, Wong RB (2011) Before and beyond divergence: the politics of economic change in China and Europe. Harvard University Press, Cambridge
Rowe WT (2005) Provincial monetary practice in eighteenth-century China: Chen Hongmou in Jiangxi and Shaanxi. In: Moll-Murata C, Song J, Vogel HU (eds) Chinese handicraft regulations of the Qing dynasty. Iudicium, München
Rowe WT (2009) China's last empire: the great Qing. Harvard University Press, Cambridge, MA
Sargent TJ, Velde FR (2014) The big problem of small cash. Princeton University Press, Princeton
Shi Yufu 石毓符 (1984) Zhongguo huobi jinrong shilüe 中國貨幣金融史略. Tianjin renmin chubanshe, Tianjin
Shimada R (2006) The intra-Asian trade in Japanese copper by the Dutch East India company during the eighteenth century. Brill, Leiden
Van de Ven H (2014) Breaking with the past: the maritime customs service and the global origins of modernity in China. Columbia University Press, New York
Vissering G (1914) On Chinese currency: preliminary remarks on the monetary and banking reform in China, vol 2. De Bussy, Amsterdam
Vogel HU (1987) Chinese central monetary policy, 1644-1800. Late Imperial China 8(2):1–52
Vogel HU (2012) Marco polo was in China: new evidence from currencies, salts and revenues. Brill, Leiden
Vogel HU, Theisen-Vogel E (1989) Der Kupferbergbau in der chinesischen ProvinzYunnan vom 18. Bis zur Mitte des 19 Jahrhunderts: Produktion, Administration, Finanzierung. Anschnitt 41(5):146–158
Von Glahn R (2007) Foreign silver coins in the market culture of nineteenth century China. Int J Asian Stud 4(1):51–78
Von Glahn R (2016) The economic history of China: from antiquity to the nineteenth century. Cambridge University Press, Cambridge
Wagel SS (1915) Chinese currency and banking. North China Daily News and Herald, Shanghai
Waley-Cohen J (2000) The sextants of Beijing: global currents in Chinese history. Norton, New York
Wang Yeh- chien [Wang Yejian] 王業鍵 (1981) Zhongguo jindai huobi yu yinhang de yanjin (1644–1937) 中國近代貨幣與銀行的演進 (1644–1937). Academia Sinica, Taipei
Wei Jianyou 魏建猷 (1986) Zhongguo jindai huobisi 中國近代貨幣史. Huangshan shushechuban
Xian Ke 獻可 (1958) Jinbainianlai diguozhuyi zai Hua yinhang faxing zhibi gaikuang 近百年來帝國主義在華銀行發行紙幣概況 [a survey of banknote issuance by imperialist banks in China in the last century]. Shanghai renmin chubanshe, Shanghai
Xiao Qing 蕭清 (1984) Zhongguo gudai huobishi 中國古代貨幣史. Renmin chubanshe, Beijing

Yang Duanliu 楊端六 (1962) Qingdai huobi jinrong shigao 清代貨幣金融史稿. Shenghuo, Beijing
Yang Lien- sheng (1954) Money and credit in China: a short history. Harvard University Press, Cambridge, MA
Ye Shichang 葉世昌 (2003) Gudai Zhongguo jingji sixiangshi 古代中國經濟思想史. Fudan daxue chubanshe, Shanghai
Zelin, Madeleine (1984) *The Magistrate's Tael: Rationalizing Fiscal Reform in Eighteenth-Century Ch'ing China*. Berkeley: University of California Press
Zhang Guohi 张国辉 (1989) Wan Qing qianzhuang he piaohao yanjiu 晚清钱庄和票号研究, Zhonghua shuju, Beijing
Zhang Jiaxiang 張家驤 (2001) Zhongguo huobi sixiangshi 中國貨幣思想史. Hubei renminchubanshe, Wuhan
Zheng Jin 鄭瑾 (2007) Zhongguo gudai weibi yanjiu 中國古代偽幣研究. Zhejiang da xuechu ban she, Hangzhou
Zhihong Shi (2016) Central government silver Treasury: revenue, expenditure and inventory statistics, ca. 1667–1899. Brill, Leiden, p 2016
Zhongguo lidai huobi 中國曆代貨幣 (1999) Comp. by The People's Bank of China. Xinhua chubanshe, Beijing

The Monetary System of Japan in the Tokugawa Period

23

Hisashi Takagi

Contents

Abstract

This chapter describes the monetary system in the Tokugawa period – from the seventeenth to the mid-nineteenth centuries – with an observation of its relationship with the periods before and after.

H. Takagi (✉)
Faculty of Letters, Yasuda Women's University, Hiroshima, Japan
e-mail: takagi-h@yasuda-u.ac.jp

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_53

The monetary system in the Tokugawa period is known as the "triple standard system," or the co-distribution of gold, silver, and *zeni*, a perforated coin made of bronze, iron, or brass. Published research has described this development process separately from the experiences in the preceding century, but this chapter emphasizes the fact that the triple standard system inherited a monetary system in the sixteenth century that was autonomously developed in this society.

Regarding the Tokugawa period, this chapter focuses on the currency policies of the Tokugawa shogunate, especially its frequent currency standard revisions.

Additionally, it highlights the continuity of the monetary system between the Tokugawa period and the subsequent Meiji Restoration, or modern Japan. This chapter specifically focuses on the distribution of various nominal metal and paper currencies and the de facto approach to the modern gold standard system in the Tokugawa period. These facts have facilitated a transition to the modern monetary system after the restoration.

Keywords
Triple standard system · Gold currency · Silver currency · *Zeni* · Paper currency

Introduction

This chapter describes the monetary system in the Tokugawa period from the seventeenth to nineteenth centuries, or the early-modern period in Japanese economic history. Conventional historical descriptions have tended to separately describe the monetary systems in the Tokugawa period and before the sixteenth century; however, this chapter emphasizes the continuity in both periods. Regarding the Tokugawa period, this chapter focuses on the currency policies of the Tokugawa shogunate, and especially on its frequent currency standard revisions. Additionally, this chapter highlights the continuity of monetary systems between the Tokugawa period and after the Meiji Restoration, or modern Japan.

Triple Standard System

The Tokugawa period's monetary system is known as the "triple standard system." A majority of published research defines the triple standard system as the co-distribution of gold, silver, and *zeni* – a perforated coin primarily made of bronze, called "*qian*" in Chinese – as currency, with no relationship between standard and subsidiary uses between the three currencies (Takagi 2017, 2018a).

In this definition, epoch-making events are listed as follows: the co-distribution of gold, silver, and *zeni* as currency during the 1570s and 1590s, and Kasuga Taisha's (Kasuga shrine, Yamato province, now Nara prefecture) decree in the early sixteenth century and Oda Nobunaga's decree in 1569, which defined the rates between gold, silver, and *zeni* (Takagi 2010). The issuance of *Keicho-gold* and *Keicho-silver*

(described later) around 1601 by the Tokugawa shogunate is also noteworthy in this definition.

More specifically, this primarily involves not only the co-distribution of the three currencies, but also their use as a standard currency or unit of account to accompany classification, regional, or usage differences; the Tokugawa shogunate's integration and designation, supply, and distribution of gold, silver, and *zeni*; the shogunate's definition of the official rates between the three currencies; the priority given to the shogunate's currencies, except those issued by the shogunate; and the shogunate's monopoly on currency issuance rights (Iwahashi 2002).

The epoch-making events in this definition are the Tokugawa shogunate's decrees in 1608 and 1609, which defined the rates between gold, silver, and *bita* (a subcategory of *zeni*, to be described later), and the issuance of *Kan'ei-tsuho*, a type of *zeni*, by the Tokugawa shogunate in 1636.

Additionally, domain currency was distributed at the beginning of the early-modern period. Domain currency was issued by the *daimyo* (feudal domain lord) in each *han* (the *daimyo*'s domain) and consisted of domain gold, domain silver, domain *zeni*, and domain note. Domain currencies and the three currencies issued by the Tokugawa shogunate never equally coexisted; in other words, the three kinds of currency issued by the Tokugawa shogunate took precedence, both politically and economically. However, there was never a perfect nationwide integration of the three currencies designated by the shogunate, and the other currencies were not perfectly eliminated (Yasukuni 2016). Further, it is well known that the three currencies' legal rates, as defined by the Tokugawa shogunate, were only applied for shogunate officials, and the shogunate approved the independent market rate.

Throughout the Tokugawa period, the triple standard system consisted of more than only three metals. The Tokugawa shogunate issued silver currencies, denominated in such gold mass units as the *bu* (approximately 4.1 g, or 1/4 *ryo*) and *shu* (approximately 1.0 g, or 1/4 *bu*). The Tokugawa shogunate also issued *Kan'ei-tsuho*, which were made of iron or brass. Moreover, the *daimyo* and merchant issued paper currencies denominated in such silver mass units as the *monme* (approximately 3.75 g) or *fun* (approximately 0.375 g, or 1/10 *monme*). Further, the *daimyo* and other subjects issued paper currencies denominated in *monme* units on the face, although the denomination substantially implied the *mon* unit (basic count unit of the *zeni*), and the specie for redemption was the *zeni* (Iwahashi 2002).

The triple standard system, in other words, evolved to consist of the *ryo, monme*, and *mon* system unit currencies, regardless of the type of material used for the currency itself (Takagi 2017).

Prehistory: Sixteenth Century

In order to explain the process of development for the monetary system in the Tokugawa period, we must first describe the currency in the sixteenth century, the century before the establishment of the shogunate (Takagi 2016).

We must first describe *zeni* among the factors that comprise the triple standard system. In the seventh century, the Japanese ancient imperial court issued *zeni* that were modeled on the Chinese dynasty's *zeni*, but the imperial court stopped producing *zeni* in the tenth century due to poor copper production and a changing financial system.

Common people used Chinese *zeni* since the eleventh century as currency imported by the private sector. However, the amount of imported *zeni* did not meet the demand for transactions. Therefore, Japanese craftsmen had manufactured imitation Chinese *zeni* on the side since the fourteenth century, and these were circulated as currency.

In the definition below, domestic *zeni* issuance took place during the ancient period; the dependence on import *zeni* in the medieval period; and the resumption of domestic currency's issuance in the early-modern period; the fourteenth century is broadly defined as the beginning of the early-modern period in Japanese monetary history, although gold and silver currencies were not yet issued (Takagi 2018b).

Entrepreneurs in the sixteenth century hired craftsmen to produce imitation Chinese and noncharacter *zeni* (Sakuraki 2009). This is contrary to the fact that until the fifteenth century, domestic *zeni* were produced by craftsmen on the side, in their own workshops. It can be said that noncharacter *zeni* were Japan's own currency independent of the Chinese monetary system, as noncharacter *zeni* were accepted without the Chinese *zeni* inscription (Nakajima 1999).

The Muromachi shogunate and its *daimyo* frequently prohibited or restricted the passing of domestic *zeni* during the late fifteenth century and into the sixteenth century. This fact paradoxically implies that domestic *zeni* were circulated in reality and compensated for the shortage of *zeni* (Takagi 2017). The domestic *zeni* production described above was the background against which the Tokugawa shogunate issued *Kan'ei-tsuho* later.

The *bita* then emerged in the 1570s as a subcategory of *zeni* and a direct predecessor to the *Kan'ei-tsuho*. Broadly, the *bita* includes both imported and domestic *zeni* and is defined as all *zeni* except for the conventional standard *zeni* evaluated as one *mon* per one piece. Among them, all except for the lowest quality *zeni* circulated (*bita*, in a narrow definition). At the beginning of its appearance, the *bita* was evaluated as less than one *mon*, but passed as one *mon* in society after the late 1570s due to a decrease in the conventional standard *zeni* supply (Takagi 2010, 2017, 2018a).

The *bita* was first politically designated as standard *zeni* by Oda Nobunaga, the re-unifier of early-modern Japan, in 1580. Toyotomi Hideyoshi, who inherited Oda's reunifying enterprise and was formerly his vassal, also inherited this policy.

Regarding gold in the sixteenth century, many gold mines were developed in eastern Japan. During this time, the oval and plate-like gold ingot, manufactured on demand, was used as a large-denomination currency and a means of stock of value, as an alternative to gold dust. The ingot had the same shape as the *koban*, a gold currency issued by the Tokugawa shogunate (to be further described later), and could be considered its prototype.

Conventional gold ingots and dust were transacted by weight, in *ryo* units. However, the ingot's mass and gold content were autonomously standardized in

society in the sixteenth century, or without governmental administration. In terms of mass standardization, this approaches the currency by table. Additionally, gold merchants signed the gold ingots in ink and incused to certify their mass and content. This act reduced transaction costs of shroffing in each transaction.

A representative gold currency was produced in the late sixteenth century by the Goto family, who were metal craftsmen in Kyoto. Toyotomi Hideyoshi designated the gold currency manufactured by the Goto family as official in the 1590s, which implies that the government approved the standardization of gold currency as generated in the private sector.

Regarding silver, a silver rush occurred in Japan in the first half of the sixteenth century with the development of the Iwami silver mine in what is now the Shimane prefecture, which has since been registered as a World Heritage site. Japan subsequently transformed into a silver exporter, and especially to China. Silver was later distributed as currency in Japan during the 1570s and the 1590s due to decreasing prices as a result of competition with South American silver and an expanding demand for domestic currency.

The silver currency was the sea cucumber-shaped ingot, which was transacted by weight in *monme* units. As with gold currency, silver ingot's mass became standardized, and silver merchants marked the ingots to certify their content. Toyotomi Hideyoshi then designated the silver currency manufactured by the Daikoku family, silver craftsmen in Sakai (now Osaka prefecture), as official in the 1590s.

The First Issuance of Gold and Silver Currencies by the Tokugawa Shogunate: The Tokugawa Ieyasu and Hidetada Regime

Now, we describe monetary policies by the Tokugawa shogunate.

The *Ryo* system unit's standard gold currency issued by the Tokugawa shogunate consisted of the *koban* (the shape is oval and the denomination is one *ryo*) and *one-bu-gold* (the shape is vertically long and rectangular, and the denomination is one *bu* equaled 1/4 *ryo* and 1/4 mass of *koban*). The gold content of *one-bu-gold* was the same as the *koban*, and the standards for the *koban* and *one-bu-gold* were simultaneously revised.

The *monme* system unit's standard silver currency was issued by the Tokugawa shogunate and consisted of the *chogin* (sea cucumber-shaped; gin means silver) and *kodamagin* (round shaped and smaller than the *chogin*), which were both of an indeterminate mass and transacted by weight. The *chogin* and *kodamagin* both had the same silver content, and both standards were simultaneously revised. Unless otherwise noted later, the phrase "(*era name*) + *gold*" indicates the set of "(era name) + *koban* and *one-bu-gold*," and "(*era name*) + *silver*" indicates the set of "(era name) + *chogin* and *kodamagin*."

Tokugawa Ieyasu was a former alliance partner of Oda Nobunaga, a provincial load in the Toyotomi regime and the first shogun in the Tokugawa period (r. 1603–1605). Ieyasu won the battle of Sekigahara in 1600 and established substantial supremacy, and his son Tokugawa Hietada inherited the post in 1605 as the second shogun (r. 1605–1623). However, even after this inheritance, Tokugawa

Ieyasu continued to significantly influence political decisions in the shogunate until his demise in 1616.

The first currencies issued by the Tokugawa regime were the *Keicho- gold and silver*, in around 1601 (*Keicho* 6th). The issuance of gold as a domestic currency was unusual in East Asia at the time, as silver currencies were circulated as an international currency (Kuroda 2014a).

Keicho-gold was manufactured under the administration by the Goto family, who had produced gold currency in the Toyotomi regime. *Keicho-gold* inherited the standard of gold currencies manufactured in *Kai* province, now Yamanashi prefecture, where Tokugawa Ieyasu had dominated in the Toyotomi regime.

Keicho-koban had no ink signature, but only marks. Unlike gold currencies in the sixteenth century, this was suitable for mass production. After this time, *koban* was issued in large amounts, and not custom produced, regardless of individual demand.

The Daikoku family was adopted as the moneyer of *Keicho-silver*, as they had produced silver currency in the Toyotomi regime. Although the *Keicho-chogin* is of an indeterminate mass, it is assumed that one piece had a standard of approximately 43 *monme*, or 161 g. This reflects the standard of silver currencies autonomously manufactured in the late sixteenth century. Additionally, the market rate in the late sixteenth century was 43 *monme* silver to 1 *ryo* gold, or 1 *koban* (Iwahashi 2002). Therefore, it can be noted that although *Keicho-chogin* is a currency issued by weight, it was also characterized similar to the currency by table.

A strengthening administration of gold and silver mines and bullion provided a foundation for the Tokugawa Ieyasu regime's issuing of gold and silver currencies. Tokugawa Ieyasu requisitioned main domestic gold and silver mines, including Iwami, soon after the battle of Sekigahara. The monopolizing of gold and silver mines was also an extension of the Toyotomi regime's policy, and in 1609, the shogunate prohibited rural provinces' smelting and forging of gold and silver currencies.

Keicho-gold and silver differ from the gold and silver currencies in the sixteenth century in terms of their amount and constant production by a permanent organization. However, the Toyotomi regime's inherited policy was reflected by the appointment of the Goto and Daikoku families as moneyers (Yasukuni 2016). Moreover, the Toyotomi regime's policy initially involved the approval of preceding social practices. Therefore, the Tokugawa shogunate's gold and silver currency-producing system is an extension of social practices in the sixteenth century.

The Tokugawa shogunate did not issue *zeni* at the time the *Keicho-gold and silver* were issued. This implies that the issuance of *Keicho-gold and silver* was primarily based on fiscal demand and not on the transaction demands of the common people. Subsequent Tokugawa regimes inherited this attitude, which did not necessarily intend to fulfill the common people's demand for currency.

The shogunate also promoted the distribution of *Keicho-gold and silver*. For example, the shogunate in 1611 prohibited Kyoto from passing silver currency, except for *Keicho-silver*.

However, *Keicho-gold and silver* did not immediately circulate nationwide. The shogunate supplied *Keicho-gold and silver* as a fiscal expenditure in direct-control

areas, such as Edo (now Tokyo), which was the Tokugawa shogunate's headquarters and the substantial capital of Japan in the Tokugawa period. In the beginning of the seventeenth century, transactions were not substantially active between the shogunate's direct control area and other provinces; thus, *Keicho-gold and silver* were difficult to distribute to rural areas.

Moreover, some *daimyos* in the seventeenth century issued their own silver currencies and *zeni* (domain currency) without necessarily obtaining permission from the shogunate. In principle, domain currency was valid only in the domain, but was circulated out of its issued domain in practice. The private sector also produced gold and silver bullion, which circulated as currency in various places. These phenomena were due to shortages of *Keicho-gold and silver* (Watanabe 2002).

There were also regional differences in the distribution of gold and silver currencies. Roughly, prices in high-value transactions were denominated in *ryo* system units in eastern Japan, including Edo, and in *monme* system units in western Japan, including the Kamigata region of Kyoto, Osaka, and the areas around these cities. This social practice continued until the Meiji Restoration.

People in eastern Japan used the *ryo* system unit because there were many gold mines in eastern Japan. Further, the shogunate inherited the gold currency system in *Kai* province before the establishment of the shogunate. Western Japanese society used the *monme* system unit due to the region's many silver mines and its social practice of silver currency usage since the sixteenth century, as well as the demand for means of trade, among other reasons.

The difference of standard currency between the two geographically distant major markets – Edo and Kamigata – developed an exchange business for different currencies: the currency changer. Further, the difference and distance created a currency remittance system between the two major markets. These mediated the transactions between Edo merchants who demanded commodities and the Kamigata merchants who supplied them.

The Tokugawa Hidetada regime then designated *bita* as its standard. In 1608 and 1609, the shogunate issued edicts that defined *bita* as every *zeni* except for those of the lowest quality – with rates between *Keicho-gold*, *Keicho-silver*, and the *bita* (1 gold *ryo* = silver 50 *monme* = 4000 *mon* in *bita*) – and approved these rates for use on tax payments and transactions. This rate was based on the domestic market rate since the fourth quarter of the sixteenth century (Iwahashi 1998).

One-bu-gold, or the smallest unit of *Keicho-gold*, equaled *bita* 1000 *mon* according to the official rate. At the time, the daily wage of an expert carpenter in Kyoto was approximately 1 *monme*, equal to *bita* 80 *mon* by the official rate. Thus, the gold currency was too expensive for common people to use in daily transactions (Takagi 2016). The common people in eastern Japan, where the *ryo* system unit's currency was standard, primarily used *zeni* after these edicts. On the other hand, the denomination of silver currency was relatively small (1 *monme* = 80 *mon*). Therefore, common people in western Japan, where the *monme* system unit's currency was standard, used silver currency in their daily transactions.

Incidentally, the aforementioned rate was only official, such as in tax payments. Separately, the market rate included those from the currency changers, as the

shogunate did not prohibit the passing of gold and silver currencies at the market price in commercial transactions.

Tokugawa Ieyasu's designation of the *bita* as the standard *zeni* is an extension of policies from the Oda and Toyotomi regimes. As described earlier, their policies reflected an approval of social practices, and Tokugawa Ieyasu's policy is also an extension of social practices in the sixteenth century. Therefore, it can be said that the social practices and *zeni* policies from the preceding early-modern regimes in the sixteenth century are a predecessor of the Tokugawa shogunate's *zeni* policy.

These edicts developed a system in which the shogunate defined the standards of three kinds of currencies and the rates between them. These edicts were epoch-making policies in developing the triple standard system. They note a system developed in which gold and silver currencies were designated as the standard, and the rates between two currencies were defined by national law, similar to modern bi-metalism.

Additionally, the rates between gold, silver, and *zeni* were already officially defined by the Kasuga Taisha shrine and Oda Nobunaga in the sixteenth century, as aforementioned. Thus, the rate was not first defined politically in the seventeenth century.

The Issuance of *Kan'ei–tsuho*: The Tokugawa Iemitsu Regime

Tokugawa Iemitsu, a son of Tokugawa Hidetada, was the third shogun (r.1623–1651) and prohibited the Japanese people's travel overseas. He allowed only the Netherlands to trade as a European partner as a national isolation policy.

Simultaneous to these policies, the shogunate issued *zeni* in 1636 (*Kan'ei* 13th), called *Kan'ei-tsuho*. Thus, a system developed in which the shogunate supplied three kinds of currencies.

The *Kan'ei-tsuho* at the time was made of bronze, with a denomination of one *mon*. The shogunate adopted its Japanese-era name for the *zeni*'s inscription: *Kan'ei*. The design of this Japanese-era name clearly reflects a withdrawal from the Chinese *zeni* system; this contrasts with the conventional domestic *zeni,* which were primarily an imitation of Chinese *zeni* (Yasukuni 2016). The era name changed from the *Kan'ei* to the *Shoji* in 1644, but the *Kan'ei*'*-tsuho* continued production afterward, until the fall of the Tokugawa shogunate.

The context for the shogunate's issuance of *Kan'ei-tsuho* includes a shortage of *zeni* since the late sixteenth century and a copper rush in seventeenth-century Japan (Sakurai 2017). The second context is the alternate attendance system instituted in 1635, in which *daimyo* were forced to alternately reside between Edo and the domain at regular intervals. The demand for small-denomination currency for food purchases and accommodation fee payments in warriors' marches and merchants' change payments propelled the issuance of *Kan'ei-tsuho*. At the time, *zeni* supply was short; thus, it appreciated against gold and silver. This was disadvantageous for *daimyo* who exchange gold or silver for *zeni* and make payment with *zeni* in the

march. In that respect, the *Kan'ei-tsuho* issuance aimed to fill the demands of the shogunate and *daimyo* and not to facilitate the common people's daily transactions.

The shogunate proclaimed the official rate of the *Keicho-gold* as one *ryo* = *Kan'ei-tsuho* 4000 *mon* = *bita* 4000 *mon*, and to use this rate for tax payments in both the shogunate's direct control area and *daimyo*'s domain. This implies that the shogunate evaluated the *Kan'ei-tsuho* and *bita* as the same and designated *Kan'ei-tsuho* as the standard *zeni*. In other words, the value of the *bita*, which appeared in the latter half of the sixteenth century and was adopted by the Oda and Toyotomi regimes as the standard *zeni*, was carried over to the *Kan'ei-tsuho*. It can be observed that the *Kan'ei-tsuho* originated as a historical experience from the sixteenth century. Further, the passing of the *bita* was not suspended at the time of the *Kan'ei-tsuho*'s issuance.

The shogunate formed temporary contractor organizations to produce *Kan'ei-tsuho* (*zeni-za*) and established them nationwide. Subsequently, the shogunate recruited contractors for every *zeni* issuance project and entrusted them to a limited production. The private sector was in charge of manufacturing as an extension of domestic imitation *zeni* production since the fourteenth century, with a special focus on pure-play production in the sixteenth century.

The shogunate prohibited the production of *zeni* except by designated contractors. Additionally, the *daimyo*'s issuance of *zeni* needed the shogunate's permission. However, the private production of *zeni* continued. It should be noted that privately produced *zeni* at the time were iron *zeni*, circulated in Ise province, now Mie prefecture, in the 1640s to 1650s (Chieda 2011a). This is the oldest iron coin circulated in Japan that can be observed in old documents, and implies that the social demand for currency had increased to the extent that even unofficial *zeni* without durability (iron has less durability than bronze) were required.

The Currency Integration Process: The Tokugawa Ietsuna Regime

Under the regime of Tokugawa Ietsuna – a son of Iemitsu, the fourth shogun (r.1651–1680) – the shogunate currency's nationwide distribution and integration progressed.

Currency, and especially silver currency, became scarce during this time due to exportation, a decrease in metal output, and an increase in demand for currency. Thus, the shogunate had implicitly approved the distribution of gold and silver currencies except for those the shogunate had issued.

However, the Tokugawa Ietsuna regime changed this policy. The shogunate prohibited the trading of bullion in the 1660s and forced people to exchange them with *Keicho-gold and silver*. These policies were implemented to acquire materials for *Keicho-gold and silver* by restricting the free trade of gold and silver.

At this time, gold and silver currencies – except for the shogunate's currency – exited the market in rural provinces, regardless of the regulations enacted by the shogunate. The *daimyos* suspended the passage of domain currencies and promoted the shogunate currency to pass due to a decrease in gold and silver productions.

Additionally, the demand for the shogunate's currency, the interdomain common currency, increased due to the development of a nationwide commerce network. This also promoted the exit of gold and silver currencies, except for the shogunate's.

The integration of *zeni* by the *Kan'ei-tsuho* also progressed due to its avoidance of the *bita*. Although the shogunate had evaluated the *bita* as equal to the *Kan'ei-tsuho*, as aforementioned, the *bita*'s market rate depreciated to half of the *Kan'ei-tsuho* in the period from the 1640s to the 1650s (Fujii 2014). This forced the *bita* to exit from the market.

Daimyos also promoted the distribution of the *Kan'ei-tsuho* in their domains. For example, the Kanazawa domain – now Ishikawa and Toyama prefectures – mandated that the *Kan'ei-tsuho* be used with a small deal in the 1650s in order to eliminate poor-quality silver currencies exported from other domains unintended (Nakano 2012). Specifically, the *daimyo* adopted the shogunate's currency to administrate the currency in the domain.

Tokugawa Ietsuna's regime increased the *Kan'ei-tsuho*'s production based on an increase in domestic copper production and the start of domestic tin mining. Thus, Japan could domestically acquire all materials of *zeni* (Kuroda 2014b). The *bita* recovered by the shogunate was also used as material for the *Kan'ei-tsuho*.

The shogunate suspended the *bita* as passable in 1670. In other words, the *Kan'ei-tsuho* was the only *zeni* permitted as passable. Imported and domestic imitation *zeni* were circulated since the twelfth to fourteenth centuries regardless of the governmental administration and sustained the Japanese currency system. However, import *zeni* was finally prohibited, as the *bita* included imported *zeni*. In a broad sense, Japanese currency's medieval period was over. As all three currency types became exclusively domestic, this policy is also epoch-making in the development of the triple standard system.

Additionally, Tokugawa Ietsuna's regime also promoted the nationwide integration of volume and mass standards. Regarding the integration of social standards, these policies are common to currency integration (Takagi 2016) and facilitated the transaction and use of currency by weight.

The Emergence of Paper Currency: Private and Domain Notes

The seventeenth century is also an epoch-making period in Japanese monetary history regarding the appearance of paper currency, similar to that in the modern era. The representative examples are the private note, issued by the private sector, and the domain note, which the *daimyo* issued as a domain currency. Both of these were primarily issued in western Japan. Consequently, the face value is mostly denominated in *monme* system units.

The private note appeared prior to the domain note, with the *Yamada-hagaki* as the first example around 1610 in Ise province. In East Asia, China had issued paper currency in the eleventh century, while Korea and Vietnam had issued them in the fifteenth century; thus, Japan's issuance was relatively late. In world history, the issuance of paper currency tends to increase in wartime, but Japanese paper currency

appeared in a time of peace, which is characteristic of paper currency in Japan (Shikano 2011).

The *Yamada-hagaki* was issued by agencies and the accommodation industry in Yoda (now Mie prefecture), a shrine city of Ise-jingu shrine (Chieda 2011b). The *Yamada-hagaki* has a vertically long, rectangular shape, similar to the shape of bills of exchange in medieval Japan. This shape was also adopted by subsequent private and domain notes.

Almost all of the face value of the *Yamada-hagaki* is less than one *monme*, or relatively small, with a format that promises a conversion with the shogunate's silver currency. While bills of exchange are issued on demand, the *Yamada-hagaki* was issued in large quantities regardless of the demand from individual users, thus approximating modern paper currency. Moreover, the issuer assumed that the *Yamada-hagaki* would be used by visitors of Ise-jingu. As a response to the common people's demand for currency, the *Yamada-hagaki* differs from the shogunate currency and domain note, in that the primary objective is to satisfy fiscal demand.

In the first half of seventeenth century, various private notes were issued in the Ise province and Kinai region, now the Kansai region, except for the *Yamada-hagaki*. Private notes in the Kinai region were primarily issued by entrepreneurs of civil engineering projects, such as river cuts, and used to pay wages for construction laborers.

The issuance occurred due to: (1) a shortage of small denomination currency used by common people, such as silver currency and *zeni*, like other paper currencies commonly appeared in world history with the shortage of metal currency; (2) a difficulty in producing silver currency less than one *monme*, as these were too small; and (3) saving on silver currency transaction costs, such as transportation or weighing (Shikano 2011).

If paper currencies are issued as a substitution for small metallic currencies, the face value should be in *mon* units, or units of *zeni*. Issuers adopted the *monme* system unit due to the social practice of using the *monme* system unit in accounts, while both *zeni* and *monme* system units' silver currencies circulated as a means of exchange.

Among the silver currencies, the fact that the shogunate's was subject to redemption implies that this particular silver currency was adopted as the standard in social practice; in other words, the shogunates' currency integration had progressed.

Another point of focus involves paper currency manufacturing costs, which are not always historically cheaper than that of metal currency. The issuance of paper currency is reasonable when papermaking and printing technologies can manufacture cheaper than metallic currencies. Further, paper currencies' endurance period is generally shorter than metal currency, and a continuous supply of paper currency is needed to meet demand. Private notes appeared as a result of solving these supply-cost problems (Takagi 2017).

Regarding the domain note, the oldest on record is a case in the Fukuyama domain, now Hiroshima prefecture, in 1630. Merchants were in charge of issuance operations in many cases of domain notes, and their own shogunate currency secured its credibility as a reserve for redemption. What the *daimyo* guaranteed as passable differed from private notes.

Domain notes were issued for similar reasons to private notes – due to a shortage of silver currency and the saving of transaction costs, such as weighing. More directly, these were issued to fiscally compensate the *daimyo*, who demanded the shogunate currency to import merchandise from outside the domain. Thus, the *daimyo* collected the shogunate's currency from those in the domain, distributed domain notes to the people in return, and enforced the use of domain notes as a substitution. Therefore, the domain note has a small face value, but its primary objective is not to fulfill the people's demand for smaller currency. This differs from private notes.

Past researchers have negatively evaluated the domain note, as a clash often occurred between price and redemption requests, creating panic due to over-issuances without sufficient reservation. However, recent research posits that few such panic cases occurred, and in many cases, circulation rather improved toward the end of the Tokugawa period (Shikano 2011).

Stable circulation occurred due to the common people's demand for currency. Many paper currencies, including private notes, were issued in western Japan, including Kinai region, where the economy had expanded to become larger than in eastern Japan. Although the domain note was issued to fulfill fiscal demand, it consequently met the common people's demand for a small currency.

While the domain note was publicly valid only in the domain, it was also distributed outside the domain. For example, domain notes from Tsuyama (now Okayama prefecture) were distributed in 1700, not only in the domain, but also in the shogunate's direct-control area in the same province (Sejima 2001). The same phenomena are observed after the eighteenth century, when the shogunate began to restrict the issuance of domain notes (Kato 2011). In other words, the common people's demand for currency determined whether the domain notes could circulate, rather than the government's regulation. As aforementioned, metal currency was integrated into the shogunate's through the seventeenth century, but the domain notes' regionalization remained in various areas.

Regarding the appearance of paper currency in the Tokugawa period, it should be noted that the private note occurred prior to the domain note. As with gold, silver, and *zeni* currencies, this is one phenomenon in which the government adopted a currency created by the private sector. For example, private notes issued by a prominent merchant generally were preferred to domain notes due to the credit for the merchant.

Standard Revision and Re-revision: The Tokugawa Tsunayoshi and Ienobu Regimes

The fifth shogun, Tokugawa Tsunayoshi (r.1680–1709), was a younger brother of Tokugawa Ietsuna. His regime revised the gold and silver currency standard.

The shogunate issued *Genroku-gold and silver* in 1695 (*Genroku* 8th), which were debased from *Keicho-gold and silver*. However, the shogunate defined *Genroku-gold and silver* as equivalent to *Keicho-gold and silver*; for example, one *ryo* of *Keicho-gold* is equivalent to one *ryo* of *Genroku-gold*. In other words, the shogunate mandated that these currencies pass at the designated face value – or the

nominal monetization of gold and silver currencies – regardless of their material composition.

This standard revision primarily aimed to: (1) increase issuance, due to exportation and a decrease in domestic gold and silver production, as well as the currency demand for transactions, as previously described, and (2) obtaining revenue from the seigniorage as a result of a fiscal deficit from the decrease of mining income after the reduction of gold and silver.

Tokugawa Tsunayoshi's regime then proclaimed the recovery of *Keicho-gold and silver* and silver bullion including domain silver and suspended the passing of *Keicho-gold and silver*. The shogunate's strengthening monopoly on gold and silver bullion had progressed from the prior regime; the exclusion of domain currency and the shogunate's currency integration were publicly completed. Old, recovered gold and silver currencies were reused as materials for *Genroku-gold and silver*. The materials to create *Genroku-gold and silver* depended heavily on old recovered currencies due to outflow from Japan and a decrease in gold and silver production.

However, the recovery and suspension of the passing of *Keicho*-gold and silver was difficult, as compositions of *Keicho-gold and silver* were superior to *Genroku-gold and silver*; thus, people hoarded the former. Additionally, both old and new currencies were exchanged by the currency changers in Edo, who were designated by the shogunate. Therefore, the currency switching was difficult, except in Edo and locations that transacted frequently with Edo (Katsumata 2004).

Consequently, old and new gold currencies continued to be co-distributed in different market prices. Forgings also became frequent, as the new gold currency could be comprised of less gold than the old one (Yasukuni 2016). The recovery of old gold and silver for re-coinage became a significant political issue, even in revisions by subsequent regimes.

The *Genroku-gold's* decreased purity from the old currency was higher than that of *Genroku-silver*. Thus, *Genroku-gold* was recused in society. After the revision, the market depreciated the *ryo* against the *monme* from their official rates (1 *ryo* = 50 *monme*).

The *ryo's* depreciation against the *monme* was disadvantageous to residents in Edo, where the *ryo* was used as the unit of account and depended on Kamigata where the *monme* was used as the unit of account and supplied commodities to Edo. Thus, the shogunate revised the official rate in 1700, as follows: 1 *ryo* = 60 *monme* = 4000 *mon*. In other words, silver depreciated. Additionally, silver currency was re-debased in 1706 (*Hoei* 3rd), causing the *monme* to further depreciate (*Hoei-silver*). The issuance of this *Hoei-silver* marked the beginning of a currency standard revision to lead the market rate to what the shogunate desired (Iwahashi 2002).

However, the *monme's* appreciation was not suppressed due to the Kamigata economy's growth in a *monme* unit economic zone, and the activation of transactions in Kamigata offset the increase in silver currencies.

Although the gold and silver currency supply increased by the 1695 revision, no additional *zeni* were supplied at the time. Thus, the supply of *zeni* remained

relatively stable in Edo and could not fulfill the demand for *zeni*. Thus, the *mon* appreciated against the *ryo* (Yoshihara 2003).

Interests regarding the *zeni* market differed among residents in Edo, as the city's wholesalers purchased commodities from Kamigata, and sold them in Edo to receive *zeni*. Thus, Edo wholesalers preferred the *ryo's* appreciation against the *monme* and *mon*. The *samurai* warrior class received its stipend in rice, exchanged this rice for the *ryo* system units' gold currency, exchanged the gold currency for *zeni*, and purchased commodities with the *zeni*. Thus, the warriors preferred rice and the *ryo's* appreciation against the *mon*. Common people received their income in *zeni*; thus, they recused the *ryo's* appreciation against the *mon*. The *mon's* appreciation increased the *zeni's* purchasing power, but the *zeni* shortage, which caused its appreciation, decreased income opportunities for ultra-small, self-employed individuals and daily workers, a majority of the common people. The *mon's* appreciation was undesirable for common people who relied on pawnbrokers to increase their actual liabilities.

The shogunate represented the interests of the warrior class. Thus, the shogunate produced *zeni* after 1697 to depreciate the *mon*. This was peak production for the one *mon* bronze *Kan'ei-tsuho* in the Tokugawa period.

The *Kan'ei-tsuho* produced in this time were smaller and lighter than those that preceded, but was accepted in society due to a shortage of *zeni* and a subsequent increasing demand (Yasukuni 1997).

Tokugawa Tsunayoshi's regime then prohibited the passing of paper currency in 1707 to avoid friction among the domains as a result of circulating domain notes issued in other domains (Yasukuni 2016). Consequently, all small currencies decreased, the demand for *zeni* increased, and the *mon* appreciated. This was an undesired side effect for the shogunate.

The Tokugawa Tsunayoshi regime's monetary policy is historically significant due to (1) the exclusion of domain and private-issued currencies (gold, silver, and paper currencies), and their integration by the shogunate currency; that is, the completion of the triple standard system; (2) the debasing and nominal monetization of the shogunate's gold and silver currencies and their acceptance by society; and (3) the appearance of repeated phenomena throughout the Tokugawa period; that is, revisions depending on old currency as materials and the difficulty of exchanging old and new currencies.

After Tokugawa Tsunayoshi's demise, his nephew Tokugawa Ienobu was named the sixth shogun (r.1709–1712) and revised the currency standard.

This shogunate proclaimed the issuance of *Hoei*-gold in 1710 (*Hoei* 7th), which was debased and lighter than *Genroku-gold*. For example, one piece of *Hoei koban* was insufficient to one *ryo* by mass, but the face value was defined as "one *ryo*." As a result, the *ryo* unit became merely a superficial monetary unit, regardless of its mass of contained gold (Yasukuni 1997).

Tokugawa Ietsuna's regime debased the *monme* system unit's silver currencies three times from 1710 to 1711. In each case, there was a higher composition of copper than silver; specifically, the nominal monetization further progressed in the same way as the *ryo* system unit's gold currency. The objectives of standard

revisions were to acquire seigniorage and increase the silver currency supply. This led to a depreciation of the *monme* and declining prices, as with the policies of Tokugawa Tsunayoshi's regime (Iwahashi 2002).

Deflation Policy by Arai Hakuseki: The Tokugawa Ietsugu Regime

In the regime of Tokugawa Ietsugu, a son of Tokugawa Ienobu and the seventh shogun (r.1713–1716), monetary policy was led by Arai Hakuseki, a Confucian, and political consultant of shogun. Arai reversed Tokugawa Tsunayoshi and Ienobu's policies. Further, Tokugawa Ietsugu's regime returned the gold and silver currency standards to those of the Tokugawa Ieyasu regime with *Shotoku-gold and silver* in 1714 (*Shotoku* 4th), that is, upgraded as *Keicho-gold and silver*.

An objective of this revision was to create deflation. Arai attempted to decrease prices by recovering bad currencies and supplying good currencies and reducing the stock of currency. This idea closely approximates the classical quantity theory of money. Another objective was justified by outside economics, or the thought of superiority against foreign countries. Arai thought that the quality of currency symbolized the nation's governance and that debasing the currency would decrease national prestige to foreign countries (Yasukuni 2016).

The currency supply decreased due to this revision and the shogunate's curtailing of expenditures, which occurred as planned for the shogunate. However, rice prices declined. As aforementioned, the *samurai* class received rice as a stipend, exchanged rice for metal currency, and purchased commodities using this metal currency. Thus, rice depreciation was disadvantageous for these warriors. It also reduced farmers' income. As a result, the warriors' and farmers' purchasing power declined, and the economy deteriorated. This was the first depression caused by policy inducement in the history of Japan (Sugiyama 2012).

Shifting Policy to Increase Currency Supply: The Tokugawa Yoshimune Regime

Tokugawa Ietsugu had no children, and after his demise, Tokugawa Yoshimune, from another house in the Tokugawa clan, was named the eighth shogun (r.1716–1745).

Initially, Tokugawa Yoshimune continued Arai's monetary policy and then increased taxation (rice render). The collected rice was exchanged for metal currency in the market, thus increasing the rice supply. These policies forced rice depreciation.

Moreover, the *monme* appreciated. The Tokugawa Tsunayoshi and Ienobu regimes debased gold and silver currencies to depreciate the *monme*, but Arai upgraded the currency to appreciate against high prices. The Tokugawa Yoshimune regime inherited the upgraded currency, and thus, the price of *monme* rebounded.

As aforementioned, the *monme's* appreciation was disadvantageous to Edo residents. Additionally, and although nominal prices declined, the rate of decline for

many goods was smaller than that of rice due to the oversupply of rice. Thus, the real price of goods remained high for Edo residents, except for rice (Sugiyama 2012). As a result, a depression disadvantageous to the warrior class continued, as in Arai's era.

Therefore, Tokugawa Yoshimune shifted monetary policy to increase the supply of currency, leading to rice appreciation. The shogunate lifted its ban on passing domain notes in 1730, and the shogunate issued *Genbun-gold and silver* in 1736 (*Genbun* 1st), which were debased from previous currencies.

Consequently, the *monme* began to depreciate, the prices of rice and general goods increased, and the shogunate's fiscal condition became positive. The currency supply also increased the common people's purchasing power, and deflation was overcome (Sugiyama 2012; Hamano et al. 2017).

At the end of the 1740s, the market exchange rate was stable as an official rate, which lasted until the end of the eighteenth century. Goods and rice prices also began to decline or remained stable until the 1810s, except for times of famine (Shinbo 1978). The prices' stability or decline, despite a currency stock increase, implies that, according to the quantity theory of money, transaction increases offset currency supply increases, and the economy grew substantially (Sugiyama 2012).

Tokugawa Yoshimune's regime attempted to increase the quantity of *zeni* to deter the *mon's* appreciation. However, this was problematic, as the production cost for one-*mon* bronze *Kan'ei-tsuho* became equivalent to its face value due to an increase in copper prices with the decrease of domestic copper production. The material of *zeni*, or copper, was in demand for exportation (Imai 2015).

Thus, the Tokugawa Yoshimune regime issued one-*mon Kan'ei-tsuho* in iron, the first iron currency issued by the government in the history of Japan. The one-*mon zeni* subsequently issued by the shogunate was made of iron, with few exceptions. The *mon* began to depreciate due to the increase in *zeni* supply in the 1740s, and the *zeni* market rate became official.

The Issuance of the *Ryo* System Unit's Silver Currency: Tokugawa Ieharu Regime

Under the Tokugawa Ieshige (the ninth shogun and a son of Tokugawa Yoshimune, r.1745–1760) and Tokugawa Ieharu regimes (the tenth shogun and a son of Tokugawa Ieshige, r.1760–1786), political affairs were led by Tanuma Okitugu, initially a lower-class officer in the shogun's secretariat and eventually the premier *roju* (standing senior councilor). The Tanuma regime attempted fiscal reconstruction by constant revenues, except for the rice render. Tanuma's monetary policy was also presented in the same context.

In 1765 (*Meiwa* 2nd), the shogunate issued the *Meiwa-five-monme-silver*, a rectangular and quantitative currency, with face value of 5 *monme* (5 *monme* = approximately 19 g). Unlike the preceding silver currencies, this was a currency by table. The shogunate anticipated that the *Meiwa-five-monme-silver* would be equivalent to a one-twelfth *ryo* according to the official rate (1 *ryo* = 60 *monme*), irrespective of the market rate. In other words, the shogunate

expected to circulate this as a subsidiary currency of the *ryo* system unit's gold currency. The *monme* at the time had somewhat depreciated from the official rate; thus, the shogunate would profit if the *Meiwa-five-monme-silver* would pass at the official rate.

However, common people would experience a loss if they received *Meiwa-five-monme-silver* at the official rate. Thus, the *Meiwa-five-monme-silver* was recused and was not widely distributed.

As a next measure to acquire seigniorage, the shogunate issued the *Meiwa-two-shu-silver* in 1772 (*Meiwa* 9th), with a face value of one-eighth *ryo,* or two *shu*. This is the first case of the *ryo* system unit's silver currency.

The *Meiwa-two-shu-silver* is historically significant because (1) it met the demand for a small denomination currency supply – as its face value is between one-*mon zeni* and one-*bu*-gold, which was the preceding gold currency's minimum unit – in an age of substantial economic growth, as previously described. (2) As a result of distribution of the *ryo* system unit's silver currency, even those in western Japan adopted the *ryo* unit in their accounting. In other words, the nation's currency unit converged to a *ryo* system unit, which indicates the approach toward the modern gold standard system (Iwahashi 2002).

The Tanuma regime increased its production of the one-*mon Kan'ei-tsuho* as a measure against the *zeni* shortage and *mon* appreciation.

Additionally, the regime issued a *zeni* with a new face value and material: the four-*mon* brass *zeni*, the first *Kan'ei-tsuho* with a face value other than one *mon*. It is assumed that brass was adopted as a material due to references to the Qing Dynasty's *zeni* made of brass. The face value of the four *mon* was equivalent to a one-one thousandth *ryo* in the official rate, and thus, it was convenient for calculation. The four-*mon zeni* met an increasing demand for small-denomination currency, primarily circulated in eastern Japan's *ryo* unit economic zone. However, the four-*mon zeni* could not be distributed as much in western Japan, as paper currencies, which were also small denomination currencies, circulated, preventing the four-*mon zeni's* distribution (Yoshihara 2003).

The Tanuma regime was negative toward paper currency and proclaimed its constraint in new issues. However, domain and private notes continued to be issued in various areas, and especially in western Japan, because transaction demand increased due to economic growth in rural areas. If only the shogunate currency had existed, serious deflation could have occurred due to currency shortages. Various paper currencies also sustained economic growth in the eighteenth century (Kato 2016).

Various Small-Denomination Gold and Silver Currencies: Tokugawa Ienari Regime

In 1786, Tanuma fell from power and Tokugawa Ieharu died. Tokugawa Ieharu had no children, and Tokugawa Ienari – Ieharu's adopted son from another house in the Tokugawa clan – was named the eleventh shogun (r.1787–1837).

Mizuno Tada-akira, a *roju*, led the Tokugawa Ienari regime's monetary policies. The regime issued various new currencies during 1818 (*Bunsei* 1st) and 1832 (*Tenpo* 3rd); in issuance order: the *Bunsei-two-bu-gold* (=the first currency with a face value of two *bu* = one-half *ryo*); the *Bunsei-gold*, or a *ryo* system unit's standard gold currency; *Bunsei-silver*, or a *monme* system unit's standard silver currency by weight; and the *Bunsei-two-shu-silver*, *Bunsei-one-shu-gold* (the first gold currency with a face value of one *shu* = one-sixteenth *ryo*), *Bunsei-one-shu-silver* (the first *ryo* system unit silver currency with a face value of one *shu*), and *Tempo-two-shu-gold*.

All of these *ryo* system unit gold currencies, *monme* system unit silver currencies, and *ryo* system unit silver currencies were debased from those preceding with the same face value, or those with lower quality than the preceding gold and silver currencies compared with the same face value. These issuances aimed to acquire seigniorage, and through this series of currency revisions, the ratio of seigniorage in the shogunate's permanent financial sources increased (Yasukuni 1997).

Former research has negatively evaluated Mizuno Tada-akira's monetary policy, which promoted high prices due to an increased money supply. However, recent research positively evaluates this policy as one that promoted economic growth. As the denominations of new currencies were less than one *bu*, which is the minimum unit of the preceding *ryo* system unit's gold currency, that is, relatively small, the opportunity increased to use such currency for common people's small transactions. This also effectively saved *zeni*, a small denomination currency. The diversification of currencies' face values increased transaction convenience and generally met the common people's increased currency demands (Iwahashi 2002).

These currency supplies developed the common people's economy, and prices increased in the long-term after these currency revisions toward the end of the Tokugawa period. In addition to the paper currencies in various regions, the total stock of such currency increased, which promoted an increase in prices (Shinbo 1978). This economic growth was a precursor to the development of the *Kasei* culture, which was represented by *kabuki* and *ukiyo-e* cultural art at this time.

After Mizuno Tada-akira's demise in 1834, Tokugawa Ienari's regime issued a new type of bronze *zeni* in 1835 (*Tenpo* 6th): the *Tenpo-tsuho*. This had a face value of 100 *mon*. The shogunate debased the standard gold and silver currencies (*Tenpo-gold and silver*) in 1837 (*Tenpo* 7th). Additionally, the shogunate issued its *Tenpo-one-bu-silver*, the first *ryo* system unit silver currency with a face value of one *bu*.

These monetary policies primarily aimed to acquire fiscal revenue by seigniorage. Despite a poor harvest in the 1830s, the seigniorage compensated for a decrease in rice render. In other words, a dependence on rice render in the shogunate's finances declined, while dependence on seigniorage strengthened (Sugiyama 2012).

Regarding paper currencies, these were restricted from passage by Tokugawa Ienari's regime, as well as other regimes in the eighteenth century. However, the policy was ignored, and various domain and private notes were issued in a range of areas. The first half and the third quarter of the nineteenth century marked a peak of paper currency issuances, both domain and private, in the Tokugawa period. This active issuance occurred due to a small currency shortage. Additionally, paper

currencies met the demand for a currency with a face value of less than the new small denomination currencies issued by the shogunate during the 1810s and 1830s.

Daimyos at the time issued domain notes for not only fiscal compensation, but also as a measure of the promotion of industry. A government monopoly policy, the *daimyo* government purchased special products in their domain by domain note and sold them exclusively outside the domain to obtain the shogunate currency (Sugiyama 2012).

Progress of Currency Integration by *Ryo* Unit: Tokugawa Ieyoshi Regime

After Tokugawa Ienari's demise in 1841, his son, Tokugawa Ieyoshi, was the twelfth shogun (r. 1837–1853), while Mizuno Tadakuni, a *roju*, led the monetary policy.

The shogunate suspended the production of *Tenpo-tsuho* to depress prices. Moreover, the regime revised the official rate as 1 *ryo* = 6500 *mon*. This appreciated the *mon* from the market rate at the time, which indicated 1 *ryo* was approximately 6800 or 6900 *mon*. Further, the regime suspended the passing of various gold and silver currencies issued under the Tokugawa Ienari regime, except for *Tenpo-two-shu-gold* and *Tenpo-one-bu-silver*, and suspended the production of all gold and silver currencies. These policies aimed to create lower prices by reducing currency supply.

Consequently, small-denomination *ryo* system unit currencies as the *Tenpo-two-shu-gold* and *Tenpo-one-bu-silver* emerged as the leading shogunate currency (Yasukuni 2016). The *ryo* system unit shogunate currencies were also circulated in the *monme* economic zone, and currency integration through the *ryo* system unit further progressed. This trend continued until the fall of the shogunate. As the Meiji Restoration ended, more than 90% of the circulating currency had face values of less than 1 *ryo* (Iwahashi 2002).

Treaty of Amity and Commerce Between the United States and the Empire of Japan, and Its Influence: The Tokugawa Iesada and Iemochi Regimes

Tokugawa Ieyoshi died in 1853, and his son, Tokugawa Iesada, was named the thirteenth shogun (r. 1853–1858). In 1858, the Harris Treaty – or the Treaty of Amity and Commerce between the United States of America and the Empire of Japan – was signed. Tokugawa Iesada died in the same year, and Tokugawa Iemochi, from another house in the Tokugawa clan, was named the fourteenth shogun (r. 1858–1866). Japan opened its ports to trade under this commerce treaty in 1859, and free trade was resumed for the first time since the seventeenth century. This commerce treaty opened Japan to the modern global market, an epoch-making event toward the start of Japan's modern economic history.

We will now discuss the process toward signing the Harris treaty. After the 1854 signing of the Treaty of Kanagawa – a Convention of Peace and Amity between the United States of America and the Empire of Japan – trade negotiations began between the two countries. One issue was the exchange rate between gold and silver currencies.

In its trade with Japan, the United States demanded to use the Mexican silver dollar, which was its trade currency in the Asian market. The shogunate considered the Mexican silver dollar as silver bullion, analyzed the material value of one Mexican dollar (approximately 24 g of silver), and evaluated this against the *Tenpo-silver* 16 *monme,* according to the shogunate's domestic purchase standard of silver bullion; 16 *monme* equaled approximately one-quarter *ryo,* or one *bu* according to the shogunate's official rate of 1 *ryo* = 60 *monme*. Thus, the shogunate noted that one Mexican dollar should be exchanged for one *Tenpo-one-bu-silver*.

Townsend Harris, the United States Consul General in Japan, was against the offer. He insisted on the principle that foreign coin shall pass for its corresponding weight of Japanese coin of the same description. This principle was adopted among western countries at the time, and according to which, Harris claimed that one Mexican dollar should be exchanged for three pieces of *Tenpo-one-bu-silver* because the silver amount per one *Tenpo-one-bu-silver* is approximately 8.5 g, or approximately 1/3 of a Mexican dollar.

Harris had misinterpreted the Japanese monetary system. *One-bu-silver* was a nominal and subsidiary currency with a face value legally defined by the shogunate as one *bu,* regardless of its bullion value, and was not a standard currency. However, Harris demonstrated no understanding of this system. As a result, the commerce treaty was signed under the conditions claimed by Harris.

Under this system, a foreign merchant would exchange one Mexican dollar in Japan with three *one-bu-silver* and would exchange these with the *ryo* system unit's gold currency (*Tenpo-gold*), or three *bu*. Shanghai's market rate at the time was 1 *ryo* = 2.6 Mexican dollars; thus, three *bu* (3/4 *ryo*) equaled approximately two Mexican dollars (Sugiyama 2012). If the merchant repeated this cycle, he could acquire arbitrage profits. This logic implies that a large amount of gold currencies would flow out of Japan and might cause confusion in Japan's currency system.

This problem occurred because the bullion value of the *ryo* system unit's silver currency in Japan was appreciated against international standards. Silver bullion appreciated in Japan because the shogunate often revised its gold and silver currency standards, while it administrated gold and silver bullion domestic transactions and monopolized the gold and silver trade with foreign countries and then divided the rates between the domestic rate of the *ryo* system unit's gold/*monme* system unit's silver/*ryo* system unit's silver currencies and the international rates of gold/silver bullion. It can be said that the shogunate could revise and enforce to pass *ryo* system unit's silver currencies (a subsidiary currency) in the face value regardless of international gold and silver bullion price fluctuations by the transaction and trade administration of bullion (Iwahashi 2002).

If the usage of foreign currency in Japan and export of Japanese currency were restricted, Japanese gold currency would not flow outward, but the commerce treaty

allowed for such an export. Japan proposed the treaty's export permission term, although economically disadvantageous, as a bargain for issues except currency. The proposal by Japan occurred due to the shogunate's traditional, arrogant diplomatic thought that Japan should trade with foreign countries as a favor. The proposal was also in return for denial of foreigners' free domestic travel and visiting in Kyoto (Yasukuni 2016; Okada 1955).

Naturally, the shogunate did not want the outflow of gold currencies; thus, the shogunate suppressed the supply of *Tenpo-one-bu-silver*, not to be exchanged with *Tenpo-gold*. Harris reacted by proposing the creation of *one-bu-silver* from Mexican silver dollars. The shogunate accepted the proposal and issued its *Ansei-one-bu-silver* in 1859 (*Ansei* 6th), which was produced until the end of the Tokugawa period.

Additionally, Harris suggested a "three-*bu*" countermark on the Mexican dollar, allowing it to pass as equivalent to three pieces of *one-bu-silver*. The shogunate accepted this suggestion. Preceding this, the shogunate had permitted the Mexican dollar's use in domestic tax payment and transactions, as this was the official permission for valid foreign currency since 1670 when the *bita* was suspended.

Man'en Revision and the Fall of the Tokugawa Shogunate

The shogunate revised its standard gold currency in 1860 (*Man'en* 1st) to solve its exchange rate problem (*Man'en-gold*, or *koban* and *one-bu-gold*). This was the shogunate's final revision of its gold standard, while *Man'en-two-bu-gold* and *Man'en-two-shu-gold* were also simultaneously issued.

The rate was designed so that *Man'en-gold* 1 *ryo* = 15 pieces of *Ansei-1-bu-silver*, based on the net mass of gold and silver contained in each currency. By debasing the standard gold currency, in other words, and by appreciating gold bullion against silver bullion, the shogunate adjusted the value of gold to international standards.

A recommendation by Harris spurred the gold currency revision. Industrial capitalist in the United States and Britain wanted to export goods to Japan, but were afraid of the outflow of gold currency from Japan due to the speculative buying that intensified a radical antiforeigner faction of the Japanese, which interfered with trade (Ono 2000). Thus, Harris proposed that gold should be revised so as to deter this outflow.

As aforementioned, *Man'en-gold* was issued due to heteronomous international factors. This differs from preceding currency revisions in the Tokugawa period, which were autonomous, such as acquisition of seigniorage (Shinbo 1978).

The gold currency revision created a rate between the *ryo* system unit's gold and silver currencies' mass based equal to the international rate as bullion. This caused the outflow of gold currencies from Japan to cease.

The amount of gold currency outflows at this time is assumed as less than that estimated by conventional research. The shogunate attempted to prevent the outflow of gold currencies by suppressing the supply of *one-bu-silver*; foreign merchants had to exchange Mexican dollars for *one-bu-silver* – the *ryo* system unit's silver currency

– to obtain the Japanese *ryo* system unit's gold currency. However, an insufficient supply of *one-bu-silver* deterred foreign merchants from obtaining gold currencies. Additionally, Japanese merchants purchased gold currencies before treaty ports opened in anticipation of foreign merchants' speculative buying of gold currencies, and the domestic market price for gold currencies increased. In other words, domestic gold currency appreciation more closely approximated international standards before the issuance of *Man'en-gold*, regardless of the shogunate's policy, but depressed the outflow of gold currencies (Yasukuni 1997).

Further, international factors restrained the outflow of gold currencies. Foreign merchants obtained the Mexican dollar in Chinese market specifically Hong Kong or Shanghai in exchange of gold currencies obtained in Japan, in order to purchase *one-bu-silver* in Japan. As the demand for the Mexican dollar increased, and the supply of gold currencies increased in China, the Mexican dollar became appreciated against gold. Consequently, arbitrage profits decreased (Ono 2000), which suppressed the outflow of gold currencies from Japan.

The revision of the gold currency standard caused prices to increase in Japan. The shogunate then matched the international rates between gold and silver by appreciating the conventional gold currency *(Tenpo-gold)* by approximately 2.7 times preceding the issuance of *Man'en-gold*. However, the domestic goods supply would not increase soon, and Japan became a more active exporter.

As a result, the domestic goods supply decreased, and commodity prices increased. Thus, the antiforeigner faction within Japan, which assumed foreigners were the reason for rising prices, further strengthened. Moreover, the increase in fiscal expenditure accompanying the shogunate's dispatch to the Choshu domain (now Yamaguchi prefecture) that began in 1864, and other political and social confusion at the end of the Tokugawa period, increased prices, and riots and strikes frequently occurred (Iwahashi 2002).

The shogunate was reluctant to produce *zeni* due to its nonprofitability. Meanwhile, bronze *zeni* was smuggled to China after treaty ports opened, despite the commerce treaty's prohibition on exporting *zeni*. This was because bronze *zeni* were appreciated in China from Japan, causing Japan's supply of *zeni* to run short. One theory assumes that the outflow of *zeni* was more serious than that of gold currency immediately after the treaty ports opened (Fujii 2016).

As a countermeasure for this *zeni* shortage, the shogunate issued an iron *Kan'ei-tsuho* in 1860, with a face value of four *mon*. This was the first four-*mon zeni* made of iron, but was soon recused in society. Thus, the shogunate ceased to produce the four-*mon* iron *zeni* and issued bronze *zeni* with a face value of four *mon* (*Bunkyu-eiho*) in 1863 (*Bunkyu* 3rd). This was the last *zeni* issued by the Japanese government. However, the *zeni* shortage continued due to hoarding and an outflow from Japan (Fujii 2016).

As another countermeasure, the shogunate defined its one-*mon* iron *zeni* as standard *zeni* (evaluated to one *mon*) and permitted it to pass under a free rate except for iron *zeni* and *Tenpo-tsuho* in 1865. The currency changers' association then appealed to the shogunate for the following price, which was approved: the 4-*mon* brass *zeni* = 12 *mon*; the *Bunkyu-eiho* = 8 *mon*; and a high-quality 1-*mon* bronze

zeni at its conventional face value = 6 *mon*, while the other 1-*mon* bronze *zeni* = 4 *mon*. The appreciated *zeni* then re-appeared and circulated in the market, creating an oversupply of *zeni* that turned into the *mon's* depreciation and encouraged inflation (Yoshihara 2003).

In 1867, under the regime of Tokugawa Yoshinobu, fifteenth shogun from another house in the Tokugawa clan (r.1866–1867), the shogunate issued a paper currency denominated in *ryo* system units. This is the first paper currency issued by the central government in the history of Japan. The issuance was primarily intended as a payment for trades and a funding for settlement developments. This paper currency did not fulfill its objectives, as the shogunate collapsed shortly after its issuance.

The Transfer of the Monetary System from the Shogunate to the Modern Era: After the Meiji Restoration

In 1868, the Meiji Restoration occurred, and the shogunate collapsed. The new government inherited the shogunate's currency system in the short term after the Restoration.

The new government suspended the passage of the *monme* system unit's silver currency and the *monme* unit's usage as money of account in 1869. This aimed to permit only currency by table to be passable and to integrate units of currency in the *ryo* system (Shikano 2011). This is an extension of the currency unit integration in the *ryo* system that had progressed in the Tokugawa period.

In 1871, the new government introduced a modern gold standard currency system and issued metal currency in yen units. The yen currency inherited the shogunate's gold currency system; the new government defined the old currency of 1 *ryo* as equivalent to 1 yen, and 1 yen is 1.5 g of gold as the same as 1 US dollar (gold). We can describe the identity equation as 1 yen = 1 US dollar ≒ 2 pieces of *Man'en-two-bu-gold*, or the primarily distributed gold currency at the end of the Tokugawa period, which = 1 *ryo* (Yamamoto 1994). Thus, the old currency was easily converted to new, as the existence of *Man'en-two-bu-gold* facilitated the transfer to the modern gold standard system.

Conclusion

The sixteenth century created a path to the triple standard system in the Tokugawa period. The Oda and Toyotomi regimes adopted the gold and silver currencies autonomously created by society and designated the *bita* as a standard *zeni* that was used as the standard *zeni* in social practice.

In the beginning of seventeenth century, Tokugawa Ieyasu, who established the shogunate, also designated the gold and silver currencies and *bita* that were adopted by the Oda and Toyotomi regimes. Additionally, the shogunate issued *Kan'ei-tsuho,* which was equivalent to the *bita*. These facts imply that the shogunate inherited monetary experience in the sixteenth century.

The shogunate's nationwide integration of currency was achieved in the late seventeenth century. However, domain and private notes were also co-distributed even after that, under the priority of the shogunate's currency; in other words, these currencies were loosely integrated.

In the eighteenth century, debasing and nominal monetization policies in the Tokugawa Tsunayoshi and Ienobu regimes were inherited, except for the reaction of Arai Hakuseki. Additionally, the *ryo* system unit's silver currency appeared in latter half of the eighteenth century, after which currency integration in the *ryo* system unit progressed. This implies an approach to the modern gold standard system, which facilitated the transition to a single currency system in yen units after the Meiji Restoration.

During the eighteenth and first half of the nineteenth centuries, small-denomination gold and silver currencies, iron and brass *zeni*, and various paper currencies – such as the domain note – sustained the life of common people and real economic growth. The experience of passing nominal metal and paper currencies facilitated a transfer to the modern monetary system after the Meiji Restoration, including the circulation of government notes and banknotes. In conclusion, the monetary experience during the eighteenth and the first half of the nineteenth centuries was a precursor to the monetary system in modern Japan.

However, this was just one result, as each policy's primary objective in the Tokugawa period was to confront one task at the time: fiscal compensation and securing profits from the warrior class.

Cross-References

- ▶ Experiments with Paper Money
- ▶ Monetary System in Ancient China
- ▶ Rise and Demise of the Global Silver Standard
- ▶ The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy

Acknowledgments This work was supported by JSPS KAKENHI Grant Numbers JP17K03118, JP16H03650, JP17H02389 and JP18H00880.

References

Chieda D (2011a) Chu-kinsei Isejingu chiiki no kahei to shogyososhiki (Money and commercial organization in Ise Shrine district from medieval to early-modern). Iwatashoin, Tokyo

Chieda D (2011b) "Matsusakashiiki wo chushin tosuru Mie-ken no koshihei no rekishi" ("History of old paper currency in Mie prefecture, centering on Matsusaka city"). In: Matsusaka City History and Folklore Museum (ed) Hansatsu to hagaki (Domain note and hagaki). Matsusaka City History and Folklore Museum, Mie, pp 29–32

Fujii J (2014) “Kinsei Kaheiron” (“Discussion of money in the early-modern Japan”). In: Otsu T, Sakurai E, Fujii J, Yoshida Y, Li S (eds) Iwanamikoza Nihonrekishi, 11 (Iwanami course history of Japan 11). Iwanamishoten, Tokyo, pp 213–245
Fujii N (2016) “Bakumatsuki no kaheikyokyu” (“Money supply in the end of the Tokugawa shogunate”). Kin’yu kenkyu (Financ Stud) 35–2:67–118
Hamano K, Ioku S, Nakamura M, Kishida M, Nagae M, Ushijima T (eds) (2017) Nihon keizaishi 1600–2015 (Japanese economic history 1600–2015). Keiogijukudaigaku Shuppankai, Tokyo
Imai N (2015) Kinsei Nihon no do to Osaka do shonin (Copper in the early-modern Japan and copper merchant in Osaka). Shibunkaku shuppan, Kyoto
Iwahashi M (1998) “Edoki kahei seido no dynamism” (“Dynamism of monetary system in the Edo period”). Kin’yu kenkyu (Financ Stud) 17–3:59–80
Iwahashi M (2002) “Kinsei no kahei·shinyo” (“Money and credit in the early-modern”). In: Sakurai E, Nakanishi S (eds) Shin Taikei Nihonshi, 12 (New Japanese history system, 12). Yamakawashuppansha, Tokyo, pp 431–469
Kato K (2011) “Kinsei kindai ikoki ni okeru Banshu Mikicho no tsuka kozo” (“Currency structure of Japan: a case of Miki town”). Ryutsukagakudaigaku ronshu (J Univ Mark Distrib Sci) 24–1:167–189
Kato K (2016) “Nihon kinsei no shisatsu” (“Private notes in early modern Japan”). Shakai keizai shigaku (Socio-Econ Hist) 81–4:25–40
Katsumata T (2004) “Nihon kinsei no kahei ryutsu ni kansuru shiron” (“Tentative assumption on money distribution in early-modern Japan”). Rekishi no riron to kyoiku (Hist Theory Educ) 117:11–22
Kuroda A (2014a) Kahei system no sekaishi (World history of monetary system). Iwanamishoten, Tokyo
Kuroda A (2014b) “Yui suzu shikan” (Historical ‘Tin’ialism). In: Hirao Y, Iinuma K, Murai S (eds) Daikokaijidai no Nihon to kinzoku koeki (Japan in the age of exploration and metal commerce). Shibunkakushuppan, Kyoto, pp 18–20
Nakajima K (1999) “Nihon no chuseikahei to kokka” (“Medieval money in Japan and nation”). In: Rekishigakukenkyukai (ed) Ekkyo suru kahei (Money crossing border). Aokishoten, Tokyo, pp 109–139
Nakano S (2012) Kagahan no ryutsukeizai to jokamachi kanazawa (Distribution economy of Kaga domain and castle city Kanazawa). Notoshuppan Insatsubu, Ishikawa
Okada S (1955) Bakumatsuishin no kahei seisaku (Monetary policy in the end of the Tokugawa shogunate and opening of the Meiji restoration). Moriyama Shoten, Tokyo
Ono K (2000) Kindai Nihon heisei to higashi Asia ginkaken (Modern Japan monetary system and eastern Asia silver currency zone). Minervashobo, Kyoto
Sakurai E (2017) Kokan, kenryoku, bunka (Exchange, power, and culture). Misuzushobo, Tokyo
Sakuraki S (2009) Kaheikokogaku josetsu (An introduction to numismatic archaeology). Keio Daigaku Shuppankai, Tokyo
Sejima H (2001) “Kinsei shoki no hansatsu” (“Domain note in the earlier period of early modern ages”). Nihonshi Kenkyu (J Jpn Hist) 471:28–52
Shikano Y (2011) Hansatsu no keizaigaku (Economics of the domain note). Toyokeizaishinposha, Tokyo
Shinbo H (1978) Kinsei no bukka to keizaihatten (Prices and economy development in early-modern). Toyokeizaishinposha, Tokyo
Sugiyama S (2012) Nihon keizaishi kinsei-kindai (Japanese history of economy in early-modern to modern). Iwanamishoten, Tokyo
Takagi H (2010) Nihon chusei kaheishiron (Discussion on monetary history in medieval Japan). Azekurashobo, Tokyo
Takagi H (2016) Tsuka no Nihonshi (Japanese history of currency). Chuokoronshinsha, Tokyo
Takagi H (2017) Kinsei no kaimaku to kahei togo (The opening of early-modern Japan and monetary integration). Shibunkakushuppan, Kyoto
Takagi H (2018a) Recent studies of bronze coin integration at the beginning of early modern Japan. Yasudajoshidaigaku kiyo (J Yasuda Womens Univ) 45:11–18

Takagi H (2018b) (in printing) Erizeni to bita ichi mon no sengokushi (Coin discrimination and bita in the warring state period). Heibonsha, Tokyo
Watanabe N (2002) Watanabe Nobuo rekishi ronshu 1 (Watanabe Nobuo's academic papers collection on History 1). Seibundoshuppan, Osaka
Yamamoto Y (1994) Ryo kara yen e (From ryo to yen). Minervashobo, Kyoto
Yasukuni R (1997) "Kaheishi ni okeru kinsei" (Early-modern in monetary history). In: Nagai K (ed) Kinsei no shutsudosen (Excavated coin in early-modern). Hyogo maizosen chosakai, Hyogo, pp 3–34
Yasukuni R (2016) Nihon kinsei kaheishi no kenkyu (Study of the monetary history of early-modern Japan). Shibunkakushuppan, Kyoto
Yoshihara K (2003) Edo no zeni to shomin no kurashi (Zeni in Edo and day-to-day life of common people). Doseisha, Tokyo

Part VII

Exchange Rate Regimes

International Monetary Regimes: The Gold Standard

24

Lawrence H. Officer

Contents

L. H. Officer (✉)
Department of Economics, College of Liberal Arts and Sciences, University of Illinois at Chicago, Chicago, IL, USA
e-mail: lofficer@uic.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_23

Abstract

The classical gold standard is the most famous monetary system that ever existed, with its heyday lasting a third of a century. By the time World War I began, the gold standard had become the predominant national and international monetary system in the world. Countries may be allocated to different groups, depending on the importance of the country to the working of the gold standard, the type of gold standard to which the country adhered, and the extent to which the country observed the standard. Whether automatic or policy-induced, there are implications for the money supply. The main theme is that the gold standard exhibited both elements that promoted stability and forces that fostered instability. Modern time-series analysis has been used to examine various facets of the gold standard, especially the roles of the core countries (Britain, France, Germany, and the United States). While there is apparent consensus on some aspects of the gold standard, controversies continue, and there remains room for further research and reflection.

Keywords

Gold · Core · Periphery · Standard · Rules

Introduction

The classical gold standard is the most famous monetary system that ever existed, with its heyday lasting a third of a century. By the time World War I began, the gold standard had become the predominant national and international monetary system in the world. Countries may be allocated to different groups, depending on the importance of the country to the working of the gold standard, the type of gold standard to which the country adhered, and the extent to which the country observed the standard. Whether automatic or policy induced, there are implications for the money supply. The main theme is that the gold standard exhibited both elements that promoted stability and forces that fostered instability. Modern time-series analysis has been used to examine various facets of the gold standard, especially the roles of the core countries (Britain, France, Germany, and the United States). While there is apparent consensus on some aspects of the gold standard, controversies continue, and there remains room for further research and reflection.

Countries on Gold Standard

Legal Versus Effective Monetary Standard

Countries effectively on the gold standard and the periods during which they were on gold are listed in Table 1. The effective monetary standard of a country is distinguished from its legal standard. For example, a country legally on bimetallism usually was on either an effective gold or effective silver monometallic standard, depending on whether the country's "mint-price ratio" (the ratio of its mint price of gold to mint price of silver) was greater or less than the world price ratio. In contrast, a country might be legally on a gold standard, but its banks (and government) have "suspended specie (gold) payments" (refusing to convert their notes into gold), so that the country is in fact on a "paper standard."

Table 1 strives to incorporate all time periods (and only time periods) when a country was on an operational, or "effective," gold standard, irrespective of the legal standard; but in some cases only beginning and ending dates on gold can be discerned. The criterion adopted is that a country is deemed on the gold standard if (1) gold was the predominant effective metallic money, (2) specie payments were in force, and (3) there was a limitation on the coinage and/or the legal-tender status of silver (the only practical and historical competitor to gold), thus providing institutional or legal support for the effective gold standard emanating from (1) and (2).

The years 1880–1913 are generally construed as "the heyday of the gold standard," because throughout this period the "core countries" (Britain, France, Germany, United States), along with Scandinavia and several Western European countries, were continuously on gold. To quote Flandreau et al. (1998), p. 150): "The big players were on gold and this is why economic history, rightly, puts the dates 1880–1913 on the gold standard."

In 1870 only Britain, two of its dependencies (Australia, Canada) and two countries also closely aligned with Britain economically and politically (Argentina and Portugal) were on the gold standard. Of all other countries, only the United States had ever been on an effective gold standard. By 1900, and even more so by 1914 (ironically, just before the gold standard collapsed, with World War I), almost every economically important country in the world had adopted gold. How did gold monometallism achieve its primacy? And, in particular, what explains the "scramble for gold" (or "rush to gold") that began in the 1870s?

Center Country

The country grouping in Table 1 reflects the importance of countries to establishment and maintenance of the standard. Consider a "core country" as a country of high importance to that end. Then Britain was the "center country," and thus the most important core country. It was the earliest country on a gold standard and was

Table 1 Countries on gold standard

Country	Type of standard	Period
Core		
Center country		
Britain	Coin	1774–1797, 1821–1914
Other core countries		
United States	Coin	1834–1861, 1879–1917
France	Coin	1878–1914
Germany	Coin	1871–1914
Inner periphery		
British colonies and dominions		
Australia	Coin	1852–1915
Canada	Coin	1854–1914
Ceylon	Coin	1901–1914
India	Exchange (British pound)	1898–1914
Western Europe		
Austria-Hungary	Coin	1892–1914
Belgium	Coin	1878–1914
Italy	Coin	1884–1894
Liechtenstein	Coin	1898–1914
Netherlands	Coin	1875–1914
Portugal	Coin	1854–1891
Switzerland	Coin	1878–1914
Scandinavia		
Denmark	Coin	1872–1914
Finland	Coin	1877–1914
Norway	Coin	1875–1914
Sweden	Coin	1873–1914
Outer periphery		
Eastern Europe		
Bulgaria	Coin	1906–1912
Greece	Coin	1885, 1910–1914
Montenegro	Coin	1911–1914
Romania	Coin	1890–1914
Russia	Coin	1897–1914
Middle East		
Egypt	Coin	1885–1914
Turkey (Ottoman Empire)	Coin	1881–1914
Asia		
Japan	Coin	1897–1917
Philippines	Exchange (US dollar)	1903–1914
Siam	Exchange (British pound)	1908–1914
Straits settlements	Exchange (British pound)	1906–1914

(*continued*)

Table 1 (continued)

Country	Type of standard	Period
Mexico and Central America		
Costa Rica	Coin	1896–1914
Mexico	Coin	1905–1913
South America		
Argentina	Coin	1867–1876, 1883–1885, 1900–1914
Bolivia	Coin	1908–1914
Brazil	Coin	1888–1889, 1906–1914
Chile	Coin	1895–1898
Ecuador	Coin	1898–1914
Peru	Coin	1901–1914
Uruguay	Coin	1876–1914
Africa		
Eritrea	Exchange (Italian lira)	1890–1914
German East Africa	Exchange (German mark)	1885–1914
Italian Somaliland	Exchange (Italian lira)	1889–1914

Sources: Bulgaria – Dimitrova and Fantacci (2010, p. 190, 194). Korea and Taiwan – Conant (1915, p. 566–568). Other countries – Officer (2008, Table 1)

Britain includes colonies (except British Honduras) and possessions without a national currency: New Zealand and certain other Oceanic colonies, South Africa, Guernsey, Jersey, Malta, Gibralter, Cyprus, Bermuda, British West Indies, British Guiana, British Somaliland, Falkland Islands, other South and West African colonies. Britain first limited legal tender of silver in 1774, terminated free coinage of silver in 1798

For precise dates and internal geographic exceptions for US period on gold standard, see Officer (1996, p. 16–17). United States includes countries and territories with US dollar as exclusive or predominant currency: British Honduras (from 1894), Cuba (from 1898), Dominican Republic (from 1901), Panama (from 1904), Puerto Rico (from 1900), Alaska, Aleutian Islands, Hawaii, Midway Islands (from 1898), Wake Island, Guam, and American Samoa

France includes Tunisia (from 1891) and all other colonies except Indochina

Canada includes Newfoundland (from 1895). India includes British East Africa, Uganda, Zanzibar, Mauritius, and Ceylon (to 1901). Austria-Hungary includes Montenegro (to 1911). Belgium includes Belgian Congo. Netherlands includes Netherlands East Indies. Portugal includes colonies, except Portuguese India. Denmark includes Greenland and Iceland. Japan includes Korea and Taiwan (both from 1904). Straits Settlements includes Borneo. German East Africa and Italian Somaliland beginning dates are approximate

For other gold-standard lists, see Bordo and Schwartz (1996, pp. 20–22), Meissner (2005, p. 391), Martin-Aceña (2007, pp. 97–100), Mitchener and Weidenmier (2015, pp. 486, 508)

indispensable to the spread and functioning of the gold standard. “London was the center for the world’s principal gold, commodities and capital markets. . .[There were] extensive outstanding sterling-denominated assets, and. . .many countries substituted sterling for gold as an international reserve currency” (Bordo 1993, p. 162).

For centuries, Britain had been on an effective silver standard under legal bimetallism. The country switched to an effective gold standard early in the eighteenth century, solidified by the (mistakenly) gold-overvalued mint-price ratio

established by Isaac Newton, Master of the Mint, in 1717. In 1774 the legal-tender property of silver was restricted, and Britain entered the gold standard in the full sense. In 1798 coining of silver was suspended, and in 1816 the gold standard was formally adopted, ironically during a paper-standard regime (the "Bank Restriction Period," of 1797–1821), with the gold standard effectively resuming in 1821 and remaining until 1914.

Other Core Countries

Lindert (1969) identifies the pound sterling, French franc, and German mark as "key currencies," the most important reserve currencies. "The role of world banker was performed by Britain, France, and Germany in these years [1900–1913] on a scale unmatched either before or since" (p. 1). Flandreau and Jobst (2005), using the criterion of international circulation of domestic currencies (measured by number of recorded geographic exchange-market quotations), also place these three currencies in the top tier. The core countries Germany and France switched from bimetallism and silver to gold in 1871 and 1878, respectively.

It is controversial whether the United States should also be considered a core country. If a large circulation of gold coin is the criterion (Gallarotti 1995, p. 23), then the United States belongs in the group. If a central bank is required, then the United States does not so belong – a judgment also reached according to the Flandreau-Jobst criterion. However, tipping the scales in favor of inclusion is the fact that the United States was a heavyweight in the world economy, with large shares of would output, trade, and investment. Tullio and Wolters (2000, p. 62) state bluntly: "by 1910 US real GDP was three times UK GDP." Indeed, most scholars show revealed preference for inclusion, because their "heyday of the gold standard" begins only after the United States returned to the gold standard in 1879, thus completing the core group.

The United States was on an effective silver standard dating back to colonial times, legally bimetallic from 1786, and on an effective gold standard from 1834. The legal gold standard began in 1873–1874, when Congressional Acts ended silver-dollar coinage and limited legal tender of existing silver coins. Ironically and again, the move from formal bimetallism to a legal gold standard occurred during a paper standard (the "greenback period," of 1861–1878), with a dual legal and effective gold standard from 1879.

Periphery

The core countries attracted other countries to adopt the gold standard, in particular, British colonies and Dominions, Western European countries, and Scandinavia. These noncore countries were generally closely aligned with one or more core countries and could be viewed as constituting the "inner periphery." The "rush to the gold standard" began in the 1870s, with the adherence of Germany, France,

Scandinavia, and other European countries. Legal bimetallism shifted from effective silver to effective gold monometallism around 1850, as gold discoveries in the United States and Australia resulted in overvalued gold at the mints. With silver discoveries in Nevada, the gold/silver market situation subsequently reversed itself, and to avoid a huge inflow of silver and stem an outflow of gold, many European countries suspended the coinage of silver and limited its legal-tender property. Some countries (France, Belgium, Switzerland – three founding members of the Latin Monetary Union) adopted a "limping" gold standard, in which existing former-standard silver coin retained full legal tender, permitting the monetary authority to redeem its notes in silver as well as gold.

So, while all noncore countries were in the broadly defined periphery, there is a narrower periphery: Eastern Europe, Middle East, Asia, some colonial Africa, and Latin America. These countries – including, for some purposes, also British colonies and Dominions – were in the "outer" periphery: acted on, rather than actors, in the gold standard, and generally not as committed to the gold standard. Some countries – China, Persia, parts of Latin America – never joined the classical gold standard, instead retaining their silver or bimetallic standards.

Flandreau and Jobst have a different division of noncore countries. The periphery consists of countries the currency of which has exchange-market representation only at home and possibly in one neighboring country: Dominions and colonies, Southeastern Europe, Latin America, and Asia. The periphery could also be defined as the set of countries which could not circulate abroad debt denominated in their own currency (Morys 2013, pp. 206–207).

For Flandreau and Jobst, an intermediate group consists of countries the currencies of which enjoy regional exchange-market quotations: the United States and various European countries. I find questionable the characterization of the United States as noncore. It was simply a matter of historical tradition that "the reach of the dollar-sterling exchange market extended beyond to encompass almost the entirety of American economic transactions. . ..All the while, the balance-of-payments strength of the United States was growing, and along with it resentment of foreign-exchange dependence on London, which financed US trade even with third parties" (Officer 1996, p. 61, 63). Throughout the gold-standard heyday, the United States became more and more economically powerful, and more and more important to the international gold standard – making characterization of the country as noncore incongruous.

Why the Scramble to Gold?

The idea of a "scramble" or "rush" to gold has also been named the "monetary chain gang" (Gallarotti 1995). There was a sequential movement to gold driven by network externalities in the form of trade and investment. Dependencies, sovereign, and nonsovereign, following Britain to the gold standard, are mentioned above. Similarly, German economic satellites (Netherlands, Scandinavia) followed Germany and French satellites (Switzerland, Belgium, Italy) followed France.

Eichengreen and Flandreau (1996) extend this thread to India and Straits Settlements, Dutch East Indies, Korea (and, logically, Taiwan), and Philippines, following Britain, Netherlands, Japan, and United States, respectively.

The role of the fall in the price of silver in the switch to gold is subject to amendment. Perhaps it was the desire to stabilize commodity prices that was the impetus for the switch (Gallarotti 1995). And it is arguable that the fall in the price of silver relative to gold was determined by shifts in demand rather than supply (Milward 1996), whence an endogenous phenomenon. These issues warrant further attention by historians and cliometricians.

Conventional scholarly wisdom is that war indemnity helps to explain some adoptions of the gold standard (e.g., France and Germany following the Franco-Prussian War, Japan after victory over China in 1895). Industrialization is also said to play a role, with its high value/weight ratio making gold the better metal than silver for transactions large in size and volume. Also, the gold standard had a "Good Housekeeping Seal of Approval" for the inflow of long-term capital (Bordo and Rockoff 1996, of which more below). Countries with fluctuating exchange rates might have been attracted to the stability of gold. There are also political theories of gold-standard adoption. Ideologically, there was the desire to follow the monetary standard of Britain and Germany, the leading economic powers. Domestically, there was the rise of urban-capitalist and industrial over agricultural interest groups, the former favoring gold for its low inflation (Gallarotti 1995), and the perennial conflict between creditors and debtors, again the former supporting gold for its purportedly deflationary power.

Empirically unscrambling the many theories of the scramble for gold is a difficult task. Meissner (2005) adopts an approach that warrants attention and extension. Using an econometric "duration model," his determined variable is the number of years (after 1870) until a country adopts the gold standard. The strongest result is that "a country would be more likely to move to gold the more it traded with other gold standard countries" (Meissner 2005, p. 400). Also "A higher gold cover ratio [gold reserves/notes outstanding] is associated with earlier adoption times" (p. 395). So is a higher spread between domestic bond yield and British consol rate, in line with the "Good Housekeeping" hypothesis. In general, "the order in which countries adopted depended on trade patterns, financial needs, and structural constraints" (p. 401).

Characteristics of Gold Standard

Domestic Gold Standard

Coin Standard

Most gold-standard countries were on a coin standard (see Table 1). From a domestic standpoint, the coin standard had four properties. First, there was a well-defined and fixed gold content of the domestic monetary unit. For example, the dollar was defined as a specified weight of pure gold. Second, gold coin circulated as money with unlimited legal-tender power (meaning it is compulsorily acceptable means of

payment of any amount in any debt transaction or obligation). Third, privately owned bullion (gold in mass, foreign coin considered as mass, or gold in the form of bars) was convertible into gold coin in unlimited amounts at the government mint or at the central bank (if one existed), and at the "mint price" (of gold, the inverse of the gold content of the monetary unit). Fourth, private parties had no restriction on their holding or use of gold (except possibly that privately created coined money could be prohibited); in particular, they may melt coin into bullion. The effect is as if coin were sold to the monetary authority (central bank or Treasury acting as a central bank) for bullion. It sometimes made sense for the authority to sell gold bars directly for coin, even though not legally required, thus saving cost of coining. The third and fourth properties in effect committed the monetary authority to transact in coin and bullion in both directions such that the mint price, or gold content of the monetary unit approximately (because of transactions costs) governed in the marketplace.

However, even under a coin standard, gold was not the only money. Rather than a "pure" coin standard, the norm was a "mixed" coin standard, with both gold coin and other money circulating. In fact, a pure coin standard did not exist in any country during the gold-standard period. There was nongold coin and also paper currency (notes) – issued by the government, central bank, or commercial banks – and demand-deposit liabilities of banks. Generally, except for a "limping" gold standard (see above), nongold (in particular, silver) coin was not officially convertible into gold and had only "token" status, meaning limited legal-tender power and face value exceeding metallic value. In contrast, government or central-bank notes and central-bank deposit liabilities were directly convertible into gold coin at the fixed established price on demand. Commercial-bank notes and demand deposits might be converted not directly into gold but rather into gold-convertible government or central-bank currency. This indirect convertibility of commercial-bank liabilities would apply certainly if the government or central-bank currency were legal tender, but also generally even if it were not.

As legal tender, gold coin was always exchangeable for paper currency or deposits at the mint price, and usually the monetary authority would provide gold bars for its coin. Again, two-way transactions in unlimited amount fixed the currency price of gold at (or approximately at) the mint price. The credibility of the monetary-authority commitment to a fixed price of gold is the essence of a successful, ongoing gold-standard regime.

Over time, gold coin declined from about 1/5 of the world money supply in 1800 (2/3 for gold and silver coin together, as silver was then the predominant monetary standard) to 17% in 1885 (1/3 for gold and silver, for an eleven-major-country aggregate), and 10% in 1913 (15% for gold and silver, for the major-country aggregate (Triffin 1964, p. 15, 56, and see Table 2). The main use of gold coin became not circulating medium but rather reserves for Treasuries, central banks, and (generally to a lesser extent) commercial banks.

Gold-Exchange Standard

As shown in Table 1, some countries in the periphery were on a gold-exchange standard, in which the monetary authority buys and sells not gold in any form but

Table 2 Structure of money: major-countries aggregate (end of year)

	1885	1913
Money supply ($ billion)	8	26
Ratio of metallic money to money supply (percent)	33	15
Ratio of official reserves to money supply (percent)	18	16
Ratio of official to official-plus-money gold (percent)	33	54

Source: Triffin (1964, p. 62)

Major countries are core (Britain, United States, France, Germany), Western Europe (Belgium, Italy, Netherlands, Sweden, Switzerland), Canada, and Japan. Money supply consists of metallic money, minor coin, paper currency, and demand deposits. Metallic money in 1885 is gold and silver coin; an overestimate, as includes commercial-bank holdings that could not be isolated from coin held outside banks by the public. Metallic money in 1913 is gold and silver coin. Official reserves are gold, silver, and foreign exchange. Official gold is gold in official reserves. Money gold is the gold-coin component of money supply

rather gold-convertible foreign exchange, that is, the currency of a country that itself is on the coin standard. Countries on a gold-exchange standard usually were colonies or territories of a country on a coin standard. In situations in which the periphery country lacked its own (even-coined) currency, the gold-exchange standard existed almost by default.

International Gold Standard

Properties

An "international" gold standard requires, in addition to the domestic properties, freedom both of international gold flows (private parties permitted to import or export gold without restriction) and of foreign-exchange transactions (an absence of exchange control). Then the fixed mint prices of any two countries on the gold standard imply a fixed exchange rate ("mint parity") between the countries' currencies. For example, the US mint price effective 1837 was $20.671835 (rounded) per fine ounce of gold, the British since 1717 £4.247727(+), whence dollar-sterling mint parity was $4.8665635 per pound sterling (Officer 1996, p. 51). There are actually several concepts of parity, for which (considering the dollar-sterling case) one may consult Officer (1996, Chap. 5; 2006). The lag of "legal parity" (for appraisal of British merchandise for tariffs) behind mint parity, catching up only in 1873, is an issue that warrants explanation by historians.

Gold Points and Gold-Point Arbitrage

A fixed exchange rate (at the mint parity) for two countries on the gold standard is an oversimplification, which is often made but is misleading. There were costs of importing or exporting gold. These costs included freight, insurance, handling (packing and cartage), interest on money committed to the transaction, risk premium (compensation for risk), normal profit, any deviation of purchase or sale price from the mint price, possibly mint charges, and possibly abrasion (wearing out or removal

of gold content of coin – should the coin be sold abroad by weight or as bullion). Expressing the exporting costs as percent of the amount invested (or, equivalently, as percent of parity), the product of 1/100th these costs and mint parity (number of units of domestic currency per unit of foreign currency, for example, number of dollars per pound) was added to mint parity to obtain the gold-export point, the exchange rate at which gold is exported. To obtain the gold-import point, the product of 1/100th of the importing costs and mint parity was subtracted from mint parity.

If the exchange rate was greater than the gold-export point, private-sector "gold-point arbitrageurs" exported gold, thereby obtaining foreign currency. Conversely, for the exchange rate less than the gold-import point, gold was imported and foreign currency relinquished. Usually the gold was, directly or indirectly, purchased from the monetary authority of the one country and sold to the monetary authority in the other. The domestic-currency cost of the transaction "per unit of foreign currency obtained" was the gold-export point. That "per unit of foreign currency sold" was the gold-import point. Also, foreign currency was sold, or purchased, at the exchange rate. Therefore, arbitrageurs receive a profit proportional to the exchange-rate/gold-point divergence.

However, the arbitrageur supply of foreign currency eliminates profit by returning the exchange rate to below the gold-export point. Therefore, perfect "gold-point arbitrage" would ensure that the exchange rate has upper limit of the gold-export point. Similarly, the arbitrageur demand for foreign currency returns the exchange rate to above the gold-import point, and perfect arbitrage ensures that the exchange rate has that point as a lower limit. It is important to note what induces the private sector to engage in gold-point arbitrage: (1) the profit motive and (2) the credibility of the commitment to (a) the fixed gold price and (b) freedom of foreign exchange and gold transactions, on the part of the monetary authorities of both countries.

Discussions of gold-point arbitrage are in Officer (1996, Chap. 8) and Canjels et al. (2004, pp. 871–875).

Spread, Gold-Point Estimation, and Gold-Effected Transfer of Funds

The "spread," the exchange-rate range over which arbitrage is unprofitable, is the difference between the gold-export point and gold-import point. It is sometimes convenient to express the gold points (and exchange rate) as percentage of parity. Then the spread becomes the sum of the gold points. Estimates of gold points and spreads involving core countries are presented in Table 3. There are many methods of obtaining or estimating gold points – Officer (1996, pp. 117–121) identifies nine techniques, which Canjels et al. (2004, p. 869) reduce to four. The main distinction is between summing cost components over time (method A in Table 3) and applying sophisticated time-series analysis to high-frequency, daily, exchange-rate data (methods B and C). Canjels et al. argue that their technique is superior to method A (exemplified by Officer 1996) and that their results – especially a narrower spread than estimated by Officer – are consistent with gold-flow data. However, the Canjels et al.'s symmetry assumption (implying gold export and import points equidistant from parity), perhaps made for analytic convenience, is at variance with historical evidence.

Table 3 Gold-point estimates

Countries	Period	Gold points (percent)		Spread (percent)	Method of computation
		Export	Import		
US/Britain	1841–1850	1.7476	3.2960	5.0436	A
US/Britain	1851–1860	1.3306	1.8631	3.1937	A
US/Britain	1881–1890	0.6585	0.7141	1.3726	A
US/Britain	1891–1900	0.6550	0.6274	1.2824	A
US/Britain	1901–1910	0.4993	0.5999	1.0992	A
US/Britain	1911–1914	0.5025	0.5915	1.0940	A
US/Britain	1879–1913	0.7706 to 0.1192		0.8898	B
US/Britain	1879–1913	0.4192 to 0.2486		0.6678	C
France/US	1877–1913	0.6888	0.6290	1.3178	D
Germany/US	1894–1913	0.4907	0.7123	1.2030	D
France/Britain	1877–1913	0.4063	0.3964	0.8027	D
Germany/Britain	1877–1913	0.3671	0.4405	0.8076	D
Germany/France	1877–1913	0.4321	0.5556	0.9877	D
Austria/Britain	1912	0.6453	0.6037	1.2490	E
Netherlands/ Britain	1912	0.5534	0.3552	0.9086	E
Scandinavia/ Britain	1912	0.3294	0.6067	0.9361	E

Sources: US/Britain, 1879–1913 – Canjels et al. (2004, pp. 879). US/Britain, other periods – Officer (1996, pp. 174). France/US, Germany/US, France/Britain, Germany/Britain, Germany/France – Morgenstern (1959, pp. 178–181). Austria/Britain, Netherlands/Britain, Scandinavia/Britain – Easton (1912, pp. 358–363).

Gold points apply to numerator country. Therefore, gold-export point is gold-import point for denominator country, and gold-import point is gold-export point for denominator country. Spread is gold-export point plus gold-import point. Scandinavia is Denmark, Sweden, and Norway

Method of computation: A sum of period-average arbitrage cost components; B exchange-rate behavior, nonparametric model; C exchange-rate behavior, smooth time-trend model; D median estimate of various authorities for various dates; E writer's estimate; B–E: converted to percent deviation from parity; B–C: Gold points symmetric and decline over time as shown, from beginning of period (maximum spread) to end of period (minimum spread); figure for spread is midpoint of maximum and minimum spread

Noteworthy in Table 3 is that the gold points and therefore the spread declined over time (evidenced by the dollar-sterling figures, whether methods A or B-C). Explanations involve technological improvements in transportation, communication, and arbitrage itself.

Almost always forgotten by economic historians is the fact that gold flows also were employed to transfer funds in lieu of a foreign-exchange transaction (rather than in combination with such transaction, per gold-point arbitrage). It is supremely ironic that contemporary accounts of such operations almost always pertain to gold-effected transfer of funds, whereas modern textbooks and scholarly articles deal exclusively with gold-point arbitrage! It is easy to demonstrate theoretically – and

Officer shows empirically – that the spread pertinent to transfer of funds was always narrower than the gold-point arbitrage spread.

Implications for Money Supply and Automatic Correctives

Consider a domestic gold standard. Under a pure coin standard, gold in circulation, monetary base, and money supply are all one. With a mixed standard, the money supply is the product of the money multiplier (dependent on the commercial-banks' reserves/deposit and the nonbank-public's currency/deposit ratios) and the monetary base (the actual and potential reserves of the commercial banking system, with potential reserves held by the nonbank public). The monetary authority alters the monetary base by changing its gold holdings and its loans, discounts, and securities portfolio (nongold assets, domestic assets). However, the level of its domestic assets is dependent on its gold reserves, because the authority generates demand liabilities (notes and deposits) by increasing its assets, and convertibility of these liabilities must be supported by a gold reserve, if the gold standard is to be maintained. Therefore, the gold standard provides a constraint on the level (or growth) of the money supply.

The international gold standard involves balance-of-payments surpluses settled by gold imports at the gold-import point and deficits financed by gold exports at the gold-export point. (Within the spread, there are no gold flows and the balance of payments is in equilibrium.) The change in the money supply is then the product of the money multiplier and the gold flow, providing the monetary authority does not change its domestic assets. For a country on a gold-exchange standard, holdings of "foreign exchange" (the reserve currency) take the place of gold. In general, the "international assets" of a monetary authority may consist of both gold and foreign exchange. Discussion of automatic correctives of a payments imbalance assumes "neutral" policy of the monetary authorities, that is, abstraction both from policies that would enhance and policies that would inhibit correction.

Traditional Mechanism

A country experiencing a balance-of-payments deficit loses gold and its money supply automatically decreases, the extent of the decrease depending on the legal or customary reserve requirements for nongold (or non-foreign-exchange) money (for the variety of legal institutions, see Martin-Aceña 2007, p. 105). Assuming that velocity does not increase (i.e., the demand for money does not decrease), money income contracts, via the equation of exchange. Then the price level and/or real income falls. If prices are fully flexible (guaranteed only by pure *and* perfect competition), then the price level bears the full force of the deflation. As long as elasticity conditions (moderate, but typically neglected in the literature) are satisfied, exports increase not only in real but also in nominal terms, and imports similarly

decrease. Symmetrically, a surplus country gains gold, the money supply increases, money income expands, the price level rises, exports decrease, and imports increase. In each case, balance-of-payments equilibrium is restored via the current account. This is called the price specie-flow mechanism; "developed in the eighteenth century, it remains the dominant approach to thinking about the gold standard today" (Eichengreen 2008, p. 24).

An extended adjustment mechanism incorporates changes in real income and interest rates. To the extent that prices are inflexible, movements of real income in the same direction as money income occur; in particular, the deficit country suffers unemployment, but the payments imbalance is nevertheless corrected.

The capital account also acts to restore balance, via the deficit-country reduced money supply increasing interest rates, inducing a net inflow of capital. The interest-rate increases also reduce real investment and thence real income and imports. Similarly, interest-rate decreases in the surplus country elicit capital outflow and increase real investment, income, and imports. This process enhances the price specie-flow current-account correction of the imbalance.

Monetary Mechanism

From a general monetarist standpoint, the traditional mechanism is unnecessary to restore payments equilibrium, because, with fixed exchange rates, gold flows simply adjust money supply to money demand. Changes in prices, real income, and interest rates are superfluous to the adjustment process. Further, under a "global-monetarist" framework, prices, interest rates, and incomes are all determined worldwide. Therefore, in logical extreme, the price-specie-flow and like mechanisms cannot even occur (on the monetarist approaches, see Kreinin and Officer 1978, Chap. 3).

For some authors (McCloskey and Richard Zecher 1976; Temin 1984, pp. 576–577; Gallarotti 1995, pp. 35–36), historical data support the monetary mechanism for the classical gold standard: Gold flows were too small to be suggestive of the traditional correctives, and prices, incomes, and interest rates moved closely in correspondence (rather than in the opposite directions predicted by the traditional adjustment mechanisms) – at least among non-outer-periphery countries, especially the core group. Hatton (1992) is skeptical of this work, while Wallace and Choudhry (1995) present evidence against global monetarism and in favor of the price specie-flow mechanism.

The "law of one price" – purchasing power in weak form – is associated with the monetary approach and contravenes price-specie flow. Examining 10 studies published during the period spanned by Enders (1989) and Catão and Solomou (2005), some of which are discussed in Officer (2012), I judge that eight provide at least partial support for PPP. However, PPP is generally found to be stronger over time, which leads to the open questions "how long is too long for the monetary approach to receive validation?" Undoubtedly, there remains scope for additional work on automatic correctives.

Sources of Instability of the Gold Standard

There were three elements making for instability of the classical gold standard. First, the use of foreign exchange as reserves increased as the gold standard progressed. Available end-of-year data indicate that, worldwide, foreign exchange in official reserves (the international assets of the monetary authority) increased by 36% from 1880 to 1899 and by 356% from 1899 to 1913. In comparison, gold in official reserves increased by 160% from 1880 to 1903 but only by 88% from 1903 to 1913 (Lindert 1969, pp. 22–25). While in 1913 only Germany among the center countries held any measurable amount of foreign exchange – 15% of total reserves excluding silver (which was of limited use) – the percentage for the rest of the world was double that for Germany (Table 4). If there were a rush to cash in foreign exchange for gold, reduction or depletion of the gold of reserve-currency countries could place the gold standard in jeopardy.

Second, Britain – the predominant reserve-currency country – was in a particularly sensitive situation. From 1899 to 1913, recorded sterling balances (mostly official) increased more than 2.5-fold (Lindert 1969, p. 22). Considering end of 1913 data, almost half of world foreign-exchange reserves was in sterling, but the Bank of England had only 3% of world gold reserves (Tables 5, 6). Defining the "reserve ratio" of the reserve-currency-country monetary authority as the ratio of (i) official reserves to (ii) liabilities to foreign monetary authorities held in financial institutions in the country, in 1913 this ratio was only 31% for the Bank of England, far lower than those of the monetary authorities of the other core countries (Table 7).

Table 4 Share of foreign exchange in official reserves: 1913 (end of year, percent)

	Including silver	Excluding silver
Britain	0	0
United States	0	0
France	0	0
Germany	13	15
Rest of world	27	31

Source: Lindert (1969, pp. 10–11)
Official reserves are gold, foreign exchange, and including or excluding silver

Table 5 Composition of World official foreign-exchange reserves: 1913

Currency	Percent
British pounds	47
US dollars	2
French francs	30
German marks	16
Other	5

Source: Lindert (1969, pp. 18–19)
End of year. Excludes holdings for which currency unspecified. "Other" is primarily Dutch guilders and Scandinavian kroner

Table 6 Official-reserves components: Percent of world total: 1913

Country	Gold	Foreign exchange
Britain	3	0
United States	27	0
France	14	0
Germany	6	5
Rest of world	50	95

Sources: Gold – Board of Governors of the Federal Reserve System (1943, pp. 544–545, 551). Foreign exchange – Lindert (1969, pp. 10–11)

Table 7 Reserve ratio of reserve-currency countries: 1913

Country	Including silver	Excluding silver
Britain	0.31	0.31
United States	90.55	64.42
France	2.38	2.02
Germany	2.11	1.75

Source: Lindert (1969, pp. 10–11, 19). Foreign-currency holdings for which currency unspecified allocated proportionately to the four currencies based on known distribution
End-of-year ratio of official reserves to official liquid liabilities (that is, liabilities to foreign governments and central banks). Proportion, not percent. Official reserves are gold, foreign exchange, and including or excluding silver

An official run on sterling could easily force Britain off the gold standard. Because sterling was an international currency, private foreigners also held considerable liquid assets in London and could themselves initiate a run on sterling.

Third, the United States, though a core country, was a great source of instability to the gold standard. The US Treasury accumulated and held a high percentage of world gold reserves (more than that of the three other core countries combined in 1913), resulting in an absurdly high reserve ratio – Tables 5, 6, 7). With a decentralized banking system composed of many banks of three distinct types (national, state, savings) – De Cecco 1984, pp. 111–113) includes loan and trust companies as a fourth group – operating under different rules, but with a New York center, interbank deposits were prevalent and financial crises involving bank failures frequent. Eichengreen (1992, p. 55) sees an analogy between interior banks maintaining balances in New York and the US financial system holding sterling balances in London banks. In addition to episodic financial shocks, there was periodic, seasonal financial stress, as monies would flow back and forth between the agricultural interior and the New York banking center. Cyclically, the US demand for money shifted greatly, but the supply was relatively inelastic. This led to episodic high interest rates in the New York money market, which attracted capital from abroad. Further, there was an upward trend in the demand for money on the part of the US private sector, which exacerbated the capital inflow.

During the heyday of the gold standard, the USA had no central bank to serve as a lender of last resort or otherwise help to stabilize the US monetary base. Provocatively, Officer (2002, pp. 115–117) has argued that the First and Second Banks of the

United States played the role of central bank, but these Banks had long vanished by the heyday of the gold standard. And the Federal Reserve had barely begun operations when the gold standard collapsed in 1914. The Treasury did not fill the void: "The US Treasury was by no means the lender of last resort of the American system; once it acquired gold, it just sat on it" (De Cecco 1984, p. 117).

Therefore, far from the United States assisting Britain, gold often flowed from the Bank of England to the United States to satisfy increases in US demand for money. Though in economic size the United States was the largest of the core countries, in many years it was a net importer rather than exporter of capital to the rest of the world – the opposite of the other core countries. The political power of silver interests (desiring to enhance the role of silver relative to gold), the accusations of farmer debtors and manufacturer exports (blaming the gold standard for deflation), and recurrent financial crises led to imperfect credibility in the US commitment to the gold standard. Runs on banks and runs on the Treasury gold reserve placed the US gold standard near collapse in the early and mid-1890s. During that period, the credibility of the Treasury's commitment to the gold standard was shaken. Indeed, the gold standard was saved in 1895 (and again in 1896) only by cooperative action of the Treasury and a bankers' syndicate, which stemmed gold exports.

Using time-series analysis on six-month commercial bank loans (deemed a "more developed" market, because that maturity is the longest series available), Tullio and Wolters (2000, p. 62, 67) conclude (based on previous work) that "the UK and London were more vulnerable to US influences in the period under study than they were to French and German influences" and (in their current study) that "All in all, the influence of US on UK interest rates is much stronger and lasts much longer than vice versa." The latter finding is stronger for 1897–1907 than for 1890–1896. I interpret these results as confirming the US unstable role in the gold standard.

In sum, the United States, by virtue of economic size and early experience with the gold standard, was a core country to be sure, but a core country that decidedly exacerbated the instability of the gold standard!

Rules of the Game

The Rules

According to the "rules of the [gold-standard] game," central banks were supposed to reinforce, rather than "sterilize" (moderate or eliminate) or ignore, the effect of gold flows on the monetary supply. A gold outflow typically decreases the international assets of the central bank and thence the monetary base and money supply. The central-bank's proper response was: (1) raise its discount rate, thereby inducing commercial banks to adopt a higher reserves/deposit ratio and therefore decreasing the money multiplier and (2) decrease lending and sell securities, thereby decreasing domestic interest-earning assets and thence the monetary base. On both counts, the money supply is further decreased. And the higher interest rate acted to increase interest rate generally and induce a capital inflow. The converse argument (involving

gold inflow, increases in the money supply, and lower interest rates) applies symmetrically to a gold inflow.

It is interesting that the "rules of the game" did not appear in the literature until a decade after the classical gold standard ended (Eichengreen 1992, p. 36). The originator was Keynes (1925, p. 18), who wrote that given the overvalued pound upon the UK return to the gold standard and the consequent payments imbalance and incipient gold loss: "The Bank of England is *compelled* to curtail credit by all the rules of the Gold Standard game." Such "credit restriction" (money-supply decrease, in today's parlance) deflates the economy (reduces nominal GDP), reducing wages (the price level) via unemployment (decreasing real GDP). External balance is maintained at the expense of internal balance. Anticipating his *General Theory*, Keynes advocates rather an "easy credit policy" (easy-money policy) to "restore prosperity" (full employment) rather than following the rules of the game and "aggravate a depression" (worsening real GDP).

Should the central bank rather increase its domestic assets when it loses gold, it engages in "sterilization" of the gold flow and is decidedly not following the "rules of the game." The converse argument (for gold inflow) also holds, with sterilization involving the central bank decreasing its domestic assets when it gains gold.

According to the monetary approach, neither the "rules of the game" nor sterilization can have any effect except in the short run. Under fixed exchange rates, gold flows simply adjust money supply to money demand; the money supply cannot be determined by policy. The central bank can control the (reserve-asset versus domestic-asset) composition of the monetary base but not the level of the base. Indeed, the rules of the game are unnecessary in the first instance, because gold flows occur only because of a disequilibrium between money demand and money supply. When gold (or any reserve) has moved sufficiently to re-equate money supply to money demand, the gold loss, or gain ceases. Thus rule (2) is unnecessary and is ineffective except possibly in the short run.

Furthermore, under global monetarism, interest rates and incomes are determined worldwide. Even core countries can influence these variables domestically only to the extent that they help determine them in the global marketplace. Therefore, rule (1) is inapplicable as well. In sum, the "rules of the game," whether followed or not, are deemed inconsequential by those who adhere to the monetary approach to the balance of payments.

Discount-Rate Rule

However, the Bank of England did, in effect, manage its discount rate ("Bank Rate") in accordance with rule (1). The Bank's primary objective was to maintain convertibility of its notes into gold, that is, to preserve the gold standard, and its principal policy tool was Bank Rate. When its liquidity ratio of gold reserves to outstanding note liabilities decreased, it would usually increase Bank Rate. The increase in Bank Rate carried with it market short-term increase rates, inducing a short-term capital inflow and thereby moving the exchange rate away from the gold-export point by increasing the exchange value of the pound. The converse would be a rise in the

liquidity ratio involving a Bank Rate decrease, capital outflow, and movement of the exchange rate away from the gold import point – but if the converse held, it was in weaker form. Nevertheless, the Bank was constantly monitoring its liquidity ratio and in response altered Bank Rate almost 200 times over 1880–1913.

Time-series analyses, such as Jeanne (1995) and Davutyan and Parke (1995), essentially support that narrative. No doubt the Bank had other objectives: certainly profitability, given that it was a commercial bank albeit with public functions, possibly, at times, economic activity ("home trade"). If Bank rate exceeded the market rate by too great a margin, the Bank's commercial business would suffer and shareholders would object. However, maintenance of note convertibility was required by law and viewed as necessary for the Bank's commercial functioning. So the goals of maintenance of convertibility and earning of satisfactory profits were not necessarily in conflict. In contrast, the studies show little concern for economic activity. The Bank's low gold holdings (which, of course, earned zero return) were viewed by contemporaries as based on an overriding concern for the interests of shareholders, that is, profitability (Gallarotti 1995, p. 115).

The Reichsbank operated in an environment similar to that of the Bank of England, except that the Reichsbank kept a greater reserve buffer and the Berlin money market was not as large as that of London. The Reichsbank, like the Bank of England, generally moved its discount rate inversely to its liquidity ratio.

However, most other central banks often violated the rule, with changes in their discount rates of inappropriate direction, or of insufficient amount or frequency. The Bank of France, in particular, kept its discount rate stable. Unlike the Bank of England, it chose to have large gold reserves (see Table 6), with payments imbalances accommodated by fluctuations in its gold rather than financed by short-term capital flows. This policy was due in part to a small money market in Paris. (Of course, the United States, lacking a central bank, had no discount rate to use as a policy instrument.)

Sterilization Was Dominant

As for rule (2) – that the central-bank's domestic and international assets move in the same direction – in fact the opposite behavior, sterilization, was dominant, as shown in Table 8. The Bank of England followed the rule more than any other central bank, but even so violated it more often than not! The sterilization policy of the Bank of France was a substitute for discount-rate policy (Bazot et al. 2016).

Was the Bank of England Supreme?

Eichengreen (1987) quotes Keynes that the "Bank of England could almost have claimed to be the conductor of the international orchestra" (p. 5) and finds that Bank rate tended to lead the Reichsbank discount rate and even the Bank of France rate. Other studies confirm that the Bank of England discount rate sometimes was followed by a change in the same direction on the part of the Reichsbank, but not

Table 8 Annual changes in international and domestic assets of central bank: 1880–1913

Country	Percent changes in same direction
Britain	48
France	26
Germany	31
Western Europe	32
Scandinavia	40
Russia	33

Source: Bloomfield (1959, p. 49)
International assets are gold, silver, and foreign exchange; domestic assets are income-earning: discounts, loans, and securities. Change in same direction implies country is following "rules of the game." Observations with zero or negligible changes in either class of assets excluded. Years when country is off gold standard excluded (see Table 1). Western Europe consists of Austria-Hungary, Belgium, and Netherlands; Scandinavia incorporates Denmark, Finland, Norway, and Sweden

the reverse. And Bazot et al. (2016), with advanced time-series analysis, find that French sterilization was ultimately due to an increase in the Bank of England rate: "the Banque de France's credit to the domestic economy (discounts and advances) correlates negatively with gold flows because it correlates positively with the discount rate of the Bank of England" (p. 2).

Morys (2013), making use of central bank archival data and sophisticated time-series analysis, examines the behavior of discount rates of 14 central banks and concludes that "a considerable amount of monetary autonomy was retained under the Classical Gold Standard, even for peripheral countries" (p. 215). However, Morys can be criticized for having no gold-flow or domestic-activity variables, for not conducting unit-root and cointegration testing, and for a principal-components solution to exchange-rate multicollinearity. Stokes and Neuburger (2016), in an ultra-sophisticated time-series analysis, determine that the London money-market rate heavily influenced the French and Reichsbank money-market rates. Their use of market rather than official interest rates brings richer data to bear on the issue of English leadership.

One concludes that the Bank of England was influential in determining foreign money-market conditions if not official rates, but Bank "leadership" or "hegemony" remains an open question and perhaps a matter of definition.

Stability of Gold Standard

How then did the classical gold standard cope with payments imbalances? Why was it a stable system?

Private-Sector Credibility in Convertibility

The fundamental reason for the stability of the classical gold standard is that there was always absolute private-sector credibility in the commitment to the fixed

domestic-currency price of gold on the part of the center country (Britain), two (France and Germany) of the three remaining core countries, and certain other European countries (Belgium, Netherlands, Switzerland, and Scandinavia). Certainly, that was true from the late-1870s onward. For the United States, this absolute credibility applied from about 1900. In earlier periods, that commitment had a contingency aspect: it was recognized that convertibility could be suspended in the event of dire emergency (such as war); but, after normal conditions were restored, convertibility would be re-established at the preexisting mint price and gold contracts would again be honored. The Bank Restriction Period is an example of the proper application of the contingency, as is the greenback period (even though the United States, effectively on the gold standard, was legally on bimetallism). An excellent discussion of "the gold standard as a contingent rule" is Bordo and Kydland (1996).

The absolute credibility in countries' commitment to convertibility at the existing mint price implied that there was extremely low, essentially zero, convertibility risk (the probability that Treasury or central-bank notes would not be redeemed in gold at the established mint price) and also exchange risk (the probability that the mint parity between two currencies would be altered or that exchange control or prohibition of gold export would be instituted).

Reasons Why Commitment to Convertibility Was So Credible

There were many reasons why the commitment to convertibility was so credible.

1. Contracts were expressed in gold; if convertibility were abandoned, contracts would inevitably be violated – an undesirable outcome for the monetary authority.
2. Shocks to the domestic and world economies were infrequent and generally mild. There was basically international peace and domestic calm.
3. The London capital market was the largest, most open, most diversified in the world, and its gold market was also dominant. A high proportion of world trade was financed in sterling, London was the most important reserve-currency center, and balances of payments were often settled by transferring sterling assets rather than gold. Therefore, sterling was an international currency – not merely supplemental to gold but perhaps better: a boon to noncenter countries, because sterling involved positive, not zero, interest return, and its transfer costs were much less than those of gold. Advantages to Britain were the charges for services as an international banker, differential interest returns on its financial intermediation, and the practice of countries on a sterling (gold-exchange) standard of financing payments surpluses with Britain by piling up short-term sterling assets rather than demanding Bank of England gold.
4. There was widespread ideology – and practice – of "metallist orthodoxy" and "monetary orthodoxy" (Gallarotti 1995), involving authorities' commitment to an anti-inflation, balanced-budget, stable-money policy. In particular, the ideology implied low government spending and taxes and limited monetization of

government debt (financing of budget deficits by printing money). Therefore, it was not expected that a country's price level or inflation would get out of line with that of other countries, with resulting pressure on the country's adherence to the gold standard.

5. This ideology was mirrored in, and supported by, domestic politics. Gold had won over silver and paper and stable-money interests (bankers, industrialists, manufacturers, merchants, professionals, creditors, urban groups) over inflationary interests (farmers, landowners, miners, debtors, rural groups).
6. There was freedom from government regulation and a competitive environment, domestically and internationally. Therefore, prices and wages were more flexible than in other periods of human history (before and after). The core countries had virtually no capital controls, the center country (Britain) had adopted free trade, and the other core countries had moderate tariffs. Balance-of-payments financing and adjustment could proceed without serious impediments.
7. Internal balance (domestic macroeconomic stability, at a high level of real income and employment) was an unimportant goal of policy. Preservation of convertibility of paper currency into gold would not be superseded as the primary policy objective. While sterilization of gold flows was frequent, the purpose was more "meeting the needs of trade" (passive monetary policy) than fighting unemployment (active monetary policy).
8. The gradual establishment of mint prices over time ensured that the implied mint parities (exchange rates) were in line with relative price levels; so countries joined the gold standard with exchange rates in equilibrium.
9. Current-account and capital-account imbalances tended to be offsetting for the core countries, especially for Britain. A trade deficit induced a gold loss and a higher interest rate, reducing capital outflow, and attracting a capital inflow. Indeed, the capital-exporting core countries – Britain, France, and Germany – could eliminate a gold loss simply by reducing lending abroad.

Rareness of Violations of Gold Points

Many of the above reasons not only enhanced credibility in existing mint prices and parities but also kept international-payments imbalances, and hence necessary adjustment, of small magnitude. Responding to the essentially zero convertibility and exchange risks implied by the credible commitment, private agents further reduced the need for balance-of-payments adjustment via gold-point arbitrage. When the exchange rate moved beyond a gold point, arbitrage acted to return it to the spread. So it is not surprising that "violations of the gold points" were rare on a monthly average basis, as demonstrated in Table 9 for the dollar, franc, and mark exchange rate versus sterling. Certainly, gold-point violations did occur; but they rarely persisted sufficiently to be counted on monthly average data. Such measured violations were generally associated with financial crises.

The number of dollar-sterling violations for 1890–1906 exceeding that for 1889–1908 is due to the results emanating from different researchers using different

Table 9 Violations of gold points

Exchange rate	Time period	Number of months	Violations	
			number	Percent of months
Dollar-sterling	1889–1908	240	1	0.4
Dollar-sterling	1890–1906	204	3	1.5
Franc-sterling	1889–1908	240	12	5.0
Mark-sterling	1889–1908	240	18	7.5

Sources: Dollar-sterling, 1890–1906 – Officer (1996, p. 235). Other, Giovannini (1993, pp. 130–131); numbers are approximate, deciphered from graph

data. Nevertheless, the important common finding is the low percent of months encompassed by violations. Canjels et al. (2004), using daily exchange-rate data, find that Officer's gold-point spread is too wide to accommodate recorded gold flows. On the other hand, Spiller and Wood (1988, p. 888), working with weekly exchange rates, conclude that "Many instances of alleged gold-point violations identified by previous authors, then, may have been nothing more than instances in which arbitrage costs may have been larger than average." This conundrum cries out for richer gold-point and gold-flow data and, of course, careful and appropriate time-series analysis.

Stabilizing Speculation

The perceived extremely low convertibility and exchange risks gave private agents profitable opportunities not only outside the spread (gold-point arbitrage) but also within the spread (exchange-rate speculation). As the exchange value of a country's currency weakened, the exchange rate approaching the gold-export point, speculators had an ever greater incentive to purchase domestic currency with foreign currency (a short-term capital inflow); for they had good reason to believe that the exchange rate would move in the opposite direction, whereupon they would reverse their transaction at a profit. Similarly, a strengthened currency, with the exchange rate approaching the gold-import point, involved speculators selling the domestic currency for foreign currency (a short-term capital outflow). Clearly, the exchange rate would either not go beyond the gold point (via the actions of speculators of this ilk) or would quickly return to the spread (via gold-point arbitrage). Also, the further the exchange rate moved toward the gold point, the greater the potential profit opportunity, for there was a decreased distance to that gold point and an increased distance from the other point.

This "stabilizing speculation" enhanced the exchange value of depreciating currencies that were about to lose gold, and thus the gold loss could be prevented. The speculation was all the more powerful, because the absence of controls on capital movements meant private capital flows were highly responsive to exchange-rate changes. Dollar-sterling data, in Table 10, show that this speculation was extremely efficient in keeping the exchange rate away from the gold points – and

Table 10 Average deviation of dollar-sterling exchange rate from gold-point-spread midpoint

Time period	Percent of parity	Percent of spread
Quarterly observations		
1881–1890	0.32	23
1891–1900	0.25	19
1901–1910	0.15	13
1911–1914	0.12	11
Monthly observations		
1890–1906	0.24	20

Source: Officer (1996, p. 272). Year 1914 ends with second quarter

increasingly effective over time. Interestingly, these statements hold even for the 1890s, during which at times US maintenance of currency convertibility was precarious. The average deviation of the exchange rate from the midpoint of the spread fell decade-by-decade from about 1/3 of 1% of parity in 1881–1890 (23% of the gold-point spread) to only 12/100th of 1% of parity in 1911–1914 (11% of the spread).

Under basic target-zone theory, credibility is 100% operationally; thus, the exchange rate never violates the spread ("target zone") and exhibits "smooth pasting" at the gold points (Duarte et al. 2013). Hallwood et al. (1996) test the assumption of full credibility and find that it is a reasonable description of the sterling-franc exchange rate but that "instances of positive devaluation expectations of the dollar are. . .common" (p. 191), though large only during 1890–1896, consistent with the discussion in the previous section "Sources of Instability of the Gold Standard".

Government Policies that Enhanced Gold-Standard Stability

Government policies also enhanced gold-standard stability. First, by the turn of the century, South Africa – the main world gold producer – sold all its gold in London, either to private parties or actively to the Bank of England, with the Bank serving also as residual purchaser of the gold. Thus, the Bank had the means to replenish its gold reserves. Second, the orthodox-metallism ideology and the leadership of the Bank of England – other central banks would often gear their monetary policy to that of the Bank – kept monetary policies harmonized. Monetary discipline was maintained.

Third, countries used "gold devices," primarily the manipulation of gold points, to affect gold flows. Consider the Bank of England. By law, the Bank had to redeem its notes in domestic gold coin (sovereigns) at a minimum price equivalent to £3 17 s. 10½d. per standard ounce of gold and purchase gold bars at a minimum price of £3 17 s. 9d. Beyond that, the Bank had tremendous discretion. It would foster gold imports by lowering the foreign gold-export point (British gold-import point, number of units of foreign currency per pound) through interest-free loans to gold importers or raising its purchase price for bars and foreign coin. The Bank would

discourage gold exports by lowering the foreign gold-import point (British gold-export point) via increasing its selling prices for gold bars and foreign coin, refusing to sell bars, or redeeming its notes in underweight domestic gold coin. These policies were alternative to increasing Bank Rate.

The US Treasury followed similar policies at times. In addition to providing interest-free loans to gold importers and changing the premium at which it would sell bars (or refusing to sell bars outright), the Treasury condoned banking syndicates to put pressure on gold arbitrageurs to desist from gold export in 1895 and 1896, a time when the US adherence to the gold standard was under stress. Officer (1996, Chap. 9) provides detailed data on the two countries' gold devices from a bilateral standpoint.

The Bank of France and Reichsbank employed gold devices relative to discount-rate changes more than Britain did. Some additional policies included converting notes into gold only in Paris or Berlin rather than at branches elsewhere in the country, the Bank of France converting its notes in silver coin rather than gold (permitted under its "limping" gold standard), and the Reichsbank using moral suasion to discourage the export of gold. Gold devices combined with a huge buffer stock of gold enabled the Bank of France to keep its discount rate stable while maintaining convertibility. In the 1900s, the Bank stopped the use of gold devices, replacing them with foreign-exchange market intervention (Bazot et al. 2016).

Also, the monetary system was adept at conserving gold, as evidenced in Table 2. This was important, because the increased gold required for a growing world economy could be obtained only from mining or from nonmonetary hoards. While the money supply for the eleven-major-country aggregate more than tripled from 1885 to 1913, the percent of the money supply in the form of metallic money (gold and silver) more than halved. This process did not make the gold standard unstable, because gold moved into commercial-bank and central-bank (or Treasury) reserves: the ratio of gold in official reserves to official plus money gold increased from 33% to 54%. The relative influence of the public versus private sector in reducing the proportion of metallic money in the money supply is an issue warranting exploration by monetary historians.

Further, while the stable environment in which the gold standard operated did not require regular central-bank cooperation, such cooperation was forthcoming when needed, that is, during financial crises. Although Britain was the center country, the precarious liquidity position of the Bank of England meant that it was more often the recipient than the provider of financial assistance. In crises, it would obtain loans from the Bank of France (also on occasion from other central banks), and the Bank of France would sometimes purchase sterling to push up that currency's exchange value. "Interestingly, it was because France cared so much more about domestic finance, that it came to care about international finance. . ..International markets represented the first line of defense for French finance; i.e. mitigating the problem at the source" (Eichengreen 2008, p. 33). Assistance also went from the Bank of England to other central banks, as needed. And cooperation went beyond the core countries. "In effect, the resources on which any one country could draw when its gold parity was under attack extended beyond its own reserves to those that could be

borrowed from other gold-standard countries" (Eichengreen 2008, p. 33). Further, the credible commitment was so strong that private bankers did not hesitate to make loans to central banks in difficulty. Cooperation during the gold standard is discussed by Gallarotti (1995, Chap. 3) and Eichengreen (1992, pp. 48–52).

In sum, "virtuous" two-way interactions were responsible for the stability of the gold standard. The credible commitment to convertibility of paper money at the established mint price and therefore the fixed mint parities were both a cause and a result of (1) the stable environment in which the gold standard operated, (2) the stabilizing behavior of arbitrageurs and speculators, and (3) the responsible policies of the authorities – and (1), (2), and (3), and their individual elements, also interacted positively among themselves.

Experience of Periphery

An important reason for periphery countries to join and maintain the gold standard was the access to the capital markets of the core countries thereby fostered. Adherence to the gold standard connoted that the peripheral country would follow responsible monetary, fiscal, and debt-management policies – and, in particular, faithfully repay the interest on and principal of debt. This "Good Housekeeping Seal of Approval," by reducing the risk premium, involved a lower interest rate on the country's bonds sold abroad, and very likely a higher volume of borrowing. The favorable terms and greater borrowing enhanced the country's economic development. However, Flandreau and Zumer (2004) argue and demonstrate that gold-standard adherence was unimportant in explaining international interest-rate spreads. Rather, a country's debt burden (ratio of interest-service to revenue) and default history were the crucial explanatory variables. This finding detracts from the alleged advantage of the gold standard to the periphery.

Furthermore, periphery countries bore the brunt of the "burden of adjustment" of payments imbalances with the core (and other Western European) countries, in four ways. First, some of the periphery countries were on a gold-exchange standard. When they ran a surplus, they typically increased (and with a deficit, decreased) their liquid balances in London (or other reserve-currency country) rather than withdraw gold from (or ship gold to) the reserve-currency country. The monetary base of the periphery country would increase, or decrease, but that of the reserve-currency country would remain unchanged. This meant that such changes in domestic variables – prices, incomes, interest rates, portfolios, etc. – that occurred to correct the surplus or deficit were primarily in the periphery country. The periphery, rather than the core, bore the burden of adjustment.

Second, the nongold (silver and inconvertible-paper) periphery was subject to substantial exchange-rate variability, which altered real exchange rates, generating core-periphery payments adjustment. "From the perspective of the core, exchange rate flexibility in the periphery facilitated international relative price adjustment, while maintaining the monetary stability required for the preservation of the gold peg" (Catão and Solomou 2005, p. 1272).

Third, when Bank Rate increased, London drew funds from France and Germany, which attracted funds from other Western European and Scandinavian countries, which drew capital from the periphery. Also, it was easy for a core country to correct a deficit by reducing lending to, or bringing capital home from, the periphery, thus bringing about "sudden stops" to the capital inflow of periphery countries.

Fourth, the periphery countries were underdeveloped; their exports were largely primary products (agriculture and mining), which inherently were extremely sensitive to world market conditions. This feature made adjustment in the periphery compared to the core take the form more of real than financial correction. This conclusion also follows from the fact that capital obtained from core countries for the purpose of economic development was subject to interruption and even reversal ("sudden stops"). While the periphery was probably better off with access to the capital than in isolation, its welfare gain was reduced by the instability of capital import.

Fifth, peripheral countries were subjected to financial crises more than the core. Bordo and Meissner (2011, p. 85) show that "higher capital inflows were strongly related to a higher probability of having any kind of crisis." They examine the roles of "original sin" (hard-currency-denomination debt), "currency mismatches" (lack of assets-liabilities offsets in foreign-currency debt), and "debt intolerance" (past defaults) in 30 countries over 1880–1913. Their results "tend to conform that it is difficult to find robust determinants of financial crises." However, they find a strange quadratic relationship between "the ratio of hard-currency debt to total debt" and debt crises. Economic historians look for patterns, of which either the absence or the weirdness is frustrating: the answer generally lies in further research.

The experience on adherence to the gold standard differed among periphery groups. The important British dominions and colonies – Australia, New Zealand, Canada, and India – successfully maintained the gold standard. They were politically stable and, of course, heavily influenced by Britain. They paid the price of serving as an economic cushion to the Bank of England's financial situation; but compared to the rest of the periphery, gained a relatively stable long-term capital inflow. The European periphery had the advantage of emigrant remittances, which, according to Esteves and Khoudour-Castéras (2009, p. 980), served as a substitute for capital inflows and "were instrumental in relieving [international] credit constraints to developing nations." Some European periphery countries "shadowed the gold standard." Even with inconvertible paper currency, they maintained relatively stable exchange rates and prices, thus largely behaving as if they were on the gold standard (see, for example, Tattara and Volpe 1997, Tattara 2003, Martín-Aceña et al. 2012).

In undeveloped Latin American and Asia, adherence to the gold standard was fragile, with lack of complete credibility in the commitment to convertibility. Many of the reasons for credible commitment that applied to the core countries were absent – for example, there were powerful inflationary interests, strong balance-of-payments shocks, and rudimentary banking sectors. For Latin America and Asia, the cost of adhering to the gold standard was very apparent: loss of the ability to depreciate the currency to counter reductions in exports. Yet the gain, in terms of

Table 11 Country-group statistics: 1881–1913 (component-country means)

	Gold-standard adherence	Money growth		Government deficit		Inflation	
Country group	Mean	Mean	SD	Mean	SD	Mean	SD
Core	100	4.0	3.8	0.8	0.9	0.3	3.4
Scandinavia	100	5.6	4.2	0.3	0.7	0.4	3.2
Western Europe	100	4.2	3.5	2.1	0.3	0.6	3.6
Dominions	100	5.4	5.5	7.8	1.6	0.4	2.6
Southern Europe	22	2.5	6.7	0.6	0.7	0.3	2.6
Latin America	32	3.3	15.9	2.3	3.1	4.0	13.6
Japan	52	7.2	14.5	−3.1	3.3	4.6	5.5

Sources: Gold-standard adherence – Table 1. Other columns – Bordo and Schwartz (1996, pp. 46–47, 52–53, 58–59). SD = standard deviation

Gold-standard adherence is percent of years 1881–1913 on gold standard. Money growth is the time coefficient from annual regression of natural logarithm of M2 on constant and time trend. Government deficit is percent of GNP. Inflation is the time coefficient from annual regression of natural logarithm of GDP deflator (or equivalent) on constant and time trend

Core: Britain, United States, France, Germany. Scandinavia: Denmark, Finland, Norway, Sweden. Western Europe: Belgium, Netherlands, Switzerland. Dominions: Australia, Canada. Southern Europe: Italy, Portugal, Spain. Latin America: Argentina, Brazil, Chile

a steady capital inflow from the core countries, was not as stable or reliable as for the British dominions and colonies.

Comparisons of periphery-country experience with the core and with each other are presented in Table 11. It is perhaps surprising that Southern Europe exhibits even less adherence to the gold standard than does Latin America. In terms of money growth, there is a schism (seen most clearly in the coefficient of variation – ratio of mean to standard deviation) between stability for the core, Scandinavia, Western Europe, and Dominions, versus instability for the remaining periphery. The figures for government deficit have some anomalies; but the core, Scandinavia, and Western Europe certainly exhibit "monetary orthodoxy." Inflation level and variability are relatively high for Latin America and Japan. The figures do not uniformly reflect the quantity theory of money; but, except for the Southern Europe anomaly, there is broadly an association between gold-standard adherence and stable money in all senses.

Performance

Performance of the gold standard is reasonably evaluated via contrast with alternative international monetary systems, whether past or future – and the possible criteria are various. Consider first, in Table 12, monetary criteria for the US heyday gold standard (1879–1913) in comparison with previous US systems: First and Second Banks (1792–1810, 1817–1838), Independent Treasury (1847–1861), intermittent paper standards (1811–1816, 1839–1846), and greenback period (1862–1878). Very revealing, but almost neglected in the historical literature, is exchange-market

Table 12 US monetary statistics: Gold standard versus other periods

	Exchange-market pressure (period mean, percent)		Ratio of monetary base to specie stock (end of year)	
Period	Algebraic value	Absolute value	Mean	Coefficient of variation (percent)
1792–1810	0.83	7.09	1.22	6.86
1811–1816	7.04	8.20	1.25	17.80
1817–1838	2.74	6.51	1.27	11.45
1839–1846	4.89	7.76	1.06	4.69
1847–1861	−17.00	17.89	1.00	10.81
1862–1878	−10.26	10.27	3.72	42.77
1879–1913	0.63	2.69	2.17	14.80

Source: Officer (2002, p. 135, 137)
Statistics of annual values

pressure (EMP) as a criterion of performance. Under certain assumptions (no money illusion, money-market equilibrium, purely monetary model, small open economy – assumptions Officer 2002, p. 134, defends for the US gold-standard period), EMP in favor of the domestic currency (US dollar) is the unweighted sum of "payments imbalance as percentage of the monetary base" and "percentage change in the foreign-currency price of the dollar." Whether taking the mean of algebraic or absolute values of EMP as the criterion, EMP is lowest for the gold standard – in fact, by a multiple in 11 of 12 comparisons. At least for the United States and for whatever reasons, the classical gold standard worked to minimize exchange-market pressure better than all previous alternatives.

The other monetary criterion is decidedly unfavorable to the gold standard. The ratio of the monetary base to specie stock ("pyramiding ratio") measures discipline in restricting the monetary base. Under a pure coin standard, the ratio is unity; so the ideal ratio is a zero coefficient of variation around a unitary mean. It is to be expected that the greenback period is least disciplined, but the gold standard follows as second (mean) or third (coefficient of variation) highest pyramiding ratio. Paradoxically, the flexible ratio may help to explain the high gold-standard EMP efficiency. It might also reflect the unstable role of the United States in the working of the gold standard!

Restricting comparisons to later and different monetary systems, Tables 13 and 14 measure inflation and real per-capita income growth for the four core countries. In mean inflation, the gold-standard is tops, with all four countries having lower inflation than under Bretton Woods or floating exchange rates. However, for no country is the variability of inflation lowest. And for no country does the gold standard entail maximum mean growth – Bretton Woods exhibits highest mean growth – or minimum standard deviation of growth. Of course, historical time periods reflect more than differential monetary systems; but the gold-standard balancing of relatively low inflation with relatively low growth is suggestive of a trade-off offered by that system.

Representative of pertinent time-series-analysis literature is the careful study for the gold-standard core provided by Bordo et al. (2010), who present a somewhat different conclusion. They distinguish "good deflation" (induced by positive supply

Table 13 Inflation in core countries: Gold standard versus later periods

Country	Gold standard (1881–1913)		Bretton Woods (1946–1970)		Floating exchange rates (1974–1995)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Britain	0.3	3.1	3.9	2.2	7.5	5.6
United States	0.3	3.0	2.5	3.5	5.0	2.4
France	0.0	4.9	5.0	3.5	6.4	3.8
Germany	0.6	2.6	2.7	4.0	3.2	1.3

Source: Bordo and Schwartz (1999, p. 205)
Mean inflation is time coefficient from annual regression of natural logarithm of GDP deflator (or equivalent) on constant and time trend. For United States, gold-standard mean is 0.4 using alternative data

Table 14 Growth of core countries: Gold standard versus later periods

Country	Gold standard (1881–1913)		Bretton Woods (1946–1970)		Floating exchange rates (1974–1995)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Britain	1.1	2.4	2.1	1.8	1.8	2.3
United States	1.8	4.9	2.0	4.6	1.5	2.3
France	1.5	4.7	4.1	2.1	1.7	1.5
Germany	1.7	2.8	5.0	3.3	1.1	4.9

Source: Bordo and Schwartz (1999, p. 205)
Mean growth is time coefficient from annual regression of natural logarithm of real per capita GDP (or equivalent) on constant and time trend. For United States, gold-standard mean and standard deviation are 1.6 and 2.7, respectively, using alternative data

shocks) from "bad deflation" (resulting from negative demand shocks). Although they do not make comparisons with other periods in this paper, they do see supply shocks (productivity improvements) as having significant effects on growth. Their time-series analysis shows a structural break around 1896, whereby deflation preceded inflation. For the European core, money is essentially neutral; but for the United States, monetary shocks significantly affect output – not unexpected for this core country.

Breakdown of Gold Standard

The classical gold standard was at its height at the end of 1913, ironically just before it came to an end. The proximate cause of the breakdown of the classical gold standard was political: the advent of World War I in August 1914. However, it was the Bank of England's precarious liquidity position and the gold-exchange standard

that were the underlying cause. With the outbreak of war, a run on sterling led Britain to impose extreme exchange control – a postponement of both domestic and international payments – that made the international gold standard nonoperational. Convertibility was not legally suspended; but moral suasion, legalistic action, and regulation had the same effect. Gold exports were restricted by extralegal means (and by Trading with the Enemy legislation), with the Bank of England commandeering all gold imports and applying moral suasion to bankers and bullion brokers.

Almost all other gold-standard countries undertook similar policies in 1914 and 1915. The United States entered the war and ended its gold standard late, adopting extralegal restrictions on convertibility in 1917 (although in 1914 New York banks had temporarily imposed an informal embargo on gold exports). An effect of the universal removal of currency convertibility was the ineffectiveness of mint parities and inapplicability of gold points: floating exchange rates resulted.

The classical gold standard possessed strong elements both of stability and instability. In the end, the shock of war led to dominance of the unstable forces. It is an open question how long the gold standard would have lasted had World War I not brought it to a close, that is, whether and when the forces making for instability would have overcome those supporting stability.

Cross-References

- ▶ Currency Unions
- ▶ International Currencies in the Lens of History
- ▶ The Evolution of US Monetary Policy
- ▶ The Evolution of Monetary Policy (Goals and Targets) in Western Europe
- ▶ The Historical Evolution of Central Banking

References

Bazot G, Bordo MD, Monnet E (2016) International shocks and the balance sheet of the Bank of France under the classical gold standard. Explor Econ Hist 62:87–107

Bloomfield AI (1959) Monetary policy under the international gold standard: 1880–1914. Federal Reserve Bank of New York, New York

Board of Governors of the Federal Reserve System (1943) Banking and monetary statistics 19140-1941. National Capital Press, Washington, DC

Bordo MD (1993) The gold standard, Bretton woods and other monetary regimes: a historical appraisal. Fed Reserve Bank St Louis Rev 75(2):123–191

Bordo MD, Kydland FE (1996) The gold standard as a commitment mechanism. In: Bayoumi T, Eichengreen B, Taylor MP (eds) Modern perspectives on the gold standard. Cambridge University Press, Cambridge, pp 55–100

Bordo MD, Meissner CM (2011) Foreign capital, financial crises and incomes in the first era of globalization. Eur Rev Econ Hist 15(1):61–91

Bordo MD, Rockoff H (1996) The gold standard as a 'good housekeeping seal of approval. J Econ Hist 56(2):389–428

Bordo MD, Schwartz AJ (1996) The operation of the specie standard. In: de Macedo JB, Eichengreen B, Reis J (eds) Currency convertibility. Routledge, London, pp 11–83
Bordo MD, Schwartz AJ (1999) Monetary policy regimes and economic performance: the historical record. In: Taylor JB, Woodford M (eds) Handbook of Macroeconomics, vol 1A. Elsevier, Amsterdam, pp 149–234
Bordo MD, Landon-Lane J, Redish A (2010) Deflation, productivity shocks and gold: evidence from the 1880–1914 period. Open Econ Rev 21(4):515–546
Canjels E, Prakash-Canjels G, Taylor AM (2004) Measuring market integration: foreign exchange arbitrage and the gold standard, 1879–1913. Rev Econom Stat 86(4):868–882
Catão LAV, Solomou SN (2005) Effective exchange rates and the classical gold standard adjustment. Am Econ Rev 95(4):1259–1275
Conant CA (1915) A history of modern banks of issue. Putnam's, New York
Davutyan N, Parke WR (1995) The operations of the Bank of England, 1890-1908: a dynamic Probit approach. J Money Credit Bank 27(4):1099–1112
De Cecco M (1984) The international gold standard. St. Martin's, New York
Dimitrova K, Fantacci L (2010) The establishment of the gold standard in Southeast Europe: convergence to a new system or divergence from an old one? In: Baubeau P, Ögren A (eds) Convergence and divergence of National Financial Systems. Pickering & Chatto, London, pp 179–196
Duarte AP, Andrade JS, Duarte A (2013) Exchange rate target zones: a survey of the literature. J Econ Surv 27(2):247–268
Easton HT (1912) Tate's modern Cambist. Effingham Wilson, London
Eichengreen B (1987) Conducting the international orchestra: Bank of England leadership under the classical gold standard. J Int Money Financ 6(1):5–29
Eichengreen B (1992) Golden fetters. Oxford University Press, Oxford
Eichengreen B (2008) Globalizing capital, 2nd edn. Princeton University Press, Princeton
Eichengreen B, Flandreau M (1996) The geography of the gold standard. In: de Macedo JB, Eichengreen B, Reis J (eds) Currency convertibility. Routledge, London, pp 113–143
Enders W (1989) Unit roots and the real exchange rate before World War I: the case of Britain and the USA. J Int Money Finance 8(1):59–73
Esteves R, Khoudour-Castéras D (2009) A fantastic rain of gold: European migrants' remittances and balance of payments adjustment during the gold standard period. J Econ Hist 69(4):951–985
Flandreau M, Jobst C (2005) The ties that divide: a network analysis of the international monetary system, 1890–1910. J Econ Hist 65(4):977–1007
Flandreau M, Zumer F (2004) The making of global finance 1880–1913. OECD, Paris
Flandreau M, Le Cacheux J, Zumer F, Dornbusch R, Honohan P (1998) Stability without a pact? Lessons from the European gold standard, 1880-1914. Econ Policy 13(26):115–162
Gallarotti GM (1995) The anatomy of an international monetary regime. Oxford University Press, Oxford
Giovannini A (1993) Bretton woods and its precursors: rules versus discretion in the history of international monetary regimes. In: Bordo MD, Eichengreen B (eds) A retrospective on the Bretton woods system. University of Chicago Press, Chicago, pp 109–147
Hallwood CP, MacDonald R, Marsh IW (1996) Credibility and fundamentals: were the classical and interwar gold standards well-behaved target zones? In: Bayoumi T, Eichengreen B, Taylor MP (eds) Modern perspectives on the gold standard. Cambridge University Press, Cambridge, pp 129–161
Hatton TJ (1992) Price determination under the gold standard: Britain 1880–1913. In: Broadberry SN, Crafts NFR (eds) Britain in the international economy. Cambridge University Press, Cambridge, pp 137–156
Jeanne O (1995) Monetary policy in England 1893-1914: a structural VAR analysis. Explor Econ Hist 32(3):302–326
Keynes JM (1925) The economic consequences of Mr. Churchill. Hogarth, London

Kreinin ME, Officer LH (1978) The monetary approach to the balance of payments: a survey. International Finance Section, Princeton University, Princeton
Lindert PH (1969) Key currencies and gold, 1900–1913. International Finance Section, Princeton University, Princeton
Martín-Aceña P (2007) The gold standard: a review from the periphery. In: Cottrell PL, Notaras G, Tortella G (eds) From the Athenian tetradrachm to the euro. Ashgate, Aldershot, pp 96–112
Martín-Aceña P, Ruiz EM, Nogues-Marco P (2012) Floating against the tide: Spanish monetary policy, 1870–1931. In: Ogren A, Oksendal LF (eds) The gold standard peripheries. Palgrave Macmillan, London, pp 145–173
McCloskey DN, Richard Zecher J (1976) How the gold standard worked, 1880-1913. In: Frenkel JA, Johnson HG (eds) The monetary approach to the balance of payments. University of Toronto Press, Toronto, pp 357–385
Meissner CM (2005) A new world order: explaining the international diffusion of the gold standard, 1870-1913. J Int Econ 66(2):385–406
Milward AS (1996) The origins of the gold standard. In: de Macedo JB, Eichengreen B, Reis J (eds) Currency convertibility. Routledge, London, pp 87–101
Mitchener KJ, Weidenmier MD (2015) Was the classical gold standard credible on the periphery? Evidence from currency risk. J Econ Hist 75(2):479–511
Morgenstern O (1959) International financial transactions and business cycles. Princeton University Press, Princeton
Morys M (2013) Discount rate policy under the classical gold standard: core versus periphery (1870s-1914). Explor Econ Hist 50(2):205–226
Officer LH (1996) Between the dollar-Sterling gold points. Cambridge University Press, Cambridge
Officer LH (2002) The U.S. specie standard, 1792–1932: some monetarist arithmetic. Explor Econ Hist 39(2):113–153
Officer LH (2006) Dollar-sterling parity: 1789–1978. In Carter SB et al (eds) Historical statistics of the United States, millennial edn. pp 5-561–563
Officer LH (2008) Gold standard. In: Whaples R (ed) EH.Net encyclopedia. EH.Net, Oxford https://eh.net/encyclopedia/gold-standard/
Officer LH (2012) Purchasing power parity in economic history. In: James J, Marsh IW, Sarno L (eds) Handbook of exchange rates. Wiley, Hoboken, pp 161–187
Spiller PT, Wood RO (1988) Arbitrage during the Dollar-Sterling gold standard, 1899–1908: an econometric approach. J Polit Econ 96(4):882–892
Stokes HH, Neuburger HM (2016) The “thin film of gold” – interest rates and gold flows in the classical gold-standard era (Unpublished)
Tattara G (2003) Paper money but a gold debt: Italy in the gold standard. Explor Econ Hist 40(2):122–142
Tattara G, Volpe M (1997) Italy, the fiscal-dominance model, and the gold-standard age. In: Marcuzzo MC, Officer LH, Rosselli A (eds) Monetary standards and exchange rates. Routledge, London, pp 229–263
Temin P (1984) Comment. In: Bordo MD, Schwartz AJ (eds) A retrospective on the classical gold standard, 1821–1931. University of Chicago Press, Chicago, pp 576–581
Triffin R (1964) The evolution of the international monetary system. International Finance Section, Princeton University, Princeton
Tullio G, Wolters J (2000) Interest rate linkages between the US and the UK during the classical gold standard. Scottish J Polit Econom 47(1):61–71
Wallace MS, Choudhry T (1995) The gold standard: perfectly integrated World markets or slow adjustment of prices and interest rates? J Int Money Financ 14(3):349–371

International Monetary Regimes: The Interwar Gold Exchange Standard

25

Olivier Accominotti

Contents

O. Accominotti (✉)
London School of Economics and Political Science, London, UK

Centre for Economic Policy Research, London, UK
e-mail: O.Accominotti@lse.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_24

Abstract

Historical accounts of the international monetary system generally oppose the classical gold standard of 1880–1914 and its interwar successor of 1925–1931. While the pre-WW1 gold standard is usually described as a paragon of international monetary and price stability, its interwar version remains associated with memories of foreign exchange market turbulence, global deflation, and, of course, the Great Depression. This chapter provides an overview of the interwar gold exchange standard system. How did this system emerge in the 1920s? How was it implemented in practice? Why did it collapse in the 1930s? And what was the link between the interwar gold exchange standard and the Great Depression?

Keywords

Gold exchange standard · Foreign exchange reserves · Currency stabilization · Devaluations · Great Depression

Introduction

Historical accounts of the international monetary system generally oppose the classical gold standard of 1880–1914 and its interwar successor of 1925–1931. While the pre-WW1 gold standard is often described as a paragon of international monetary and price stability, its interwar version remains associated with memories of exchange market turbulence, global deflation, and, of course, the Great Depression. Scholars have described the interwar monetary system as a severely flawed arrangement and argued that its poor design largely contributed to the global economic slump of the 1930s (Choudhri and Kochin 1980; Eichengreen and Sachs 1985; Temin 1989; Bernanke and James 1991; Eichengreen 1992).

In a gold standard system, monetary authorities are required to fix the value of the domestic currency in terms of a given weight of gold and to sell and purchase banknotes against gold at that fixed parity. The exchange rate between two currencies simultaneously on the gold standard fluctuated within a narrow band around the ratio of their gold parities. While most countries had left the gold standard during the First World War, they progressively returned to this regime in the postwar years. The adoption of the gold standard by a majority of countries in the 1920s therefore resulted in the emergence of a *global fixed exchange rate system*.

In essence, the interwar gold standard was not different from other historical fixed exchange rate regimes such as the pre-WW1 classical gold standard or the post-WW2 Bretton Woods system. For many countries, the restoration of the gold standard was simply the re-establishment – with significant amendments – of a monetary regime that had been in place since the mid-nineteenth century and

which had been only temporarily suspended during the war and postwar years. But, when it comes to monetary arrangements, the evil is often in the details. In its implementation, the interwar gold standard differed from its nineteenth century predecessor in several dimensions. In particular, several countries chose to implement a *gold exchange standard* in which central banks backed their note circulation with foreign currency reserves in addition to gold. The international political context of the interwar years was also markedly different from that of the pre-WW1 and post-WW2 periods, and this greatly affected the operation of the international monetary system.

This chapter provides a general overview of the history of the interwar gold exchange standard. I first present the progressive steps towards the appearance of this system as well as the historical context of its emergence. I then describe how the gold standard was implemented in practice in the interwar years, and the events leading to its collapse in the 1930s. Finally, the chapter discusses the link between the interwar gold exchange standard and the Great Depression.

The Emergence of the Interwar International Monetary Regime

The First Word War and Collapse of the Classical Gold Standard

The interwar gold exchange standard did not emerge in one day but was the outcome of a slow process of monetary construction that took place during the 1920s. The First World War had precipitated the collapse of the classical gold standard system and marked the end of a long era of fixed exchange rates. On 28 July 1914, the archduke and heir of the Austro-Hungarian crown Franz-Ferdinand was assassinated in Sarajevo, triggering a conflict in which most European powers were involved. As the European conflict unfolded into a long-lasting war, the mobilization of economic and financial resources became critical for achieving final victory. In order to finance their military expenses, belligerent countries' governments increased taxation and sold bonds to domestic citizens. But this proved insufficient, and they soon also had to borrow directly from their national note-issuing institute – or central bank – which financed these loans through monetary issuance. Since it implied issuing additional banknotes relative to the central bank's gold reserves, debt monetization jeopardized the maintenance of the gold standard, and its operational principles had to be adapted. Central banks restricted the conditions under which they delivered gold against banknotes and attempted to restrict exports of specie, either through moral suasion or official embargoes. They also imposed restrictions on foreign exchange activities and capital outflows (Eichengreen 1992, pp. 70–73).

These policies had important consequences on the foreign exchange market. European belligerent countries saw their currencies depreciate against the US dollar as the volume of their note issuance increased (Hall 2004). In 1916–1919, the US government directly purchased sterling at the exchange rate of 4.76 ½ US dollars per pound (as opposed to the pre-war parity of 4.86) and also supported the French franc (Atkin 2005, pp. 25–26). This helped to stabilize these currencies temporarily.

But when foreign exchange interventions were suspended in March 1919 following the end of the hostilities, European currencies all depreciated sharply relative to the US dollar and gold.

The Return to the Gold Standard: International Coordination Versus Individual Initiatives

The following period was marked by severe inflation in Europe and dramatic foreign exchange market instability. Currency investors initially expected European currencies to re-appreciate and be stabilized at their pre-war gold parities (Einzig 1937). However, as the period of floating exchange rates was prolonged, uncertainty regarding the timing and conditions at which each country would return to gold contributed to exchange rate volatility. Figure 1 displays the share of countries (out of a 42 countries sample) which were officially on the gold standard and did not apply exchange controls, from January 1919 to December 1936 (the data are from Crafts and Fearon 2010). The reconstruction of the international gold standard took place progressively from the early 1920s and was not fully achieved until the year 1928.

To what extent was this process coordinated? In the early 1920s, efforts were made to bring governments of various countries together in order to discuss the terms under which they should stabilize their currency. The first of these endeavors took place in Brussels, where representatives of 39 countries as well as independent experts met in September 1920. The meeting was actively prepared by

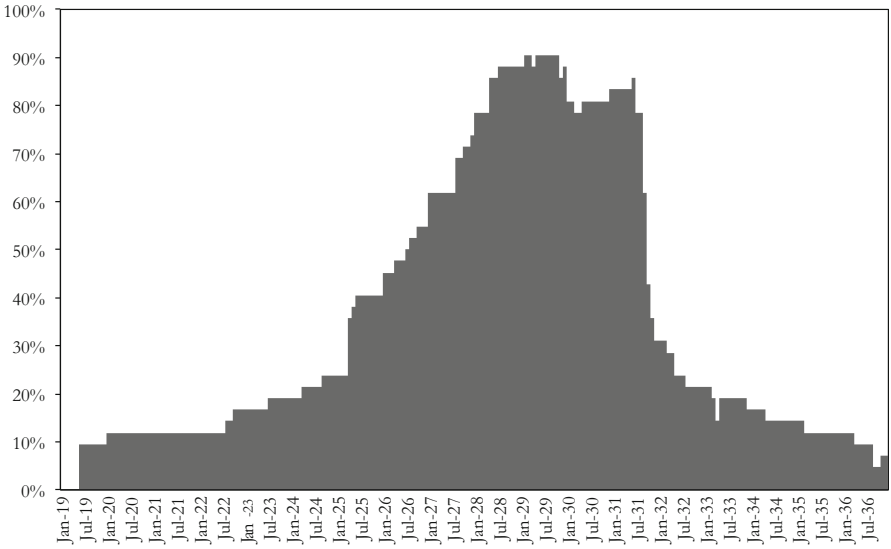


Fig. 1 Share of countries officially on the gold standard, 01/1919–12/1936. (Source: Constructed using data reported in Crafts and Fearon (2010))

the Secretariat of the Economic and Financial Section of the League of Nations, and no less than 14 statistical volumes and memoranda were prepared for the occasion (Siepmann 1920). In its final statement, the Currency and Exchange Committee mentioned the desirability of a return to the gold standard, but there was little in the way of concrete recommendations. The Committee simply noted that conditions needed to be "sufficiently favourable" and that "it [was] impossible to say how or when all the older countries would be able to return to their former measure of effective gold standard or how long it would take the newly formed countries to establish such a standard." Contemporaries mocked the vagueness of these statements. For example, German Finance Minister Wirth noted that "it was of little use for the doctors to find out what was the matter with [a sick man] if they did not prescribe for his return to health" (*The Times*, 28 October 1920 quoted in Siepmann 1920, pp. 458–459).

A second international conference took place in Genoa two years later, in April–May 1922, where a series of 12 more specific resolutions on currency were adopted (Hawtrey 1922). In particular, resolution 11 recommended national governments to allow their central banks to hold foreign exchange as a complement to gold in the reserves used to back notes in circulation – a system that came to be known as the gold exchange standard.

However, there was nothing binding in the Genoa resolutions, and each country eventually decided individually of the timing of its return to the gold standard and of the way of actually implementing it. Exchange rate developments of the immediate postwar years were critical in this respect. Figure 2 displays monthly exchange rates of selected European currencies against the US dollar from December 1919 to June 1928, expressed as a percentage of the pre-WW1 exchange rate parities. Two distinct paths can be identified in 1919–1920. While all European currencies experienced a fall in their gold value, depreciation remained relatively limited for the Swiss franc, Netherlands guilder, and pound sterling. These currencies later re-appreciated and were eventually stabilized at their 1914 gold parity. By contrast, the Belgian franc, French franc, and Italian lire fell much more sharply just after the end of the war, and, at the end of 1920, they had all already lost between 67% and 82% of their pre-war gold value. The situation was even more extreme in Germany where the mark had depreciated by almost 95% relative to gold in December 1920. For those countries, re-appreciating the currency to its pre-WW1 parity revealed impossible, and they all had to reduce its gold content considerably before stabilizing.

Restoring Pre-WW1 Gold Parities: The United States and United Kingdom

The first country to officially return to the gold standard was the United States. While the dollar had remained convertible into gold during the war, the US government de facto suspended the gold standard as it strictly limited domestic residents' ability to export gold abroad and imposed restrictions on foreign exchange market transactions (Crabbe 1989; Eichengreen 1992). During the war and postwar years,

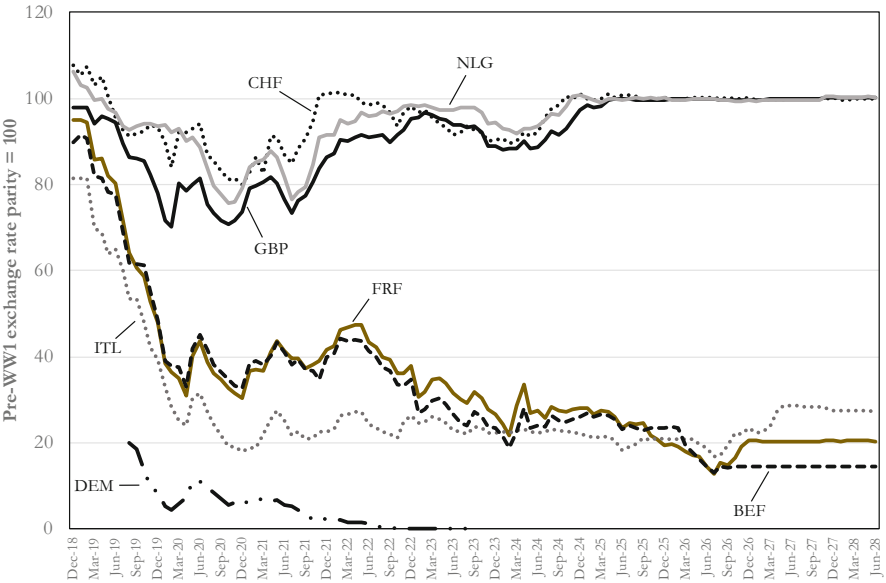


Fig. 2 Selected European currencies' exchange rates against the US dollar 12/1918–06/1928. (Note: *BEF*, Belgian franc; *CHF*, Swiss franc; *DEM*, German mark; *FRF*, French franc; *GBP*, pound sterling; *ITL*, Italian lira; *NLG*, Netherlands guilder. (Source: Monthly exchange rates data are from Accominotti and Chambers (2016) and were originally collected from the Financial Times)

European countries ran sizable trade deficits against the United States, which were financed through gold shipments and borrowing. The US monetary gold stock increased by more than 80% between 1914 and 1917 (Crabbe 1989, pp. 425–426), and the United States emerged from the war as the world's largest creditor. This new status – combined with the moderate inflation the United States had experienced during the war – paved the way for an early restoration of the gold standard. Restrictions on gold exports were eventually lifted in June 1919.

The road toward the restoration of the gold standard was more cumbersome in the United Kingdom. The pound's dollar price had fallen from a high of 4.765 in January 1919 to a low of 3.414 in February 1920, and Britain's economic policies in the following years were all oriented toward re-appreciating the currency in order to stabilize it at its pre-war dollar exchange rate of 4.866. Achieving this goal required reducing the domestic price level through deflationary monetary and fiscal policies. The pound first appreciated rapidly between August 1921 and February 1923 but then fell again between March 1923 and January 1924. Appreciation then resumed, and on 28 April 1925, the Chancellor of the Exchequer Winston Churchill finally declared that the 1920 parliamentary act that suspended the gold standard would not be prolonged.

The decision to restore the pound's pre-war gold parity was severely criticized by John Maynard Keynes in a 1925 pamphlet, *The Economic Consequences of*

Mr. Churchill. Based on an analysis of consumer price indices, Keynes argued that this parity overvalued sterling relative to the US dollar. Maintaining the fixed dollar/sterling exchange rate would require the United Kingdom to engage in further deflationary policies. Because nominal salaries were slow to adjust, such policies were bound to increase real wages and fuel unemployment (Keynes 1925). Among the skeptics also stood the chairman of one of the largest British commercial banks (the Midland Bank) and former Chancellor of the Exchequer Reginald McKenna. Although not opposed in principle to the re-establishment of the gold standard, McKenna thought this regime should not be restored too quickly and certainly not before relative price levels had become consistent with the old exchange rate parities (Arthmar and McLure 2016).

To what extent did the $4.86 parity actually overvalue sterling? This question has been subject to debate in the literature. While Moggridge (1972) and Redmond (1989) have argued that Britain's monetary problems of the 1920s were mostly the consequence of an overvalued exchange rate, authors such as Matthews (1986) and, more recently, James (2001) have disputed the view that sterling was much overvalued. Once restored however, sterling's gold parity proved difficult to maintain. Only six years after Churchill's announcement, Britain left the gold standard once again – and that time definitively.

Devaluing Before Stabilizing: France and Italy

France made a different choice than Britain and did not restore its currency's pre-war gold parity in the 1920s. After the US government stopped supporting European currencies in 1919, the French franc depreciated rapidly and lost more than 65% of its gold value between January 1919 and April 1920. Attempts to stabilize the currency in the following years failed in the face of political instability, persistent budget deficits, and high inflation. The French government placed its hopes of balancing the budget in the prospect of receiving war reparations from Germany and, in the meantime, it asked the Bank of France to monetize new debt issues. However, as reparation payments were not coming due, the French President of the Conseil (Prime Minister) Raymond Poincaré decided to invade the Ruhr region in January 1923, where the heart of Germany's coal, steel, and iron industry was located, in order to directly seize German production of raw materials. The policy revealed a complete failure and did not help to resolve French budgetary problems because German workers responded through protests, strikes, and an attitude of "passive resistance." Fears that Poincaré's successor, the center-left Edouard Herriot, might impose a wealth tax resulted in capital flight and a resurgence of exchange rate instability. France underwent two severe speculative attacks in 1923–1924 and 1925–1926, and currency instability only ended when Raymond Poincaré came back to responsibilities in 1926 as the head of a national unity government. The new government announced its willingness to adopt a drastically different fiscal policy, therefore contributing to stabilize inflationary expectations (Sargent 1984). In addition, Poincaré's return to power removed the prospect of a capital levy

(Prati 1991). From July to December 1926, the franc rose relative to the US dollar as investors repatriated their capital. French authorities however worried that foreign exchange speculation could destabilize the franc and result in its excessive appreciation (Sicsic 1993). In December 1926, the government mandated the Bank of France to purchase gold, pounds sterling, and US dollars in order to stabilize the currency at its current value corresponding to 20% of its pre-war parity. The Bank accumulated enormous amounts of sterling and dollar reserves in the following years, and France officially restored its currency's gold convertibility through the monetary law of 25 June 1928. The new gold parity significantly undervalued the French franc relative to sterling.

The Italian lire also experienced large swings on the foreign exchange market in the 1920s. After an initial depreciation in 1919–1920, the currency stabilized against the US dollar in 1921–1924 but then fell prey to another speculative attack in 1925–1926 coinciding with the troubles of the French franc (Cohen 1972). Fascist leader Benito Mussolini reacted strongly to this episode. In a speech held in Pesaro on 18 August 1926, he affirmed his willingness to defend the lire "to the last breath, to the last drop of blood" (cited by Cohen 1972, p. 649). The Fascist government's economic policy was subsequently oriented toward the goal of re-appreciating the lire to the symbolic exchange rate of 90 lire per pound sterling (or 19 lire per US dollar) – the so-called *Quota Novanta* – and deflationary policies were implemented in the following year (Cohen 1972). With the help of a USD 75 million credit from the Federal Reserve and Bank of England and a private bank syndicate's loan of 50 million US dollar, the Bank of Italy replenished its international reserves. A Royal Decree-Law officially restored the lire's gold convertibility on 21 December 1927 (The Economist, 31 December 1927). While the government had decided not to appreciate the lire to a level as high as its pre-war gold parity, the exchange rate appreciation undergone in the final year before stabilization nevertheless left the lire overvalued relative to sterling and the dollar, and it resulted in severe deflationary pressures on the Italian economy.

Post-hyperinflation Stabilization: Germany, Austria, and Hungary

Whereas France and Italy had struggled to reduce inflation in the 1920s, central European countries fell into hyperinflation. In those countries, the return to the gold standard came as part of a comprehensive package of monetary and fiscal reforms that aimed to stabilize domestic prices and restore confidence in the currency.

In Germany, inflation had its roots in the First World War when the lack of a large domestic market for government debt and difficulties in increasing taxation led authorities to engage into more debt monetization than other belligerents (Balderston 1989). In the 1920s, the country's economic policy was crippled by persistent budget deficits and the problem of war reparations. Following the end of the war, the German mark fell at a vertiginous rate in comparison to other European currencies (Fig. 2). Unable to close its budget, the German government borrowed extensively from the Reichsbank which monetized its debts. The resulting inflation in turn

eroded the real value of tax revenues and contributed to dig the deficit even further, therefore requiring more debt monetization and note issuance. Fiscal and monetary problems reinforced each other in a vicious spiral that eventually led the country toward hyperinflation (Holtfrerich 1986; Sargent 1982; Eichengreen 1992). Between January 1922 and October 1924, the German wholesale price index was multiplied by 35.6 billion, and the monthly inflation rate reached more than 10,000% in November 1923 (Sargent 1982, Table G1).

As emphasized by Sargent (1982), agents' expectations of future price movements played a crucial role in Germany's hyperinflation and were self-fulfilling. Agents expecting a rise in domestic prices required increases in their nominal wages, which, in turn, contributed to raise German firms' costs of production and induced them to actually increase their prices. A fiscal shock was needed in order to break this process and reverse inflationary expectations. In November 1923, a monetary reform made the central bank effectively independent from the government and introduced a new currency, the Rentenmark (Sargent 1982, pp. 83–84). In order to anchor inflationary expectations, the gold standard was reintroduced. The selected gold parity was symbolically chosen as being the same as the old Reichsmark's pre-WW1 parity (although one new mark corresponded to only one trillion old gold marks). These monetary and currency reforms – in conjunction with the rescheduling of reparation debts obtained under the Dawes plan – allowed Germany to escape hyperinflation.

Austria and Hungary underwent similar – although less dramatic – hyperinflation episodes in the early 1920s. Just like in neighboring Germany, the Austrian and Hungarian inflations arose from difficulties at balancing the budget. In order to put their public finances into order and stabilize prices, both countries obtained international loans arranged by the League of Nations. These loans came conditionally to the implementation of fiscal and monetary reforms conducted under the scrutiny of an external *commissioner-general* appointed by the League. The Austrian and Hungarian crowns' gold prices were stabilized in, respectively, July 1923 and July 1924, and both countries officially joined the gold standard in, respectively, March and April 1925. Austria's adoption of the gold standard coincided with the introduction of a new currency, the schilling, while the pengö replaced the Hungarian crown as of January 1927 (Garber and Spencer 1994).

Peripheral Countries

Due to their small size, greater openness to trade, or large foreign debts, countries at the periphery of the international monetary system enjoyed little autonomy when deciding of the timing of their return to the gold standard.

The experience of small European countries in the 1920s illustrates this phenomenon. In those countries, exchange rate policy was largely determined by the behavior of the major international currencies on the foreign exchange market (Straumann 2010). In 1921–1922, optimism regarding a return of the United Kingdom to the gold standard resulted in an appreciation of the Swiss franc, Dutch

guilder, and Swedish krone against the US dollar. The Swedish krone soon reached its pre-war parity, while the Swiss franc appreciated even above its 1914 gold value for a few months (Straumann 2010, pp. 75–76). While Sweden decided to restore the krone's gold convertibility on 1 April 1924, Switzerland and the Netherlands waited until the United Kingdom returned to the gold standard in April 1925. The Norwegian and Danish krone had depreciated much more sharply than these other currencies in the early 1920s. However, they started appreciating rapidly following Britain's restoration of the gold standard until their stabilization at their pre-war level in, respectively, January 1927 and May 1928.

Belgium, by contrast, did not restore its currency's pre-war gold parity, but its monetary experience in the 1920s mirrored that of France. The country had also been severely hit by the First World War, and Belgian authorities hoped that reparations from Germany would help to eliminate the budget deficit. The Belgian franc's behavior on the foreign exchange market closely followed that of the French franc, and both currencies experienced speculative attacks at roughly the same time in 1923–1924 and 1925–1926. Exchange rate stabilization eventually occurred in October 1926 for Belgium after a few months of currency appreciation following the end of a currency crisis in July. With the help of an international loan, Belgium restored the franc's gold convertibility, at only 14% of its pre-war parity (Eichengreen 1992, p. 170).

As a peripheral country in the international monetary system, Japan also faced significant constraints on its exchange rate policy. The Japanese economy was shackled by several episodes of bank runs and stock market collapses in the 1920s, and the country ran persistent trade deficits depleting its international reserves. However, when Osachi Hamaguchi became Prime Minister in July 1929, he strongly committed to bring the yen back to the its pre-WW1 gold parity, and this move finally took place in January 1930 just after the onset of the Great Depression (Shizume 2012). The government's decision to appreciate the yen resulted in deflation and the Japanese economy contracted. However, Japan was highly dependent on foreign capital and, before rolling over their credits to the Japanese government, foreign banking creditors – especially JP Morgan - had insisted that the yen be stabilized (Metzler 2006, pp. 192–195).

Staying Out: Spain and China

While the restoration of the gold exchange standard was almost universal in the 1920s, a few countries did not adopt the new regime. One important European exception was Spain whose currency was never fixed to gold in the interwar years. The Spanish peseta underwent considerable depreciation against the US dollar in 1927–1931 at a time when most European currencies had been stabilized. However, it would be inaccurate to say that Spanish monetary authorities deliberately renounced to pegging the peseta in order to enjoy the benefits of an autonomous monetary policy. Both the Spanish government and Bank of Spain in fact remained seriously worried about the peseta's depreciation, and significant efforts were made

to stabilize the currency. In March 1931, the Bank of Spain received a USD 60 million credit from foreign banks in order to assist it in its attempt at restoring the gold standard (Financial Times, 27 March 1931). However, the political crisis of April 1931 triggered a severe banking and currency crisis and marked the end of this project (Jorge-Sotelo 2018). It is therefore mostly against its will that Spain stayed out of the gold standard in the interwar years.

China's interwar monetary experience was also untypical in that the country was on a silver rather than on a gold standard. Since the price of silver declined relative to gold in the late 1920s and early 1930s, the Chinese dollar depreciated against currencies of countries that had restored the gold standard. Exchange rate depreciation benefitted the Chinese economy when most countries were sliding into the Great Depression in 1929–1931. However, considerable deflationary pressures were felt subsequently in 1934–1935 when silver appreciated following the devaluation of the dollar and the silver standard was abandoned in 1935 (Friedman 1992).

The Functioning of the Interwar Gold Standard

What's in a Word? The Gold *Exchange* Standard

How did central banks implement the gold standard in practice? In a so-called *gold coin standard*, gold coins constitute the bulk of a country's money supply. However, the progressive development of fractional reserve banking systems over the eighteenth and nineteenth centuries had resulted in the increased use of paper money (banknotes). This allowed for the emergence of a *gold bullion standard*, a system in which gold coins only marginally circulated as currency, but notes issued by the government, central bank, or private banks were convertible into gold at a fixed parity. During the First World War, gold coins almost entirely disappeared from circulation and were replaced by paper money.

The prospect of currency stabilization after the First World War caused fears that a return to the gold standard would lead to a rise in the demand for gold, therefore pushing up its price relative to other commodities and leading to deflation. Monetary experts such as Hawtrey (1919) and Cassel (1920) suggested methods to prevent a scarcity of monetary gold. They first recommended governments to restrict the circulation of gold as currency and to encourage the use of paper money (Cassel 1920, p. 40). They also suggested that central banks should hold foreign exchange instead of gold in the reserves backing their note circulation. This system was known as the *gold exchange standard* and was defined by British economist Ralph Hawtrey (1922, p. 294) as follows:

> The principle is that the currency of each participating country, instead of being convertible into gold, may be convertible at par into the currencies of the other.

Representatives at the Genoa conference of 1922 made concrete proposals for the implementation of an international monetary system based on this principle. They

recommended the creation of an "international convention … to coordinate the demand for gold" (Resolution 9). A few countries would "establish a free market in gold and therefore become gold centres," while other members of the convention, "in addition to any gold reserve held at home, may maintain in any other participating country reserves of approved assets in the form of bank balances, bills, short-term securities, or other suitable liquid resources" (Resolution 11).

While economists such as Hawtrey, Cassel, and Keynes strongly favored this scheme, several others remained skeptical. Mlynarski (1929) argued that movements in central banks' foreign exchange reserves were potentially destabilizing. Those reserves would typically be placed in the form of short-term bank deposits and other credit instruments, and their sudden withdrawals could disrupt credit conditions in the gold centers as well as make it difficult for these centers' central banks to control domestic credit (Mlynarski 1929, p. 89). The risks of such reserves reversals were especially high if most central banks' foreign reserves were concentrated in a few gold centers. These centers' external liabilities would increase relative to their own gold reserves, and their ability to convert central banks' foreign exchange reserves into gold would therefore be jeopardized – a phenomenon later emphasized by economist Robert Triffin in the context of the Bretton Woods dollar system. Therefore, Mlynarski argued, a relatively large number of reserve currencies/gold centers was needed in order to mitigate the risks inherent to the gold exchange standard.

Legal Provisions

Although the "international convention" recommended by the Genoa experts never saw the light of day, many countries decided to follow the conference's recommendations regarding the gold exchange standard. The practical implementation of this system varied greatly across countries. As noted by the League of Nations (1931, p. 9), "never in history has the gold standard been simultaneously applied in exactly the same manner in all the countries which are roughly classified as adhering to it." Typically, the return to the gold standard was accompanied with the passage of a new monetary law which defined the central bank's legal reserve requirements and set a legal *cover ratio*, i.e., the minimum share of the monetary circulation (defined as the sum of banknotes in circulation and sight liabilities of the central bank) that the issuing institute had to cover with international reserves. Legislators also decided what type of assets could be included in the cover reserves and, in particular, whether foreign currency assets were eligible.

Table 1 describes central banks' legal reserve requirements in 45 countries in the year 1930. Cover ratios were very heterogenous and ranged from 24% of the monetary circulation in Hungary to 100% in India and Siam (Thailand). Several countries like the United Kingdom, Finland, Japan, and Norway adopted a *fiduciary issue* system in which the central bank was required to cover all banknotes issued above a certain amount with either specie or foreign assets.

Table 1 Legal reserve requirements of 45 countries on the gold standard, 1930

Country	Type	Legal reserve requirements
Albania	GES	*33.3%* of the National Bank's banknotes to be held in *(a) gold and silver* coin and bullion or *(b) foreign values* in currencies not subject to abnormal fluctuations (up to 66.6% of total cover)
Austria	GES	*24%* of the National Bank's note circulation and sight liabilities to be held in *(a) gold, silver and token money* and *(b) foreign banknotes and bills of exchange* in currencies not subject to extraordinary fluctuations[a]
Belgium	GES	*40%* of the National Bank's total sight liabilities to be held in *(a) gold* (at least 30% of sight liabilities) and *(b) foreign exchange* convertible into gold
Bolivia	GES	*50%* of the Central Bank's notes to be held in *(a) gold* coins and bars, *(b) deposits payable in gold* on demand in banks of high standing in London or New York, c) Bolivian silver coins (up to 10% of banknotes in circulation)
Bulgaria	GES	*33.3%* of the National Bank's notes and demand liabilities to be held in *(a) gold* coin and bullion, *(b) gold foreign exchange* in currencies convertible into gold or foreign exchange at a fixed price (including balances with foreign central banks, foreign currency bills of exchange, and Treasury bills maturing within 3 months)
Canada	GS	Minister of Finance required to keep a minimum *gold* reserve of 25% of Dominion notes. Banks of issue required to keep a minimum *gold* reserve of 5% of issued banknotes with the Minister of Finance
Chile	GES	*50%* of the Central Bank's note circulation and deposits to be held in (a) *gold* coin or bar or (b) *deposits in first-class banks in New York and London* payable at sight in gold
Colombia	GES	*60%* of the Republic of Colombia's note circulation and deposits to be held in *(a) gold* or *(b) demand deposits payable in gold* in banks of high standing in foreign financial centers (up to 2/5 of total cover)
Costa Rica	GS	*50%* of the private banks' note circulation to be held in domestic or foreign *gold* coin. International Bank of Costa Rica not required to hold any fixed reserve against its notes
Cuba	GS	No bank of issue. *Gold peso coins* and *US dollars* legal tender
Curaçao	GS	*33.3%* of the Bank of Curaçao's notes to be held in *specie and bullion*
Czechoslovakia	GES	*25%* of the National Bank's note circulation and other sight engagements to be held in *(a) gold* coin or bullion (at least 50% of total cover), *(b) foreign banknotes*, or *(c) foreign exchange* in the form of drafts or recoverable balances held with unquestionably good banks in the principal banking centers of Europe and America[b]
Danzig	GES	*33.3%* of the Bank of Danzig's note circulation to be held in *(a) gold coin*, *(b) notes of the Bank of England*, or *(c) sight assets in sterling* with the Bank of England. Any notes issued over 40 million gulden have to be covered in full
Denmark	GES	*50%* of the National Bank's note circulation to be held in (a) *gold* coin or bullion (at least 60% of total cover), (b) *balances with the Bank of Norway and Swedish Riksbank*, or (c) *balance on giro account with the German Reichsbank*

(*continued*)

Table 1 (continued)

Country	Type	Legal reserve requirements
Ecuador	GES	*50%* of the Central Bank's notes and deposits to be held in *(a) gold* coins or bars, *(b) deposits in first-class banks in New York or London* payable at sight or at 3-day notice in gold
Estonia	GES	*40%* of the Eesti Pank's notes and demand liabilities to be held in *(a) gold* coin or bullion, *(b) foreign exchange* convertible into exportable gold including balances with banks abroad, foreign bills of exchange up to 3-month maturity, and obligations of foreign governments up to 6-month maturity
Finland	GES	*100%* of the Bank of Finland's note circulation and sight liabilities in excess of 1200 million marks (fiduciary issue) to be held in *gold or undisputed foreign balances*
France	GS	*35%* of the Bank of France's note circulation and current credit accounts to be held in *gold*
Germany	GES	*40%* of the Reichsbank's note circulation to be held in *(a) gold* (at least 75% of total cover) or (b) *foreign exchange* including banknotes, bills of exchange maturing in less than 14 days, demand deposits with banks of known solvency in foreign financial centers
Greece	GES	*40%* of the Bank of Greece's note circulation and demand liabilities to be held in *(a) gold* coin and bullion, *(b) foreign gold exchange* in currencies convertible into gold or foreign exchange (including balances with foreign central banks, bills of exchange, and Treasury bills maturing within 3 months)
Guatemala	GES	*40%* of the Central Bank's note circulation to be held in *(a) gold* or *(b) deposits in first-class credit institutions abroad*. At least one third of total cover must consist of coin, of which 90% must be gold
Hungary	GES	*24%* of the National Bank's note circulation and sight liabilities to be held in *(a) gold, silver*, *and token money* or *(b) foreign exchange* in currencies not subject to extraordinary fluctuations (including foreign banknotes, foreign bills of exchange, and credit balances and cash deposits abroad available at sight and held in banks of unquestioned solvency)
Iceland	GES	*37.5%* of the National Bank's note circulation to be held in *(a) gold* (at least 75% of total cover) or *(b) demand deposits in foreign banks* which are considered by the Board of Governors to be perfectly safe and are sanctioned by the Minister of Finance. Reserve may not be allowed to fall below two million kronur
India	GS	*100%* of the note circulation to be held by the government in *(a) rupees*, *(b) silver half rupees*, or *(c) gold bullion*
Italy	GES	*40%* of the Bank of Italy's note circulation and sight liabilities to be held in *(a) gold* or *(b) foreign exchange* convertible into gold
Japan	GS	*100%* of the Bank of Japan's note circulation in excess of 120 million yen (fiduciary issue) to be held in *(a) gold* coin or bullion (at least 75% of total cover) or *(b) silver* coin and bullion
Latvia	GES	*50%* of the Bank of Latvia's note circulation below 100 million lats, *75%* of the circulation between 100 and 150 million lats, and *100%* of the circulation above 150 million lats to be held in (a) gold coin or bullion, (b) silver coin or bullion, or (c) stable and sure foreign currencies

(*continued*)

Table 1 (continued)

Country	Type	Legal reserve requirements
Lithuania	GS	*33.3%* of the Bank of Lithuania's note circulation to be held in *gold*
Mexico	GS	*50%* of the Bank of Mexico's note issue to be held in *gold* coin and bullion
Netherlands	GS	*40%* of the Netherlands Bank's banknotes, bank assignations, and current accounts to be held in *coin and bullion*
New Zealand	GS	Reserve holdings of note-issuing banks governed by their own charters
Norway	GS	100% of the Bank of Norway's banknotes in excess of 250 million kroner (fiduciary issue) to be held in *gold* coin and bullion
Paraguay	GS	Exchange office not required to hold any fixed gold reserve
Peru	GES	*50%* of the Reserve Bank's note circulation and sight liabilities to be held in *(a) gold* coin and bars or *(b) balances in dollars* in New York or in *sterling* in London
Poland	GES	*40%* of the Bank of Poland's note circulation and sight liabilities to be held in *(a) gold* coin and bullion (at least 75% of total cover), *(b) silver* at its gold value (up to 5% of the gold holdings), or *(c) 30-day balances*, checks, and bills of exchange accepted or endorsed by foreign banks *in currencies convertible into gold*
Romania	GES	*35%* of the National Bank's sight liabilities to be held in *(a) gold* (at least 25% of the Bank's sight liabilities) or *(b) foreign exchange*
Salvador	GS	*40%* of the banks of issue's notes and 20% of their sight deposits to be held in US gold coin
Siam (Thailand)	GES	*100%* of the government's note circulation to be held in *(a) gold*, *(b) gold securities*, *c) cash* placed at call or short notice at any bank approved by the Minister and payable *in the currency of a gold standard country*, or *d) Baht coins*
South Africa	GS	*40%* of the Reserve Bank's note circulation to be held in *gold* or *gold specie*
Sweden	GS	*50%* of the Riksbank's note circulation in excess of 250 million kronor (fiduciary issue) to be held in *gold* coin or bullion
Switzerland	GS	*40%* of the National Bank's note circulation to be held in *gold* coin or bullion
United Kingdom	GS	*100%* of the Bank of England's note circulation in excess of 260 million pounds (fiduciary issue) to be held in gold coin or bullion
United States	GS	*40%* of the Federal Reserve banks' notes and *35%* of their deposits to be held in *gold* or lawful money
Venezuela	GS	*33.3%* of the banks of issue's to be held in *gold* coin
Yugoslavia	GES	*33.3%* of the National Bank's note circulation to be held in *(a) gold*, *(b) silver*, or *(c) deposits in foreign countries*

Source: League of Nations (1930), *Legislation on Gold*
GS, gold standard; *GES*, gold exchange standard
[a]The monetary law foresaw a rise in the reserve ratio to 28% in January 1933 and to 33.3% in January 1938
[b]The monetary law foresaw a rise in the reserve ratio to 30% at the end of 1930 and 35% at the end of 1935

Nearly half of the countries in the world adopted the gold exchange standard. It is however difficult to identify a clear pattern in its adoption. In Europe, Albania, Austria, Belgium, Bulgaria, Czechoslovakia, Danzig, Denmark, Estonia, Finland, Germany, Greece, Hungary, Iceland, Italy, Latvia, Poland, Romania, and Yugoslavia were on the gold exchange standard, while Lithuania, the Netherlands, Norway, Sweden, and Switzerland only included gold or silver in their legal reserves. In Latin America, Bolivia, Chile, Colombia, Ecuador, Guatemala, and Peru allowed their central banks to hold foreign exchange in their legal reserves, whereas Costa Rica, Cuba, Curaçao, Mexico, Paraguay, Salvador, Uruguay, and Venezuela remained on a pure gold standard. Countries which were – or aspired to be – gold centers such as the United States, the United Kingdom, or France were also all legally on a strict gold standard.

Central banks of countries that adopted the gold exchange standard were allowed to replace their unproductive gold reserves with interest-bearing foreign assets, and this boosted their profitability – a far from negligible benefit as many of them were privately held institutions accountable to their shareholders. At the same time, they were subject to both currency and credit risk on their foreign holdings. The monetary law usually attempted to limit these risks through various provisions. In Albania, Belgium, Germany, Guatemala, and Poland, central banks could only hold a fraction of their cover reserves in international currencies, while the rest had to be held in gold. This fraction was set to only one fourth in Belgium, Germany, and Poland and to two thirds in Albania and Guatemala. In Austria and Hungary, legislation required that foreign reserves covering note circulation be held in currencies not subject to "violent" or "extraordinary" fluctuations (League of Nations 1930, pp. 159 and 250). In Belgium, Bulgaria, Estonia, Greece, Italy, and Poland, international reserves had to be denominated in currencies convertible into gold, while the central banks of Bolivia, Chile, Ecuador, and Guatemala could only hold foreign deposits "payable in gold" on demand. Finally, in order to limit credit risk, governments often restricted the range of assets used for covering monetary circulation to short-term bank deposits with reputable foreign banks, as well as 3-month bills of exchange or Treasury bills. In Hungary, the central bank's foreign deposits had to be held with "institutes of unquestionable solvency in the leading money markets of Europe and America" (League of Nations 1930, p. 250). In Bolivia, Chile, Ecuador, Guatemala, and Peru, they could only be placed with first-class banks in New York or London.

The Significance of Foreign Exchange Reserves

The practice of holding reserves in foreign centers was not brand new in the interwar years, and central banks in the nineteenth century already held a significant proportion of their external reserves in the form of foreign exchange (Lindert 1969). One important example is the National Bank of Belgium which held a large foreign portfolio in the 1850s (Ugolini 2012). Based on evidence for 35 countries, Lindert (1969) estimates that around 19% of central banks' international reserves consisted of foreign currency assets on the eve of the First World War. The widespread

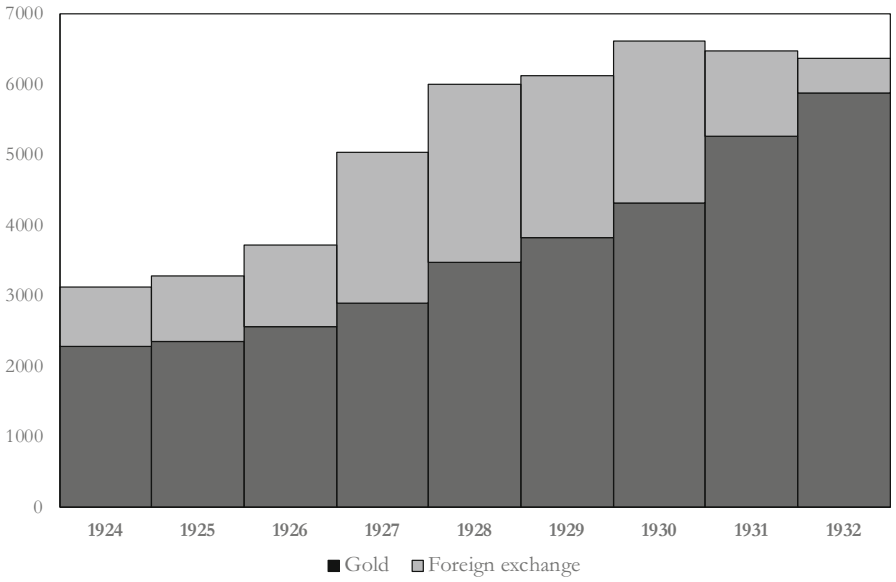


Fig. 3 Gold and foreign exchange reserves of 24 European countries, 1924–1932. (Source: Nurkse 1944)

adoption of the gold exchange standard however strongly reinforced this practice (Nurkse 1944).

Figure 3 displays the total holdings of gold and foreign exchange reserves of 24 European countries from 1924 to 1932. The proportion of central banks' reserves held in foreign exchange increased in the second half of the 1920s and reached 42% in 1927. Central banks often held much more international reserves than they were legally bound to in order to cover their note circulation (Nurkse 1944, p. 32). Such international reserves held in excess of the legal cover ratio were not subject to legal restrictions, and a large part of them was in the form of foreign exchange.

Table 2 reports the share of foreign exchange reserves in total international reserves for selected European countries from 1924 to 1932. Central European countries that had experienced hyperinflation were among those that exhibited the largest proportions of foreign exchange reserves in their total reserves. In 1924, this proportion was as high as 97% in Austria and reached 83 and 63% in, respectively, Hungary and Germany. However, while Austria maintained a high share of foreign reserves until the year 1930, both Hungary and Germany significantly reduced their holdings of foreign exchange in 1925–1928. Even in countries not legally on the gold exchange standard, central banks sometimes held substantial amounts of foreign currency reserves (Nurkse 1944, p. 31). For example, around one half of the Swedish Riksbank's international reserves were held in foreign currencies in 1925–1929, whereas Norway, the Netherlands, and Switzerland held approximately one third of their international reserves in foreign exchange. In France, the monetary law of June 1928 only allowed the central bank to cover its banknotes with gold.

Table 2 Share of foreign exchange reserves in central banks' total international reserves (in %)

Year	1924	1925	1926	1927	1928	1929	1930	1931	1932
Central and Eastern Europe									
Austria	97.1	97.5	93.2	89.7	81.0	79.5	79.0	42.6	27.6
Bulgaria	46.7	33.3	38.5	47.1	66.7	44.4	35.3	15.4	8.3
Czechoslovakia	42.6	57.1	69.7	70.6	68.5	64.8	61.0	38.8	37.0
Germany	63.1	45.8	34.5	20.3	16.2	26.3	25.6	23.7	23.9
Hungary	83.3	85.9	58.9	51.4	32.7	32.6	29.3	18.2	15.0
Poland	71.0	3.7	47.1	63.3	53.3	42.8	42.2	26.4	21.1
Yugoslavia	84.1	82.6	79.3	79.8	71.4	74.3	54.8	20.5	11.4
Nordic countries									
Denmark	18.8	30.0	11.1	34.7	40.3	34.3	37.0	14.3	9.8
Finland	63.6	79.5	77.1	80.0	70.4	68.0	75.0	51.9	50.1
Norway	26.4	32.8	38.1	29.1	22.0	31.6	32.8	12.5	17.0
Sweden	36.0	46.6	48.3	53.0	47.9	51.8	61.8	19.1	50.9
Southern Europe									
Greece	75.0	69.8	69.6	70.8	87.3	80.0	82.1	56.0	61.9
Portugal	62.1	62.1	52.6	47.1	64.0	65.4	50.0	61.8	45.5
Spain	1.4	1.2	1.4	1.4	3.5	3.7	3.3	11.9	10.9
Other continental Europe									
Belgium	10.3	10.2	41.9	42.2	38.5	34.3	41.4	0.0	0.0
France	1.9	1.8	14.0	47.1	50.6	38.5	32.8	23.8	5.1
Italy	8.7	22.5	41.2	62.5	54.4	49.8	45.0	27.8	18.4
Netherlands	18.1	35.7	31.1	29.4	33.5	32.8	36.7	8.9	6.5
Switzerland	27.4	32.3	32.1	27.5	32.2	37.2	38.1	4.2	3.4

Source: Nurkse (1944), *International Currency Experience: Lessons from the Interwar Period*

However, the Bank of France had accumulated large amounts of foreign exchange reserves following the stabilization of the franc in 1926–1928. When gold convertibility was restored, the Bank held much more international reserves than it was legally required to in order to cover monetary circulation, and more than half of these reserves were in foreign exchange (Mouré 2002; Accominotti 2009). As the largest holder of foreign currency reserves worldwide, France was de facto – if not de jure – on a gold exchange standard.

Relying on archival records for 16 different countries for the year 1929, Eichengreen and Flandreau (2009) have estimated that 97 percent of central banks' foreign exchange reserves were denominated in only two currencies, the pound sterling and US dollar, and that the latter overtook the former around 1924–1926. Mlynarski's (1929) fears of an excessive concentration of foreign exchange reserves in a few gold centers were therefore proven correct. Doubts about the two international reserve currencies at the beginning of the 1930s eventually undermined the gold exchange standard and shackled the international monetary system.

Central banks' foreign currency reserves also constituted an important component of the international credit architecture of the interwar years. A large share of these

reserves was held in the form of sight deposits with London and New York banks, and they provided a significant source of funding for these banks. Central banks also held short-term sterling and dollar bills, especially Treasury bills and bankers' acceptances – an instrument which was used for the purpose of financing international trade. Through the holdings of such instruments, central banks were important providers of resources for the financing of international trade at the end of the 1920s.

Exchange Rate Parities

By the end of the 1920s, the restoration of the gold standard had given way to a global fixed exchange rate system. Bilateral exchange rate parities between the different currencies were the outcome of the various governments' decisions to stabilize their currency's gold price at a certain level, itself chosen for specific economic, financial, and political reasons. In his memorandum prepared for the 1920 Brussels conference, Gustav Cassel noted that "there [was] no reason to believe that exchanges would ever be restored, generally, to their old parities" and that "every country should decide . . . what internal value it is going to give to its money" (cited by Siepmann 1920, p. 440). However, given the weak coordination across countries in returning to the gold standard, there was also no reason to believe that the new exchange rate parities that would emerge from the process of monetary reconstruction would correspond to fundamental equilibrium exchange rates.

Throughout its existence, the interwar gold exchange standard was undermined by misaligned exchange rate parities. In theory, such deviations of nominal exchange rates from their equilibrium values should only have been short-lived under a gold standard system. According to the so-called *price specie flow model*, a country whose currency was overvalued (undervalued) should have experienced trade deficits (surpluses) and reserves outflows (inflows). Changes in reserves should in turn have led to deflation (inflation) until real exchange rate equilibrium was restored (Hume 1777). However, governments' reluctance to engage in deflationary or inflationary policies as well as the stickiness of nominal prices and wages prevented this adjustment from taking place smoothly, and real exchange rate disequilibria, rather than being temporary, lasted persistently.

On the one hand, the UK government was criticized for restoring a gold parity which overvalued sterling relative to the US dollar and French franc. On the other hand, contemporaries reproached the French government's stabilization of the franc at an undervalued exchange rate, a decision which contributed to the pound's difficulties. Even among countries that did not restore pre-war gold parities, overvaluation could be a problem. For example, the Italian stabilization of 1927 left the lire overvalued relative to both sterling and the US dollar, even though the currency had been devalued by 73% compared to its pre-war level. It would however be misleading to blame the misfunctioning of the gold exchange standard system on individual countries' mistakes in the choice of their exchange rate parity. Governments choosing the level at which to stabilize their currency did so in response to a set of economic, financial, and political constraints. Real exchange

rate disequilibria, therefore, mostly arose from the absence of any binding mechanism to induce coordination in the setting of the gold parities.

The Collapse of the Gold Exchange Standard

The Capital Boom and Sudden Stop

The early years of operation of the gold exchange standard were characterized by a boom in world trade and international financial integration. After the monetary chaos of the postwar years, a period of relative financial stability opened up in 1924–1928, and the world economy expanded. Capital started flowing from a few creditor countries to a large number of borrowing countries especially in Central Europe and Latin America (Royal Institute of International Affairs 1937; Feinstein and Watson 1995). Governments and corporations issued long-term bonds in the main international financial centers, especially New York and London (Accominotti and Eichengreen 2016). At the same time, firms could borrow short term to finance their international trading activities by drawing bills on London and New York City. Germany ran trade deficits for most years in 1924–1928 and paid reparations, making for large current account deficits, which were mostly financed through inflows of private capital. Capital also flowed to Austria and Hungary following their exchange rate stabilization, and Latin American countries were also large recipients of foreign investments.

This era of expansion was however short-lived. In 1928–1929 already, a decline in capital inflows marked the beginning of the difficulties for the capital-importing countries. This "sudden stop" continued following the US stock market crash of October 1929 until private international capital flows reversed in the year 1931. Debtor countries suddenly needed to reduce their current account deficits, and doing so while remaining on the gold standard required engaging into internal deflation.

The Crisis in Central Europe

The first countries to leave gold following the onset of the Great Depression were producers of primary goods hit by both the decline in commodity prices and the drying up of foreign lending. In December 1929, Australia, Argentina, and Brazil all severely restricted gold convertibility and effectively suspended the gold standard (Eichengreen 1992, pp. 232–241).

Central European countries were also severely hit by the sudden stop due to their dependence on foreign borrowing. In the spring and summer of 1931, Austria, Germany, and Hungary were shaken by severe financial crises. In May 1931, Austria's largest commercial bank, the Credit-Anstalt, announced significant losses triggering a bank run and government bailout (Schubert 1991). A banking and currency crisis ensued ultimately leading to the devaluation of the shilling and the

imposition of capital controls in October. A few days after the failure of the Credit-Anstalt, banking troubles propagated to Hungary. Both countries experienced typical emerging market "twin" crises in which banking problems were concomitant with pressure on the currency. These crises even became "triplet" ones as fiscal difficulties were added to the foreign exchange and banking problems when governments intervened to rescue the banking system (James 2001, p. 53).

In June 1931, the financial crisis reached Germany (Ferguson and Temin 2003; Schnabel 2004). Domestic depositors and foreign creditors started running on the country's large commercial banks as well as on the German mark. Following the failure of the Darmstaedter and National Bank, the German government severely restricted foreign payments on 15 July. This involved a ban on foreign exchange market transactions and a suspension of the mark's convertibility. Although the German currency's gold parity was officially maintained, the absence of a free foreign exchange market had transformed the gold standard into a mere fiction.

The Fall of Sterling

The difficulties soon spread from Central Europe to Britain. Ever since its return to the gold standard, the pound sterling had displayed significant weaknesses on the foreign exchange market, especially in 1928–1931 (Accominotti 2009). The introduction of capital controls in Germany in mid-July 1931 was immediately followed by a run on the Bank of England's gold reserves and expectations of a pound's devaluation mounted. The Bank of England received international assistance from the Federal Reserve Bank of New York and Bank of France in order to support sterling on the foreign exchange market and raised its interest rate – although only moderately. In a last attempt to save the pound, a national unity government was formed at the end of August under the lead of Ramsay MacDonald with the aim of balancing the budget. However, this proved insufficient to reverse investors' expectations and stop the run on the British currency. On 21 September 1931, the UK government amended the Gold Standard Act of 1925 and suspended sterling's gold convertibility. The pound immediately depreciated on the foreign exchange market.

Several factors contributed to the pound's collapse. Following the restoration of the pre-war gold parity in 1925, the British government proved reluctant to engage into a policy of price deflation and wage reduction in order to overturn the pound's overvaluation (Moggridge 1972). The difficulties at balancing the budget over the years 1925–1931 also resulted in fiscal deficits which were hardly compatible with an exchange rate peg (Sayers 1976). In addition, the UK suffered from an unemployment problem in the 1920s, and its persistence made the maintenance of the gold standard even more costly (Eichengreen and Jeanne 2000). Finally, British financial institutions known as merchant banks or acceptance houses were heavily invested in Germany due to their trade finance activities. Through the merchant banks, the German and Central European financial crisis directly transmitted to the London money market in mid-July 1931, and this contributed to further undermine the currency (Accominotti 2012, 2018).

The pound's devaluation triggered a shock wave in the international monetary system. From September 1931 to May 1932, Bolivia, Canada, Chile, Columbia, Denmark, Ecuador, Finland, Greece, Japan, New Zealand, Nicaragua, Norway, Peru, Salvador, and Sweden all officially suspended the gold standard (League of Nations 1934/35). Several of these countries had strong trade relationships with the United Kingdom, and the pound's devaluation affected their trade balance. In Sweden, currency problems had however started well before September 1931. The German financial crisis of the spring 1931 had triggered a run on Swedish commercial banks because they had indirectly lent to the German government through the Swedish financier Ivar Kreuger's financial conglomerate. The sterling crisis added to these difficulties until the Swedish government decided to suspend the gold standard (Straumann et al. 2017). The wave of devaluations of 1931–1932 resulted in a complete disintegration of the international monetary system.

The United States Off the Gold Standard

Speculative pressure also mounted on the US dollar as of September 1931. The US currency traded at a substantial discount on the forward exchange market in the fall and the Federal Reserve lost large amounts of gold reserves. US monetary authorities however resisted speculative pressure and raised interest rates in October 1931. This tightening of monetary policy avoided the United States an immediate devaluation but had severe domestic financial repercussions. In the following months, a severe banking panic hit the country leading to numerous bank failures (Richardson and van Horn 2018).

The election of Franklin D. Roosevelt as President in November 1932 changed the country's exchange rate policy. An advocate of economic interventionism, Roosevelt had however promised during his campaign that he would not use currency devaluation as a tool against economic recession and that the United States under his leadership would remain on the gold standard. However, in March 1933, the country experienced a nationwide run on bank deposits as well on the Federal Reserve's gold reserves (Wigmore 1987). The day after his inauguration, Roosevelt declared a bank holiday and imposed a ban on gold exports. While the US dollar remained stable in March, it depreciated sharply from April 19 onward until its stabilization in January 1934 at only 40% of its previous gold parity.

The Liquidation of Foreign Exchange Reserves

The crisis of the international monetary system of 1931–1933 was accompanied by the widespread abandonment of the gold exchange standard scheme which rested on central banks' holdings of foreign exchange reserves. At the end of the 1920s, fears over the sustainability of the pound had already led central banks to sell off part of their sterling reserves. Following the pound's devaluation of September 1931, several of them however made substantial losses on their remaining sterling assets.

As currency troubles propagated to the United States, central banks now started liquidating their dollar holdings and converted them into gold. The share of foreign currency reserves in European central banks' total international reserves fell from a high of 42% at the end of 1927 to 19% at the end of 1931 and only 8% at the end of 1932 (Nurkse 1944).

As the world's largest holder of foreign reserves, France played an important role in the demise of the gold exchange standard, and the large volume of its foreign holdings put it into a difficult predicament when confidence in the international monetary regime collapsed. As a privately held institution, the Bank of France was seriously concerned about currency risk on its foreign reserves and attempted to sell off its sterling assets when the pound showed signs of weakness in 1929–1930. However, monetary officials were confronted with the difficulty of getting rid of such a large sterling portfolio. When the pound's difficulties intensified at the end of 1930, the Bank of France had to renounce to its policy of sterling liquidation in order to avoid further destabilizing the British currency on the foreign exchange market and increasing the risk of an exchange loss for itself (Accominotti 2009). The Bank eventually made a considerable loss when sterling was devalued in September 1931 and had to be bailed out at a high cost by the French government. To avoid further losses, French monetary officials subsequently decided to get rid of their dollar holdings and converted them all into gold (Mouré 2002; Accominotti 2009). This, in turn, put intense pressure on the gold exchange standard system.

The risks associated with a gold exchange standard system – as described by Mlynarski (1929) – had therefore materialized at the beginning of the 1930s. Central banks participating in this system responded to the weaknesses of the two main international currencies by liquidating their foreign reserves, and, in so doing, they weakened these currencies further. In addition, the unloading of official sterling and dollar assets contributed to reduce the funding available on the London and New York money markets at a critical time. Designed as a tool to prevent global deflation, the gold exchange standard eventually became a destabilizing factor for the global monetary system when the credibility of reserves currencies was put into question.

The Collapse of the Gold Bloc

By mid-1933, only a few countries had resisted devaluing their currencies: France, Belgium, Italy, the Netherlands, Poland, and Switzerland. At the London World Economic Conference of June–July 1933, representatives of these countries made a joint statement confirming "their intention to maintain the free functioning of the gold standard in their respective countries at the existing gold parities and within the framework of existing monetary laws" (The Economist, July 8th 1933, p. 64). They formed the "gold bloc."

The economies of the gold bloc countries contracted severely over subsequent years as a consequence of exchange rate appreciation, and their strong commitment to the gold standard soon became difficult to honor. Italy was the first to defect when its government introduced foreign exchange restrictions in May 1934. In March

1935, Belgium suspended the gold standard following a banking and governmental crisis (Straumann 2010, p. 163). The Belgian devaluation immediately triggered a run on the Bank of France's gold reserves. France initially resisted the speculative attack, but the accession of the left-wing Popular Front coalition to power in May 1936 resulted in further pressure on the franc, which was eventually devalued in September. Immediately afterwards, Switzerland and the Netherlands also suspended gold convertibility. At the end of 1936, the interwar gold standard system was dead and buried.

The Gold Exchange Standard and the Great Depression

The Gold Standard and Global Deflation

Between 1929 and 1932, when the interwar gold standard was in operation, the world economy experienced the most severe price deflation and economic recession of the modern era. The question of the link between the gold standard, global deflation, and the Great Depression has attracted considerable attention from scholars.

There are several reasons why deflation and depression were closely related. Due to the slow adjustment of nominal wages, the decline in consumer prices resulted in an increase in real wages which depressed aggregate supply (Bernanke and Carey 1996). Deflation also contributed to raise the real value of nominal debts, a mechanism known as debt deflation. This led to more debt defaults and banking fragilities (Fisher 1933; Bernanke and James 1991). Researchers looking for causes to the Great Depression have therefore principally focused on explaining the spectacular decline in prices that took place in that period. In so doing, they have singled out the international monetary regime as an important culprit.

Several authors have described the link between the gold standard and global deflation as almost mechanical. For example, Mazumder and Wood (2013) argue that the restoration of pre-WW1 gold parities in the 1920s was bound to lead to the "Great Deflation" of 1929–1932. In a standard closed-economy model such as that developed by Barro (1979), the price level in a gold standard regime corresponds in equilibrium to the cost of producing gold relative to other commodities. If the official gold parity undervalues gold, its production will be discouraged, and the demand for gold for other purposes than monetary use will increase. This will result in a decline in central banks' gold reserves leading to a contraction in the money supply and deflation. Because of the dramatic increase in wholesale and consumer prices during the First World War and postwar years, a generalized return to pre-war gold parities would have undervalued gold and ineluctably led to global deflation (Mazumder and Wood 2013).

However, not all countries had returned to the gold standard at their pre-war gold parities in the 1920s. In addition, the dynamics of reserves and prices under the interwar gold exchange standard system did not directly follow that of the textbook

model because several central banks could complement their gold reserves with foreign currency assets and expand the money supply accordingly. As long as real exchange rates did not depart too much from equilibrium, global deflation was not inevitable after the restoration of the interwar gold standard. However, avoiding it would have required a much higher degree of international coordination in the choice of exchange rate parities and, subsequently, in the implementation of monetary policy (Dabrowski 2015).

Credibility and Cooperation

Eichengreen (1992) points toward two main weaknesses in the interwar gold standard that explain its failure. The monetary regime suffered from its inception from a lack of *credibility* and from weak *cooperation* between central banks. Due to the progress of enfranchisement, labor unions, and political parties representing the interests of the working class, governments of the interwar years faced increased pressure to use fiscal and monetary policy in order to fight against unemployment and attached a lower weight to fixed exchange rates in their objective function (Eichengreen 1992, p. 31). Investors understood this well and were prompt to trigger speculative attacks on currencies at the first signs of weakness. In this low-credibility environment, central banks were reluctant to engage in monetary expansion because they wanted to maintain confidence in their currency. They also competed actively with each other for gold reserves through raising interest rates. Political tensions and the persistence of the reparations problem undermined any attempt at international monetary cooperation. The outcome was a set of excessively restrictive monetary policies and a "deflationary bias" responsible for the Great Depression (Eichengreen 1984).

This analysis suggests that monetary contraction and deflation did not simply arise from individual governments' or central banks' monetary policy mistakes but were the consequence of the interaction between countries in a poorly designed global monetary system. Of course, deflation might have been avoided if central banks had behaved more cooperatively. However, one might doubt whether this is a realistic counterfactual. Flandreau (1997) has argued that cooperation between central banks was equally absent in the pre-WW1 and interwar periods. Central banks and governments responded to existing incentives, and there were few or no mechanisms that incentivized cooperation in the international monetary system. In many dimensions, the interaction between central banks under the interwar gold standard can be described as a prisoners' dilemma and be analyzed through the tools of game theory. While all countries would have achieved a better economic outcome if they had engaged into more cooperation, each of them also had strong incentives to defect. In the absence of a mechanism to prevent defection, cooperation was not an equilibrium outcome in the game played by central banks under the gold exchange standard, and one should not have expected them to cooperate.

The Role of France

The case of France illustrates these issues perfectly well. The Bank of France has been described as the ultimate example of a noncooperative player in the interwar monetary regime. Following the restoration of the franc's gold convertibility in June 1928, the Bank of France accumulated large amounts of gold reserves. France's gold accumulation attracted much criticism from contemporaries abroad who described the country as a "gold sink." Einzig (1932) accused French monetary authorities of destabilizing the pound through selling sterling reserves on the foreign exchange market. Johnson (1997) attributes the French policy of gold accumulation to a series of idiosyncratic and almost deliberate "errors" which largely contributed to global deflation. More recently, Irwin (2011) has argued that France – together with the United States – withdrew a large share of the world's monetary gold from circulation and that the French and US mismanagement of the gold standard and failure to monetize gold inflows were responsible "for nearly 40 percent of the worldwide deflation experienced between 1929 and 1931" (Irwin 2011, pp. 37–38).

The main reproach made to France in this context is not the gold accumulation per se but the fact that its gold inflows were "sterilized." Sterilization policy consists in offsetting the effects of changes in the central bank's international reserves on the domestic money supply. When a central bank accumulates (loses) international reserves, the domestic money supply automatically increases (decreases) in the same proportion. A central bank can however sterilize the effect of reserves changes on the money supply by selling (buying) domestic assets – for example, government bonds – on the open market. In 1919, the UK Cunliffe Committee on Currency and Foreign Exchanges argued that the smooth adjustment of balance of payments under a gold standard system required central banks to abstain from sterilizing movements in their international reserves. Countries that did not engage in sterilization were said to follow "the rules of the game" of the gold standard. However, it appears that central banks never integrated these alleged rules. Nurkse (1944) and Bloomfield (1959) have demonstrated that the "rules of the game" were in fact rarely followed in either the pre-WW1 or interwar periods.

Figure 4 displays the ratios of the Bank of France's international reserves (defined as the sum of gold and foreign exchange reserves) to its domestic liabilities (defined as the sum of the note circulation and sight deposits with the central bank) at the end of each year from 1928 to 1932. It is apparent that the Bank of France did not actively sterilize reserves inflows as the ratio of its international reserves to domestic liabilities remained fairly constant over those years. This means that the Bank did not attempt to counteract the effect of international reserves inflows on its monetary circulation through the sale of domestic assets. Had the rules of the game of the gold exchange standard ever existed, one might say that the Bank of France closely followed them (Bernanke and Mihov 2000).

However, the ratio of the Bank's gold reserves to domestic liabilities increased considerably over those years due to a massive rebalancing of its reserves in favor of gold and at the expense of foreign exchange. These movements *within* the international reserves had no effect on the domestic money supply and could therefore not be sterilized. Nevertheless, France's massive liquidation of its foreign exchange

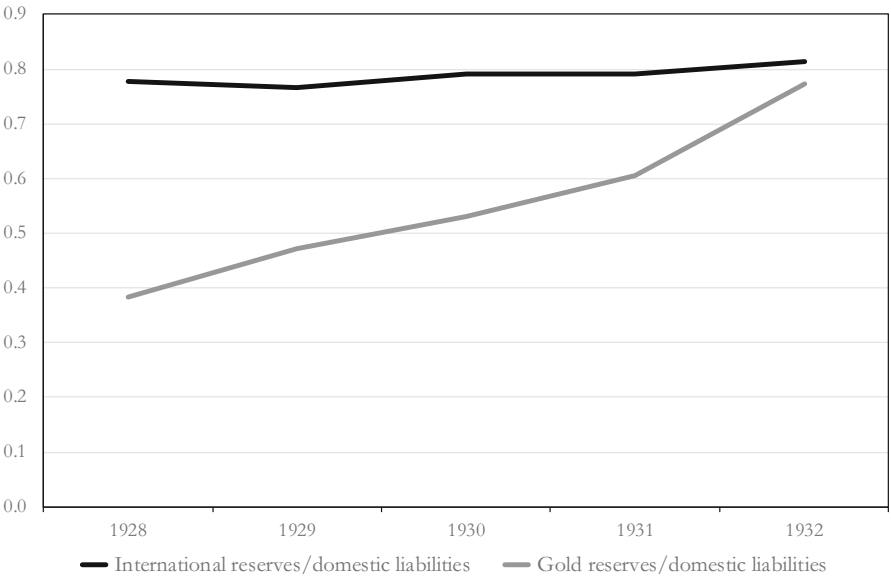


Fig. 4 Bank of France's ratio of international reserves to domestic liabilities. (Source: Board of Governors of the Federal Reserve (1943). International reserves are defined as the sum of gold and foreign exchange reserves. Domestic liabilities are defined as the sum of note circulation and deposits with the Bank of France)

reserves did contribute to destabilize the international monetary system because it undermined the functioning of the gold exchange standard mechanism and put pressure on its main international currencies (Bernanke and Mihov 2000). In this respect, the Bank of France's reserves policy was similar to that of several other central banks that implemented a gold exchange standard in the 1920s but decided to get rid of their foreign reserves in order to minimize the risks of exchange loss when doubts arose about the future of the two reserve currencies (Accominotti 2009). This was a negative-sum game, and a joint effort from countries to support international currencies would probably have yielded a better outcome for each of them. However, given the low credibility of the gold exchange standard system, central banks lacked the proper incentives to engage in such cooperation, and their combined actions led to the disappearance of the gold exchange standard system. Therefore, there was nothing idiosyncratic or accidental in the Bank of France's reserves policy of 1928–1931, but its behavior was predictable and representative of the noncooperative equilibrium that prevailed in the international monetary system.

The Persistence of the Gold Standard

Given the apparent problems of the interwar gold exchange standard regime, one might wonder why so many countries longed to join this regime and why they did not abandon it earlier when the world economy slid into depression at the beginning

of the 1930s. There is ample empirical evidence that countries which devalued early in the 1930s managed to escape deflation and recover more quickly than others from the Great Depression (Choudhri and Kochin 1980; Eichengreen and Sachs 1985). Yet, many countries remained faithful to the gold standard for several years after the beginning of the global economic slump. The abandonment of the gold standard was usually not a deliberate decision but came as a consequence of a currency crisis. There are several factors that explain the occurrence of such crises including the presence of banking problems, the size of the central bank's international reserves, the type of political regime each country adopted, as well as the intensity of trade links with other devaluing countries (Wandschneider 2008; Wolf 2008).

One of the reasons advanced for countries' persistent willingness to be on the gold standard is ideology. Eichengreen and Temin (2000) argue that interwar policymakers were prey to a "gold standard mentalité"; their intellectual framework prevented them from considering any other way of organizing the creation of money. However, for many countries, there might also have been an economic rationale behind the decision to join and, then, remain on the gold standard. In particular, small countries at the periphery of the international monetary system as well as those with a history of hyperinflation often have little autonomy in the choice of their exchange rate regime. The dependence on international trade or foreign currency borrowing as well as the need to build credibility in the domestic currency provided strong incentives to be on the gold standard and maintain fixed exchange rates with the major creditor countries' currencies.

From that perspective, the monetary experience of several countries in the interwar years was not fundamentally different from that of developing and emerging countries in the post-1980 era whose exchange rate policy has often been characterized by a "fear of floating" (Calvo and Reinhart 2002). Interestingly, even countries that enjoyed the benefits of flexible exchange rates up to the early 1930s such as Japan, Spain, and Portugal eventually attempted to restore the gold standard – with or without success – in 1930–1931, at a time when the depression was already underway.

Conclusion

This chapter has reviewed the main developments in the international monetary system of the interwar years, from the emergence of the gold exchange standard system in the 1920s up to its collapse in the 1930s.

The interwar gold exchange standard was fundamental in shaping economists' views about money and exchange rates, and it influenced both the theory and practice of monetary policy. Scholars of various times have however retained different lessons from this episode. Economists of the 1940s such as Ragnar Nurkse (1944) thought that the regime's main failure consisted in its inability to bring back exchange rate stability following the dramatic experience with floating exchange rates of the 1920s. When working on the design of new monetary institutions at the end of the Second World War, representatives of the victor countries at the Bretton

Woods Conference wished to avoid at any price a repeat of the exchange rate instability and competitive devaluations of the 1930s. They opted to reconstruct a global fixed exchange rate system but combined it with restrictions on cross-border capital flows in order to give countries more monetary policy autonomy.

By contrast, when revisiting the interwar monetary experience, many scholars of the 1980s and 1990s have reinterpreted it as evidence that fixed exchange rates regimes are fundamentally flawed. The interwar gold exchange standard was often invoked in the policy debate surrounding the creation of the Euro, and its poor record led many economists to be skeptical toward this project.

Drawing policy lessons from the gold exchange standard era is however more difficult than it appears at first sight. The literature of the past 40 years has demonstrated that the dramatic monetary contraction of the 1930s cannot be attributed to individual countries' mistakes in the conduct of their monetary policies but that those policies themselves responded to the existing incentives and constraints central banks faced in the global monetary system. International monetary regimes are the outcome of long historical processes, and institutions producing the right incentives for countries to coordinate their policies do not emerge easily, as the recent crisis in the eurozone has shown.

Cross-References

- ▶ Deflations in History
- ▶ International Monetary Regimes: The Gold Standard
- ▶ The Evolution of US Monetary Policy
- ▶ The Evolution of Monetary Policy (Goals and Targets) in Western Europe
- ▶ The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy

References

Accominotti O (2009) The sterling trap: foreign reserves management at the Bank of France, 1928–1936. Eur Rev Econ Hist 13:349–376

Accominotti O (2012) London merchant banks, the Central European panic and the sterling crisis of 1931. J Econ Hist 72:1–43

Accominotti O (2018) International banking and transmission of the 1931 financial crisis. Econ Hist Rev (forthcoming)

Accominotti O, Chambers D (2016) If you're so smart: John Maynard Keynes and currency speculation in the interwar years. J Econ Hist 76:342–386

Accominotti O, Eichengreen B (2016) The mother of all sudden stops: capital flows and reversals in Europe, 1919–1932. Econ Hist Rev 69:469–492

Arthmar R, McLure M (2016) On Britain's return to the gold standard: was there a 'Pigou–McKenna School'? Econ Rec 92:1–14

Atkin J (2005) The foreign exchange market of London. Routledge, New York

Balderston T (1989) War finance and inflation in Britain and Germany, 1914–1918. Econ Hist Rev 42:222–244
Barro R (1979) Money and the price level under the gold standard. Econ J 89:13–33
Bernanke B, Carey K (1996) Nominal wage stickiness and aggregate supply in the Great Depression. Q J Econ 111:853–883
Bernanke B, James H (1991) The gold standard, deflation, and financial crisis in the Great Depression: an international comparison. In: Glenn Hubbard R (ed) Financial markets and financial crises. University of Chicago Press, Chicago
Bernanke B, Mihov I (2000) Deflation and monetary contraction in the Great Depression: an analysis by simple ratios. In: Bernanke B (ed) Essays on the Great Depression. Princeton University Press, Princeton, pp 108–160
Bloomfield A (1959) Monetary policy under the international gold standard, 1880–1914. Federal Reserve Bank of New York, New York
Board of Governors of the Federal Reserve (1943) Banking and monetary statistics, 1914–1941. Federal Reserve System, Washington, DC
Calvo G, Reinhart C (2002) Fear of floating. Q J Econ 117:379–408
Cassel G (1920) Further observations on the world's monetary problems. Econ J 30:39–45
Choudhri E, Kochin L (1980) The exchange rate and the international transmission of business cycle disturbances: some evidence from the Great Depression. J Money Credit Bank 12:565–574
Cohen J (1972) The 1927 revaluation of the lira: a study in political economy. Econ Hist Rev 25:642–654
Crabbe L (1989) The international gold standard and U.S. monetary policy from World War I to the new deal. Fed Reserv Bull, pp 423–440
Crafts N, Fearon P (2010) Lessons from the 1930s Great Depression. Oxf Rev Econ Policy 26:285–317
Dabrowski M (2015) Did the Great Deflation really have to happen? A reconsideration of the inevitability of the Great Deflation view. Eur Rev Econ Hist 19:235–254
Eichengreen B (1984) Central Bank Cooperation under the interwar gold standard. Explor Econ Hist 21:64–87
Eichengreen B (1992) Golden fetters: the gold standard and the Great Depression, 1919–1939. Oxford University Press, Oxford
Eichengreen B, Flandreau M (2009) The rise and fall of the dollar (or when did the dollar replace sterling as the leading reserve currency. Eur Rev Econ Hist 13:377–411
Eichengreen B, Jeanne O (2000) Currency crisis and unemployment: Sterling in 1931. In: Krugman P (ed) Currency crises. Chicago University Press, Chicago, pp 7–46
Eichengreen B, Sachs JD (1985) Exchange rates and economic recovery in the 1930s. J Econ Hist 45:925–946
Eichengreen B, Temin P (2000) The gold standard and the Great Depression. Contemp Eur Hist 9:183–207
Einzig P (1932) Behind the scenes of international finance. Macmillan, London
Einzig P (1937) The theory of forward exchange. Macmillan, London
Feinstein CH, Watson K (1995) Private international capital flows in Europe in the inter-war period. In: Feinstein CH (ed) Banking, currency, and finance in Europe between the wars. Oxford University Press, Oxford, pp 94–130
Ferguson T, Temin P (2003) Made in Germany: the German currency crisis of July 1931. Res Econ Hist 21:1–53
Financial Times (London, UK), various issues
Fisher I (1933) The debt-deflation theory of Great Depressions. Econometrica 1:337–357
Flandreau M (1997) Central bank cooperation in historical perspective: a sceptical view. Econ Hist Rev 50:735–763
Friedman M (1992) Franklin D. Roosevelt, silver and China. J Polit Econ 100:62–83
Garber PM, Spencer MG (1994) The dissolution of the Austro-Hungarian Empire: lessons for currency reform. Essays in international finance. Princeton University, Princeton, p 119

Hall GJ (2004) Exchange rates and casualties during the First World War. J Monet Econ 51:1711–1742
Hawtrey RG (1919) The gold standard. Econ J 29:428–442
Hawtrey RG (1922) The Genoa resolutions on currency. Econ J 32(127):290–304
Holtfrerich C-L (1986) The German inflation, 1914–1923. Causes and effects in international perspective. Walter de Gruyter, New York
Hume D (1777) On the balance of trade. In: Essays, moral, political, and literary. Longmans, Green, London. Available at http://oll.libertyfund.org/titles/hume-essays-moral-political-literary-lf-ed
Irwin DA (2011) The French gold sink and the great deflation of 1929–1932. Cato papers on public policy, Cato Institute, Washington DC, 2
James H (2001) The end of globalization: lessons from the Great Depression. Harvard University Press, Cambridge, MA
Johnson C (1997) Gold, France and the Great Depression, 1919–1932. Yale University Press, New Haven
Jorge-Sotelo E (2018) The gold standard and the Great Depression in Spain. mimeo London School of Economics
Keynes JM (1925) The economic consequences of Mr. Churchill. L. and G. Woolf, London
League of Nations (1930) Legislation on gold. League of Nations, Geneva
League of Nations (1931) Second interim report of the gold delegation of the Financial Committee. League of Nations, Geneva
League of Nations (1934/35) Statistical year-book. League of Nations, Geneva
Lindert P (1969) Key currencies and Gold, 1900–1913. Princeton studies in international finance, Princeton University, Department of Economics, Princeton, New Jersey, vol 24
Matthews KGP (1986) Was Sterling overvalued in 1925? Econ Hist Rev 39:572–587
Mazumder S, Wood JH (2013) The great deflation of 1929–33: it (almost) had to happen. Econ Hist Rev 66:156–177
Metzler M (2006) Lever of Empire: the international gold standard and the crisis of liberalism in prewar Japan. University of California Press, Berkeley
Mlynarski F (1929) Gold and Central Banks. Macmillan, New York
Moggridge DE (1972) British monetary policy 1924–1931. Norman's conquest of $4.86. Cambridge University Press, Cambridge
Mouré KJ (2002) The gold standard illusion: France, the Bank of France and the International Gold Standard, 1914–1939. Oxford University Press, Oxford
Nurkse R (1944) International currency experiences: lessons from the inter-war period. League of Nations, Geneva
Prati A (1991) Poincare's stabilization: stopping a run on government debt. J Monet Econ 27(2): 213–239
Redmond J (1989) Was sterling overvalued in 1925? A comment. Econ Hist Rev 42:87–89
Richardson G, Van Horn P (2018) In the eye of a storm: Manhattan's money center banks during the international financial crisis of 1931. Explor Econ Hist 68:71–94
Royal Institute of International Affairs (1937) The problem of international investment. RIIA, London
Sargent T (1982) The ends of four big inflations. In: Hall R (ed) Inflation: causes and effects. University of Chicago Press, Chicago, pp 41–98
Sargent T (1984) Stopping moderate inflations: the methods of Poincare and Thatcher. In: Dornbush R, Simonsen MH (eds) Inflation, debt and indexation. MIT Press, Cambridge, MA
Sayers R (1976) The Bank of England, 1891–1944. Cambridge University Press, Cambridge
Schnabel I (2004) The German twin crisis of 1931. J Econ Hist 64:822–871
Schubert A (1991) The Credit-Anstalt crisis of 1931. Cambridge University Press, Cambridge
Shizume M (2012) The Japanese economy during the interwar period: instability in the financial system and the impact of the world depression. In: Ogren A, Oksendal LF (eds) The gold standard peripheries. Monetary policy, adjustment and flexibility in a global setting. Palgrave Macmillan, New York

Sicsic P (1993) Was the Franc Poincaré Undervalued? Explor Econ Hist 29:69–92
Siepmann HA (1920) The international conference at Brussels. Econ J 30:436–459
Straumann T (2010) Fixed ideas of money: small states and exchange rate regimes in twentieth century Europe. Cambridge University Press, Cambridge
Straumann T, Kugler P, Weber F (2017) How the German crisis of 1931 swept across Europe: a comparative view from Stockholm. Econ Hist Rev 70:224–247
Temin P (1989) Lessons from the Great Depression. MIT Press, Cambridge, MA
The Economist (London, UK), various issues
Ugolini S (2012) The origins of foreign exchange policy: the National Bank of Belgium and the quest for monetary independence in the 1850s. Eur Rev Econ Hist 16:51–73
Wandschneider K (2008) The stability of the interwar gold exchange standard: did politics matter? J Econ Hist 68:151–181
Wigmore BA (1987) Was the bank holiday of 1933 caused by a run on the dollar? J Econ Hist 47:739–755
Wolf N (2008) Scylla and Charybdis. Explaining Europe's exit from gold, January 1928–December 1936. Explor Econ Hist 45:383–401

International Monetary Regimes: The Bretton Woods System

26

Peter Kugler and Tobias Straumann

Contents

Abstract

The Bretton Woods system was the first attempt to create an international monetary arrangement with fixed exchange rates based on international cooperation of central banks and supervised by a newly created international institution, the International Monetary Fund (IMF). Negotiated in 1944 and effective from 1946, the system collapsed in 1971 because of two fundamental problems, the adjustment and the confidence problem. The adjustment problem resulted from the difficulty to reduce international imbalances by demand-management policies and changes of the par values. Once convertibility for current account transactions was restored in 1958, divergent economic policy preferences caused

P. Kugler
University of Basel, Basel, Switzerland
e-mail: peter.kugler@unibas.ch

T. Straumann (✉)
University of Zurich, Zürich, Switzerland
e-mail: tobias.straumann@econ.uzh.ch

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_25

centrifugal capital movements that proved impossible to contain over time. The confidence problem emerged because the Bretton Woods system, against the will of its architects, turned into a gold dollar standard as a result of the economic and political supremacy of the USA. As the dollar became the world reserve currency, fueled by sizeable US balance of payment deficits in the 1960s, holders of dollars became increasingly skeptical toward the stability of the dollar. Due to the threat of an imminent run on US gold reserves by foreign investors and central banks, President Richard Nixon suspended the gold convertibility of the dollar, thus effectively ending the basis of the Bretton Woods system.

Keywords
International monetary system · Bretton Woods system · Fixed exchange rates · Gold · US dollar

Introduction

The Bretton Woods system was the first attempt to create an international monetary arrangement with fixed exchange rates based on international cooperation of central banks and supervised by a newly created international institution, the International Monetary Fund (IMF). The system was based on 20 articles, which were agreed upon at the Bretton Woods conference in July 1944 held by the allied nations at war against the axis powers. It was designed in reaction to the disastrous economic development of the interwar years and aimed at reestablishing the exchange rate stability as in the gold standard without the deflationary adjustment pressure in deficit countries (Bordo 1993; Kugler 2016).

To achieve this goal, the Bretton Woods agreement introduced a set of rules which distinguished it from the gold standard. Central banks were required to intervene in the foreign exchange market. The IMF provided liquidity to member states struggling with a current account deficit. Free convertibility at a fixed rate was mandatory only for current transactions, while restrictions on capital movements were allowed. Devaluations were possible in the case of "fundamental disequilibrium" and after consultation with the IMF. By contrast, the classical gold standard saw no role for institutionalized central bank cooperation, was not supervised by an international authority, required full convertibility both for current account and capital account transactions, and prohibited devaluations.

Although the Bretton Woods system had fundamental flaws, it contributed to the recovery of the world economy after WWII. Paradoxically, one reason for its surprising resilience was that the 1944 Articles of Agreement were repeatedly violated by the member states. Most importantly, the transition to convertibility for current transactions was delayed until December 1958 which provided more room for domestic monetary and fiscal policies. Once convertibility was restored, the weakness of the system became quickly apparent, however. Three problems emerged: the adjustment process, the provision of liquidity, and the stability of the gold price of the dollar. As US monetary policy became expansionary after 1965,

eventually leading to international imbalances and inflation, the collapse of the Bretton Woods was only a matter of time. In 1971, US President Nixon closed the gold window, and in 1973, the adjustable peg system ceased to exist.

The chapter is organized as follows. First, we summarize the design of the Bretton Woods system according to the 1944 agreement (section "The Design of the Bretton Woods System"). Then, we give an account of the evolution of the system, covering the first years up to the establishment of convertibility from 1946 to 1958 (section "The Long Path to Convertibility, 1946–1958"), the period of convertibility from 1958 to 1968 (section "The Bretton Woods System Under Convertibility, 1958–1968"), and the closing phase between 1968 and 1971 (section "The Collapse of the Bretton Woods System, 1968–1971"). Finally, we deal with the position of the developing world within the Bretton Wood System (section "The Bretton Woods System and the Developing World"). The article ends with an overall assessment of the Bretton Woods System (section "An Assessment of the Bretton Woods System").

The Design of the Bretton Woods System

The Bretton Woods Articles of Agreement were aimed at creating a new international monetary system with fixed but adjustable exchange rates. It was supposed to avoid the monetary problems of the interwar years, as they were perceived by the architects of the postwar order. These problems included destabilizing speculation and competitive devaluations, the subordination of monetary policy to external balance, the asymmetry between creditor and debtor countries, and the ensuing deflation as well as bilateral trade clearing system and multiple exchange rates for different international transactions. In retrospect, some of the interpretations serving as the basis for the negotiations at Bretton Woods were mistaken. For instance, Eichengreen (1982) found no empirical evidence of destabilizing speculation and beggar-thy-neighbor competitive devaluations. But at the time, there was a strong consensus that pegged exchange rates were essential for international trade (Nurkse 1944), and this conviction persisted throughout the postwar years (Straumann 2010).

The postwar system negotiated in July 1944 in the US resort of Bretton Woods (New Hampshire) was a novelty in the history of international monetary regimes in that it was based on a set of rules agreed upon by the member states and created a new body, the International Monetary Fund (IMF). But the planners did not start from scratch. After the final collapse of the interwar gold exchange standard in 1936, France, Great Britain, and the USA had signed the short-lived tripartite agreement, later joined by Belgium, the Netherlands, and Switzerland, which was intended to provide fixed exchange rates based on central bank cooperation. This attempt may be considered the predecessor of the Bretton Woods system, even though it lacked a strong institutional foundation.

The Bretton Woods system was based on the convertibility of the dollar to gold (at $35 per ounce) and on a peg of all other currencies to the dollar. In addition, the 1944 agreement foresaw full convertibility for current account transactions in a

multilateral payment system but permitted governments to retain controls on capital account transactions. Transitory external imbalances were supposed to be smoothed out by international liquidity support provided by the IMF and capital flow controls. Finally, exchange rates should be adjustable in the presence of fundamental imbalances by the means of international coordination. Thus, the Bretton Woods system was supposed to restore the exchange rate stability of the classical gold standard without reintroducing its rigid market-driven adjustment mechanism (Horsefield 1969a).

The list of the 44 countries participating in the negotiations of the agreement shows that the international coverage of countries was very restricted. For obvious reasons, the axis powers were not invited to the conference, but even the neutral countries were excluded. Moreover, a sizable share of countries was represented by their exile governments as they were fully or partly occupied by Germany or Japan. Interestingly, the Soviet Union was attending the conference but did not sign the agreement and participate in the Bretton Woods system. The emerging Cold War overshadowed and eventually inhibited international monetary cooperation.

The Bretton Woods Articles of Agreement were a compromise between two plans formulated by the British economist John Maynard Keynes and Harry Dexter White, a senior official at the US Treasury (Steil 2013). The Keynes Plan stipulated a supranational institution, the International Clearing Union (ICU), that was supposed to provide a new international bank money, bancor, which is fixed (but not unalterably) in terms of gold and accepted as the equivalent of gold by the British Commonwealth and the USA and all members of the Union for the purpose of settling international balances. Moreover, generous overdraft facilities (total $25–30 billion at the beginning, adjustable to foreign trade growth) of national central banks were planned. Debtors with ICU were supposed to pay interest in favor of creditors, and the overdraft should be conditional on policy measures aimed at reducing external imbalances (parity change, capital controls). The White Plan stipulated a United Nations stabilization fund ($5 billion) financed by own currency and gold by member central banks allowing to keep exchange rates fixed in the presence of temporary imbalances. Parity changes were allowed in case of a fundamental disequilibrium on approval of three quarters of members, if larger than 10% of the initially fixed parity, and, besides capital controls, exchange controls were accepted for scarce currencies. The stabilization fund had the option to declare a currency "scarce," when the demand of other countries for it was considered to be large in relation to the fund's holdings of this currency. The Bretton Woods system included elements of both the Keynes and the White Plan but had a strong bias in favor of the US proposal. This was not surprising given that in 1946 the USA owned nearly three quarters of the world's stock of monetary gold and accounted for more than a third of international trade.

The 1944 agreement contained 20 articles (Horsefield 1969c). They comprised a series of obligations for members with respect to their currency and their treatment of international payments. According to Article III, they had to pay their quota of 25% in gold and of 75% on an account of the IMF at the member central bank. Article IV obliged them to keep their exchange rate and thus the price of gold within a 1% band around par values. Article VI allowed member states to introduce capital controls

when facing sustained capital outflows. Article VIII forbade restrictions on current payments and multiple currency practices without consent of the IMF. Moreover, it made exchange contracts violating exchange controls consistent with the Agreement legally unenforceable in any member country.

The 1944 agreement also stipulated that most members had little power to shape the decisions of the IMF. The paragraphs of Article XII on the rules of voting lead to a dominant influence of large countries on IMF decisions. The USA and the UK had a quota of $2,750 million and $1,300 million of a total of $8,800 million implying 27,750 and 13,250 votes, respectively. The quota of Canada, for instance, was $300 million resulting in 3,250 votes. This dominance was increased by the fact that the Soviet Union, which was present at the conference and had a planned quota of 1,200, did not sign the agreement resulting in a 53% US and UK share in the total quotas. Moreover, Article XVIII allowed the IMF to decide on all questions of interpretation without any possibility to appeal. This was particularly important as many aspects of IMF policy were not fixed in the Agreements and were decided at a later moment of time by the Fund. These conditions did not look favorable to most member countries even if it was easy to withdraw from the Fund formally: according to Article XV, a country simply had to send a letter to the IMF, and the withdrawal became effective once the Fund received the notice.

The incompleteness of the articles can be exemplified by two important examples. The first one concerns the regulations of drawings of IMF resources in Article V. Besides the requirement of a cap on the annual rate of change (25%) and the total (200% of quota) of IMF's holdings of the drawer's currency, there were no restrictions on the access to IMF resources for a member fulfilling its obligation. This vacuum led to the establishment of additional rules: in February 1952, 25% of the quota were declared unconditional (gold tranche), whereas drawings over 25% were subject to negotiations, resulting into the introduction of standby arrangements and credit tranche policy which provided the condition for drawings (policy measures recommended by the IMF) above the 25% (Gold 1969). The second example concerns the determination of par values. According to Article IV, these could be changed if a member faces "a fundamental disequilibrium." However, the Article was silent on an operational meaning of this term, and the IMF had to establish criteria for decisions on par values. The same applies to margins from par for forward exchange transactions which were left what the IMF deemed reasonable or cross rates when only the dollar exchange rate is kept within a 1% band. Thus, the 1944 agreement lacked clarity, when it came to adjustment and liquidity issues. Eventually, these unresolved issues would contribute to the downfall of the Bretton Woods system in the late 1960s.

The Long Path to Convertibility, 1946–1958

In the first years after it began to operate, the Bretton Woods system bore little resemblance to what the planners had envisaged in 1944. Europe was still suffering from the consequences of WWII. International trade was hampered by tariffs, quotas, and exchange controls. A return to the traditional European division of labor was

made impossible by bilateral payment agreements. The reason for this cumbersome system was the shortage of gold and foreign exchange. The dollar, the only major convertible currency, was in short supply. Europe went through difficult postwar years (Eichengreen 2007b).

The economic and social situation became particularly critical during the exceptionally cold winter of 1946–1947. In addition, political tensions were rising as the Soviet Union aimed at dominating Eastern Europe by installing Communist parties. In January 1947, the Communists took power in Poland. Greece was mired in a civil war between nationalist forces and Communist partisans. The situation in Asia was similarly unstable. The Chinese Communist Party under Mao Zedong was gaining ground in its campaign against government forces under Chiang Kai-shek. The Korean peninsula was divided between North and South, with the former occupied by Soviet and the South by US troops. The Cold War was in the making and shaped the evolution of multilateral cooperation.

The IMF which began to operate in May 1946 did not have the power to ease international tensions and to provide a boost to the Western European economy. Only once, in May 1947, a member country, France, made a drawing from the Fund amounting to $25 million which provided a temporary relief. But as France introduced a multiple currency system, the IMF in January 1948 barred France from further drawings. The only power that could change the economic and political dynamic of Europe was the USA. In early June 1947, Secretary of State George Marshall, formerly Chief of Staff of the United States Army, outlined a new plan in a speech in Harvard. The European Recovery Program foresaw grants and loans amounting to more than $13 billion and a program aimed at rebuilding ravaged regions, liberalizing trade, abandoning war-related regulations, modernizing industry, and pushing back the influence of Communists parties, especially in France and Italy. As, starting in April 1948, the dollar funds were distributed to Western European countries, the IMF decided to bar member states from making drawings of their quotas in order to protect its resources.

The Marshall Plan, as it was dubbed soon after it was launched, proved successful. It led to trade liberalization, political stability, investment, deregulation, and recovery and eased the problem of dollar shortage (Eichengreen 2007b). The Organization for European Economic Cooperation (OEEC) made sure that the plan was implemented. Especially trade liberalization fostering the reemergence of the traditional European division of labor made a crucial difference. Trade relations were further enhanced by the General Agreement on Tariffs and Trade (GATT) taking effect in early 1948 and the European Payments Union (EPU) established in 1950 in coordination with the OEEC. The latter arrangement reintroduced a multilateral clearing system for current account transactions. The external balance of Western European countries was improved by the 1949 devaluation of sterling and the currencies of Australia, Belgium, Canada, Denmark, Egypt, Finland, France, Greece, Iceland, India, Iraq, Luxembourg, the Netherlands, Norway, and South Africa. The postwar period characterized by hunger, black markets, and political instability came to an end. By the early 1950s, the current account deficit of Western Europe turned into a surplus. As a

result, the problem of dollar shortage was solved by the rebalancing of trade relations. In this process, the IMF played a negligible role. By the early 1950s, almost all observers considered it a moribund institution (James 1996).

The introduction of free convertibility for current account transactions took much longer, however. It was not until December 1958 that 14 Western European countries took this step toward a liberal international payment system and abandoned the regional EPU that had helped to promote the European recovery: Austria, Belgium, Denmark, Finland, France, the Federal Republic of Germany, Ireland, Luxembourg, the Netherlands, Norway, Portugal, Sweden, and the UK (Eichengreen 1996, pp. 111–115). Fifteen other members, most of them former members of the British Empire, adjusted their exchange control regulations to the new conditions; these were Australia, Burma, Ceylon (Sri Lanka), Ghana, India, Iraq, Libya, Malaya (Malaysia and Singapore), Morocco, New Zealand, Pakistan, the Sudan, Tunisia, and South Africa. Japan introduced convertibility of current account transactions in 1964.

The long duration of the transition period may have been crucial for the economic recovery of Western Europe after WWII, providing some breathing space for full-employment policies, but the delay ran clearly against Article VIII of the Bretton Woods agreement. Similarly, multiple exchange rates did not disappear as rapidly as the architects of postwar monetary cooperation had hoped for. As a matter of fact, from 1947 to 1955, the number of countries running multiple rates even increased. Only after the IMF in 1955 urged members to simplify their exchange rate systems, the trend turned (de Vries 1986).

The reason for the long transition period was that some Western European countries found themselves in a vulnerable situation. Notably, Great Britain was struggling to find its way back to a new equilibrium after the premature introduction of convertibility in 1947 that had to be reversed after a few months. The British currency was freely convertible within the sterling zone encompassing the former colonies and the Commonwealth (Schenk 2010). On the other hand, postwar British governments focused on policies favoring full employment requiring expansionary policies. As a result, Great Britain suffered from large capital outflows. Not surprisingly, the IMF's demand to expand convertibility to all current account transactions outside the sterling area met strong resistance in London. What changed the mind of British officials was the Suez Crisis in late 1956 which triggered a substantial capital outflow and hereby required a larger financial assistance by the IMF. In return, London agreed to proceed with restoring convertibility for current account transactions.

Similarly, successive French governments favored expansionary policies over concerns about the external account, not least because of its costly wars in Indochina and ambitious social programs. In addition, the monetary overhang resulting from war and occupation had not been eliminated in the late 1940s which required all sorts of wage and price controls in order to contain inflation (James 1996). As in the case of Great Britain, the Suez Crisis was the tipping point. France, losing reserves at an alarming rate, increased import tariffs, received an IMF loan, and eventually engineered a devaluation by introducing a complicated scheme of trade subsidies.

After another devaluation in 1958 and fiscal stabilization program, France was ready to lower tariffs and introduces convertibility for current account transactions.

Germany was more forthcoming in liberalizing its payment system after suffering from a severe balance of payments crisis in the early 1950s. The strong current account surplus that emerged after the crisis made it easier to combine full-employment policies and external stability. As a result, Germany began to liberalize its capital account as early as 1957 (James 1996). By contrast, Japan, the future economic challenger of US hegemony, needed more time to make the transition toward convertibility. The economy had grown at a fast rate since the late 1940s but repeatedly suffered from balance of payments crises, notably in 1961. The ensuing involvement of the IMF led to discussions that finally convinced Japanese officials to pursue the path of liberalization and to subscribe to Article VIII.

There was one country, Canada, that took a completely different path in the 1950s. Instead of delaying the adoption of convertibility, the government liberalized external payments at an early stage and used its exchange rate to deal with balance of payments disequilibria. The decisive step was taken in 1950 in the wake of huge capital inflows from the USA. By letting the Canadian dollar appreciate, the authorities hoped to regain monetary policy autonomy. Though strongly criticized by the IMF, the strategy worked quite well until 1956, inspiring academic economists and public officials elsewhere to advocate flexible exchange rates. But after 1956, the monetary fiscal policy mix proved inadequate (Bordo et al. 2010). The experiment was finished in 1962.

The Bretton Woods System Under Convertibility, 1958–1968

The introduction of convertibility for current account transactions had a profound impact on the functioning of the Bretton Woods system. Maintaining exchange rate stability and preserving monetary autonomy became more difficult, as capital began to flow more freely between countries. Officially, convertibility was confined to current account transactions, but investors found many ways to shift capital across frontiers outside of imports and exports of goods and services. They could either prepone or delay a payment ("lead and lag"), or they pretended to make payments for imports and exports that never occurred. The larger the global trade volume became, the more difficult it was for the authorities to control capital movements.

Contemporaries were aware of the new constraints caused by increasing capital mobility. In 1964, 32 economists published a report which identified the three major problems of the Bretton Woods system: liquidity, adjustment, and credibility (Machlup and Malkiel 1964). The liquidity problem was stemming from the lack of a mechanism that provided international reserves sufficient to finance the growth of international trade and finance. Neither the IMF's quotas nor its holdings of gold and members' currencies were large enough. De facto the dollar served as world reserve currency which loosened the constraints on the US balance of payments which undermined the stability of the dollar and the entire monetary system.

The adjustment problem resulted from the difficulty to reduce international imbalances by demand-management policies and changes of the par values. Depending on the fiscal and monetary stance, capital movements reinforced or counteracted the intended effect of a specific economic policy. The problem of lacking exchange rate flexibility had to do with domestic politics, the IMF's ambiguous definition of a "fundamental disequilibrium" of the current account, and the unpractical procedure required in the case of a change of the par value. A case in point was the revaluation of the deutschmark and the Dutch guilder in 1961 by 5%. An effective rebalancing would have required a much stronger revaluation, but opposition of the exporting sectors prohibited a more effective adjustment. Of course, the malfunctioning of adjustment mechanisms reinforced the liquidity problem. Countries delaying the policies required to reduce imbalances developed an increasing demand for liquidity.

Finally, the confidence problem was a consequence of the liquidity problem. As the dollar became the world reserve currency which was fueled by sizeable US balance of payment deficits in the 1960s, holders of dollars became increasingly skeptical toward the stability of the dollar. The interaction between the liquidity and the confidence problem became also known as the Triffin dilemma, named after the Yale economist Robert Triffin (1960).

In the early 1960s, policymakers concentrated on the liquidity problem for several reasons (de Vries 1987). First, they believed that the costs of adjustment may be too high for deficit countries, as it undermined the policy of full employment. Second, they believed that solving the liquidity problem would also mitigate the adjustment and the confidence problem. Third, the adjustment problem was considered less urgent, as the USA continued to run a current account surplus and the balance of payments deficit did not worsen until the mid-1960s. Many expected that international imbalances would automatically be reduced by strong economic growth and trade expansion.

Two measures were taken to address the liquidity problem (Bordo 1993; Eichengreen 1996; James 1996). The first one consisted in expanding the resources of the IMF. The first general increase in Fund quotas became effective in September 1959, a few months after the introduction of convertibility in December of that year. Members' quotas were expanded by 50%, resulting in an increase of the Fund's resources from $9.2 billion to $14.0 billion. In February 1966 the Fund quotas were raised by another 25% for all members. The Fund's resources were increased to $21 billion. In addition, in 1961 the ten largest industrial members concluded the General Arrangements to Borrow (GAB) in which they declared themselves ready to provide the IMF with another $6 billion. Out of this cooperation grew the Group of Ten (G10) that played an important role in the crisis management.

The second measure was the creation of a new reserve asset, the Special Drawing Right (SDR). Talks began in the early 1960s; the first SDR became available in 1969. The value of the SDR was equivalent to 0.888671 g of fine gold which was the gold content of one dollar. The SDR is a potential claim on the freely usable currencies of IMF members. Holders of SDRs can obtain these currencies in exchange for their SDRs in two ways: first, through the arrangement of voluntary exchanges between members, and second, by the IMF

designating members with strong external positions to purchase SDRs from members with weak external positions.

Giving priority to the liquidity problem did not imply that nothing was done to lessen the adjustment and the credibility problem. Yet, it was not the IMF which initiated these measures but the US government with the support of Western European countries. In 1961, the USA, the UK, France, Germany, Italy, Belgium, the Netherlands, Sweden, and Switzerland agreed on a mechanism to stabilize the gold price at $35 per ounce in the London gold market. They pooled part of their gold reserves, with the USA providing 50% of the total, and intervened in the market whenever the price of gold threatened to rise or slide. The purpose of the Gold Pool was to maintain confidence in the dollar peg to gold serving as the anchor of the international monetary and financial system.

In the same year, central banks of the G10 countries and Switzerland began to set up a system of swap agreements which aimed at stabilizing the gold reserves of the USA. By these swap agreements, the Fed acquired foreign currencies which it used to absorb forward sales of dollars by foreign central banks hedging exchange risk on their dollar holdings. To obtain additional foreign currencies in order to repay the Fed's swap drawings, the US Treasury issued nonmarketable foreign currency-denominated bonds to foreign central banks. These bonds named after Treasury Undersecretary for Monetary Affairs Robert Roosa ("Roosa bonds") bore a higher interest rate than equivalent medium-term treasury bonds and protected foreign central banks against the risk of a dollar devaluation, since both the principal and the coupon were not based on the dollar but on the currency of that particular country (Bordo 1993; Eichengreen 1996).

In addition, the US government introduced measures containing capital outflows in order to reduce the balance of payments deficit. The Kennedy administration increased taxes on foreign earnings of US corporations and put in place the Interest Equalization Tax which taxed the earnings of foreign securities by 1%. It also reduced public spending abroad by setting new rules for defense government purchases and foreign aid programs. The Johnson administration expanded the system by introducing a tax on the earnings of foreign loans, imposing guidelines on direct investment and limiting the growth of bank lending to foreigners.

All these measures helped to prolong the life of the Bretton Woods system (Schenk and Straumann 2016). But they failed to solve the adjustment and the confidence problem. The US balance of payments deficit persisted. As Fig. 1 shows, the change in net reserves was mostly negative from 1958 to 1971, indicating a decrease in reserves and thus an export of US dollars. In 1965, with the escalation of the Vietnam War, US monetary policy became more expansionary, thus generating inflationary pressures that eventually spilled over to the whole Bretton Woods system and exacerbating the conflict between the USA and the surplus countries Germany and Japan (Bordo 1993; Bordo and Eichengreen 2013). The expansion of monetary aggregates also undermined the functioning of the London Gold Pool, central bank swap agreements, and the issuance of Roosa bonds. Thus, increasingly, the Bretton Woods system was suffering from a liquidity surplus, not a liquidity shortage. In such an environment, the SDRs which became effective in 1969 were largely obsolete.

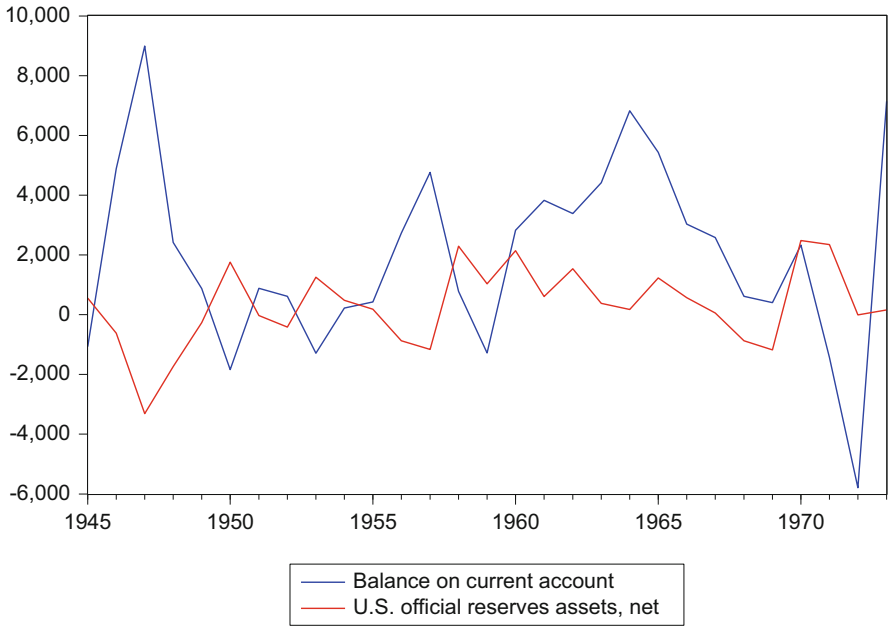


Fig. 1 US balance of payments (million dollars): balance of current account and net official reserves assets

The administrative measures taken by the Kennedy and the Johnson administration fostered the rise of the euro dollar markets in London. And annoyed by the privileged position of the USA as the issuer of the major reserve currency, France began to accelerate its gold purchases in 1965 and quietly left the Gold Pool 2 years later. When in November 1967 the French press reported the exit, confidence in the dollar declined (Eichengreen 2007a, p. 57). The forced devaluation of sterling by 14% occurring in the same month further undermined confidence in the sustainability of the Bretton Woods system.

The Collapse of the Bretton Woods System, 1968–1971

The final phase of the Bretton Woods system was initiated by the end of the Gold Pool in 1968. After the devaluation of sterling in November 1967, the gold holdings of the Gold Pool declined by $3 billion in the following 4 months, due to deteriorating confidence in the stability of the dollar and gold scarcity. Policymakers reacted with the introduction of a two-tier system in March 1968. The gold price for private demand in London jumped, while the price of official transactions was kept at $35 per ounce. At the same time, US President Johnson proposed a bill removing the 25% gold cover of banknotes. This measure was a clear signal that gold convertibility of the dollar was no longer warranted with increasingly expansionary

US monetary and fiscal policy. As a result, the Bretton Woods system turned into a de facto dollar standard (Bordo 1993).

Meanwhile, France was destabilized by strikes and student protests in May 1968. The government reacted with expansionary fiscal and monetary policies which triggered capital flight. An international rescue package, the introduction of capital controls, and a shift to restrictive economic policies temporarily alleviated the pressure, but eventually, in August 1969, the French franc was devalued by 11% (Bordo et al. 1995). By contrast, Germany, a surplus country receiving funds fleeing from France, let its currency float and revalued it in September 1969 by 9%.

To bridge future balance of payments crises, the IMF in February 1970 agreed to a further rise of the Fund's resources to $28.9 billion. But predictably, providing more liquidity to IMF members did little to solve the adjustment and the confidence problem, as US monetary policy turned strongly expansionary in 1970–1971, resulting in a record high decrease of net reserves (see Fig. 1). Increasingly, speculators turned their dollar assets into gold, prompting the Federal Reserve to increase the discount rate. The measure proved too weak to stop the run on the dollar, however. In August 1971, President Nixon "closed the gold window," i.e., suspended the gold convertibility of the dollar, introduced a 10% import tariff, and imposed a wage and price freeze. In December 1971, the USA devalued the dollar by 8%, thus raising the price of gold from 35 to 38 dollars per ounce, while the currencies of the Benelux, Germany, Japan, and Switzerland were revalued. The fluctuation bands of currencies were widened from 1% to 2.25%.

The Smithsonian Agreement failed to restore confidence. The currency adjustments proved inadequate, and the USA continued to pursue expansionary policies. The run on the dollar resumed in 1972. Expecting the near end of the Bretton Woods system, the members of the European Community (EC) in April 1972 introduced smaller margins for their exchange rates ("snake in the tunnel"). In June 1972, sterling under the pressure of a speculative attack became the first currency to leave the Bretton Woods system and to float. In January 1973, the Swiss franc suspended its peg and began to float. In February, the yen followed. Eventually, in March 1973, the Bretton Woods completely collapsed, as the EC members introduced a joint float of their currencies against the dollar. A new era began.

The Bretton Woods System and the Developing World

The architects of the Bretton Woods system, especially the US officials under the guidance of William Dexter White, were well aware of the importance of a broader approach that included the less advanced economies (Helleiner 2014). Among the 44 countries represented at the Bretton Woods conference in 1944, there were 19 Latin American countries (all but Argentina), 4 African countries (Egypt, Ethiopia, Liberia, South Africa), 5 Asian countries and colonies (China, India, Iran, Iraq, the Philippines), and 4 Eastern-European countries (Czechoslovakia, Greece, Poland, and Yugoslavia). With the exception of the USSR, all of them signed the agreement, and over the years, especially in the wake of decolonization,

nearly all developing countries joined the IMF and maintained their membership. Between 1944 and 1973, only three countries withdrew (Czechoslovakia, Cuba, and Poland), as the IMF denied access to its resources. Indonesia withdrew and was readmitted in 1954 and 1967, respectively.

Yet, the broader development goals discussed in Bretton Woods did not result into a comprehensive framework and played only a minor role in the IMF's policy after 1946. The Truman administration had other priorities, and Eastern Europe's subjugation under Soviet rule and the communist revolution in China in the late 1940s profoundly changed the debate about which economic policy was best for promoting economic growth and welfare. Nevertheless, the IMF was relevant for developing countries during the period of the Bretton Woods system. Balance of payments crises were frequent, calling for financial support and economic advice. The experiences the Fund staff made before 1973 served as basis for adjustment programs implemented after the collapse of fixed exchange rates, the era of oil price hikes, and the Latin American debt crisis.

Of course, developing countries found themselves in a fairly different position than the industrial countries. While the latter group in the 1950s was on the path to convertibility for current account transactions, developing countries were struggling with more fundamental problems. Their primary concerns were fluctuations of exports, adverse terms of trade, and slow growth. In such a difficult environment, many policymakers in developing countries believed that enhancing current account convertibility was not a priority. On the contrary, they preferred all sorts of trade barriers, multiple exchange rates, and import substitution policies to promote domestic industrial development. Yet, as their policies often resulted in high inflation and balance of payments crises, they were repeatedly forced to seek financial assistance by the IMF. Between the mid-1940s and the early 1960s, five Latin American countries (Argentina, Brazil, Chile, Colombia, and Mexico) as well as India and Indonesia made greatest use of the Fund's resources. In addition, 13 smaller Latin American members, 3 Asian countries (Burma, the Philippines, Sri Lanka), 4 Middle Eastern countries (Iran, Israel, Pakistan, Syria), 1 African country (Sudan), and 2 European countries (Turkey, Yugoslavia) asked the IMF for financial support (de Vries 1987).

The Fund tied financial assistance to stabilization programs comprising credit, fiscal, and exchange policies mainly aimed at containing inflation by curtailing monetary financing of budget deficits. The success of the Fund's programs was mixed and met increasingly strong criticism. Some countries managed to bring down inflation but suffered from strong recessions. In addition, it was increasingly recognized in Washington DC that poor countries needed a set of economic policies to spur growth that went beyond securing monetary stability (James 1996). Accordingly, in the 1960s, the US and international institutions focused more strongly on growth strategies. The IMF in its 1963 annual report emphasized the importance of promoting development. It also adopted a more holistic view regarding the underlying problems of the developing world. In addition, the IMF established a new form of financial support without conditionality. Yet, according to its primary mandate, financial assistance in combination with a stabilization program remained in the

center of the Fund's task, as balance of payments crises remained frequent problems. Between 1963 and 1972, the IMF prescribed nearly 80 stabilization programs of which 72% were fully implemented, with most of them leading to an improvement of the external balance. Of course, these emergency interventions were not suited to address the underlying structural weaknesses of developing countries. Between 1960 and 1970, the aggregate current account deficit of developing countries being member of the IMF rose from less than $5 billion to $8 billion. In 1971, it increased by another $1.5 billion. The rising deficit was mainly due to higher payments for international services, in particular for the servicing of foreign debt (de Vries 1987). The gap between the industrial creditor countries and the developing debtor countries remained significant from the early days to the final phase of the Bretton Woods system.

An Assessment of the Bretton Woods System

The Bretton Woods system operated for more than 25 years. This appears to be a long period for a system with a lot of obligations but a negligible influence on IMF decisions for most member countries. There are several reasons for this. First, there was the incentive to use the resources of the Fund to finance balance of payment deficit. The economics of this incentive to cooperate instead of deviate from the obligations as implied by noncooperative game theory is analyzed by Dominguez (1993). Secondly, the voting power got more balanced by the increase in the number of members from 29 at the beginning to 104 in 1966. Despite the general quota increases in 1959, the USA share decreased from 36% to 25% (Horsefield 1969b). Besides the strong qualitative increase in membership, this reflects the admission of major former WWII enemy countries like Germany, Italy, and Japan not represented at the Bretton Woods conference. Correspondingly the number of Executive Directors was increased from 12 to 20 in 1966.

Thirdly, IMF decisions in the Executive Board were mostly reached by negotiations and consensus, and the asymmetry in voting power was partly neutralized (Gold 1969). Finally, the IMF became very tolerant with respect to the obligations of its members (Gold 1969). There were devaluations against the rules, in particular the 30% devaluation of the pound in 1949, which was only quickly approved ex post by the IMF, and instances of floating exchange rates (Canada from 1950 to 1961 being the most well-known example). Indeed, in 1966 only 27 of 104 members fulfilled the obligations of Article VIII (Horsefield 1969b). Article XIV, which aimed at a transition period of 5 years, was extensively used in order to legalize all kind of violations from multiple currency practice to floating exchange rates. Sanctions (denying access to the IMF resources) were rarely imposed, a notable exception being France in 1948 in response to unauthorized multiple currency practices (Gold 1969).

Yet, on the other hand, 25 years appear to be a short period given the good macroeconomic performance provided by a series of good policies and circumstances that were not directly related to the Bretton Woods system. Instead of enforcing the repayment of war debts and reparation as after WWI, the USA fostered

the recovery of Europe by the Marshall Plan. In addition, the very low level of production in industrialized countries strongly involved in WWII allowed a strong catch-up effect as we see it today in the high growth of the emerging economies. By contrast, the bad performance in the post 1973 period is to some extent caused by problems piled up during the postwar era, notably the high levels of inflation.

Two fundamental problems explain the collapse of the Bretton Woods system. The first one was to build the international monetary regime on fixed but adjustable exchange rates. Contrary to the international gold standard of 1880–1914, which clearly constrained monetary policy autonomy, Bretton Woods was ambiguous on the option chosen and did not anchor expectations: giving up freedom of capital movement was seen as one of the instruments against transitory imbalances, whereas adjustment of exchange rates was considered as a remedy for fundamental imbalances. However, capital controls became increasingly inefficient and difficult to enforce with current account convertibility. Thus, in line with the well-known "impossible trinity" theorem positing the impossibility of fixed exchange rates, monetary (and to some extent fiscal) policy autonomy, and freedom of capital movements, the goal of fixed exchange rates became unsustainable. Moreover, devaluation in response to fundamental problems provided a one-way bet to speculators. This problem clearly showed up in the foreign exchange crises of the 1960s resulting in devaluations and revaluations (for instance, devaluation of sterling and the French franc 1967/1969, revaluation of the DM 1961/1969).

The second fundamental problem was that the Bretton Woods system turned, against the intention of its architects, into a gold dollar standard. This development was caused by the initial conditions of the system after WWII: the USA was the world's largest creditor with, in addition, two thirds of world gold reserves, a current account surplus, deep financial markets, and a convertible currency. Most other countries were debtors with depleted gold reserves, current account deficits, and repressed financial markets and exchange controls. Thus, the USA was the only country which fulfilled Article IV by pegging the price of gold at $35, and all other currencies were pegged to the dollar with a 1% margin. Moreover, drawings of IMF members were mainly in dollar which was the only major convertible currency until 1960: 87% of total drawings from 1946 to 1960 of 3683.5 million were purchases of dollar (Horsefield 1969b). Therefore, the deposits of the IMF at the central bank of most members were not used for financing balance of payments deficit as intended. This composition of drawings changed in the 1960s: the dollar share was reduced to 43%, and the DM share increased 16% for the years 1961–1965, but the leading role of the dollar as reserve currency was not really challenged.

Moreover, par exchange rate values for European currencies fixed mostly at the prewar level in 1946 which was clearly overvalued after the war. Given these structural problems, the IMF resources, which were sized for temporary balance of payments problems, were too small in order to allow a quick transition to convertibility. Instead the US and Canadian Loans to the UK (1945 $3.75 billion and $1.2 billion respectively) and Marshall Plan (1948–1951, $13 billion) provided the funds for recovery and current account surpluses in Western Europe which were also supported by the 1949 devaluations of European currencies. The Marshall Plan

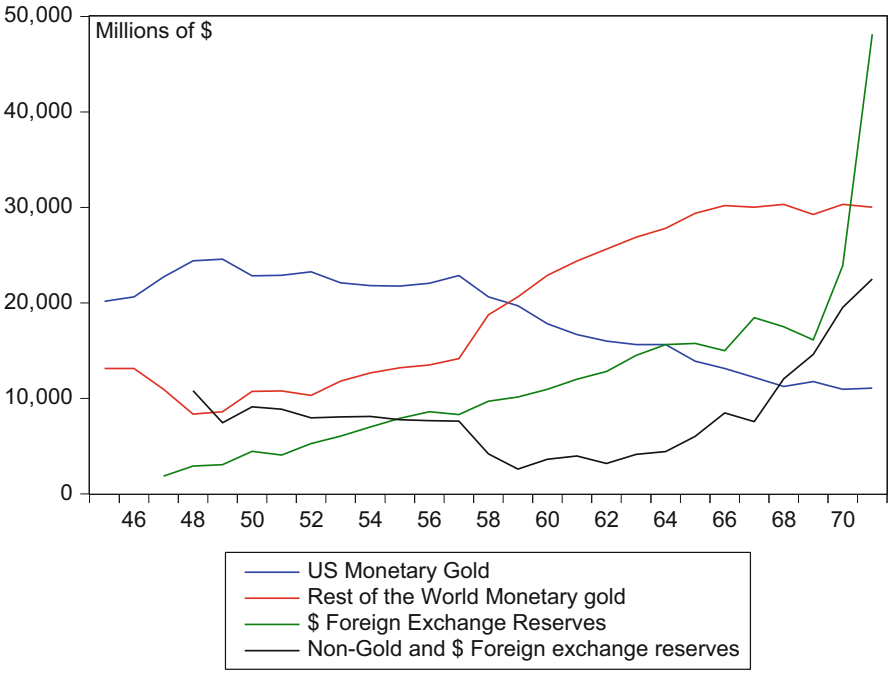


Fig. 2 World international monetary reserves during the Bretton Woods period

and private US long-term capital exports provided the dollar reserves which allowed the introduction of convertibility in 1958 and provided a short run solution of the "liquidity problem." The USA became the "bank" of the system with long-term foreign assets and short-term liabilities and the corresponding potential liquidity problem.

Figure 2 shows the development of international reserves, based on data from Horsefield (1969c) and International Financial Statistics of the IMF. Gold is disaggregated into US monetary gold stock and that of the rest of the world. Non-gold foreign exchange reserves are disaggregated into dollar reserves and the rest (sterling and other currencies, reserve position with IMF, Special Drawing Rights 1970/1971). We see that gold was the main international reserve asset in the 1940s, and most of it belonged to the USA. Non-gold reserves were mainly sterling, and the dollar played a minor role in this respect. We see a steady growth of dollar reserves (liabilities of the USA to foreign monetary authorities) which accelerates tremendously in 1968. The monetary gold stock shows only a minor increase (given high postwar economic growth), and there is a lot of redistribution from the USA to the rest of the world. In 1964 the US gold reserves became smaller than the US total (official) foreign liabilities. This development made the "confidence problem" noted firstly by Triffin (1960) obvious as the convertibility of official foreign dollar holdings into gold became questionable.

The confidence problem can also be illustrated with the development of the London gold price. The first essential deviation from the official $35 occurred in late 1960 after the presidential election of JFK. In reaction to these problems, the G10 gold pool was formed to keep the gold price at $35, but nevertheless some central banks (in particular the Banque de France) substituted dollar reserves by gold. Heavy gold losses of the pool at the beginning of 1968 led to a two-tier system (free market price for private transactions, $35 official transactions). In 1968 the USA removed the 25% gold cover of banknotes which was an important signal that gold convertibility of the dollar was no longer warranted with increasingly expansionary US monetary and fiscal policy. These developments led to a US inflation rate higher than the low values of the 1950s and early 1960s, which was transmitted to other countries with to some extent low inflation preference as in Germany. Under these circumstances the conflicts on the amount of international liquidity between surplus and deficit countries with different inflation preferences allowed for only a meager volume of IMF Special Drawing Rights as an inferior international reserve asset (restricted use for financing balance of payments deficits).

The adjustment problem of deficit and surplus countries proved to be asymmetric, and many attempts to keep the parity were unsuccessful in the end, despite IMF/G10 loans (General Agreement to Borrow, 1961), import surcharges, tight capital controls, and discrimination of foreign deposits. Relatively high inflation and low productivity in the UK produced current account deficits and capital outflows resulting in repeated sterling crises which finally led to a 14.3% devaluation in 1967. Relatively low inflation and high productivity in Germany produced current account surpluses and capital inflows which finally resulted in DM revaluations in 1961 (5%) and 1969 (8.15%). Other deficit countries as France and surplus countries as the Netherlands made similar experiences as the UK and Germany, respectively.

Given these developments the final collapse of the Bretton Woods system became unavoidable. European countries were not willing to accept a dollar standard and were themselves divided by their different economic developments and inflation preferences, which were fueled by the “Phillips illusion” of a persistent trade-off between inflation and unemployment. In August 1971, President Nixon closed the gold window, and the dollar could no longer be converted at $35 per ounce in official transactions.

The Bretton Woods system was similar to a bimetallic system with a legally fixed relation. The only difference was that the Bretton Woods system was based on gold and dollars at a ratio of $35 per ounce of gold, not on gold and silver. In general, a bimetallic system is more flexible than a monometallic standard and should lead to a lower degree of price level variability in a world without transaction costs (Niehans 1978; Friedman 1990; Flandreau 2002; Velde and Weber 2000). The problem with the bimetallic standard is that large changes in the relative volume of the two metals by discoveries of large deposits as in the Californian gold rush in 1848 lead to pressure on relative price of the two metals. When the relative market price deviates from the legal ratio, the legally overvalued metal will be used as money and may completely crowd out the other metal as money, as posited by Gresham’s law.

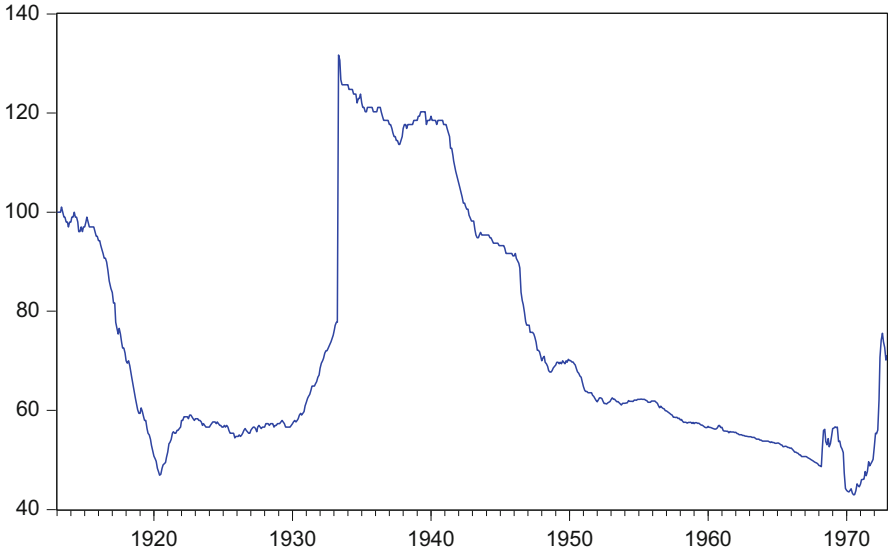


Fig. 3 Real $ price of gold, January 1913–December 1975. (Sources: Official gold price from 1912 to 1949, London market price from 1950, US CPI is from Federal Reserve Economic Data (http://research.stlouisfed.org/fred2/))

The Bretton Woods system was different from a bimetallic standard in that only gold and not the dollar had a nonmonetary use. However, gold may be crowded out of as money, if its relative price is falling and it becomes more attractive for nonmonetary use and gold production gets less profitable. Figure 3 plots the real dollar gold price (gold price divided by the consumer price index) from 1913 to 1972. During inflationary periods (WWI, WWII, and Bretton Woods), the real gold price strongly declined while climbing during periods of deflation, especially in the Great Depression. As a result of the inflationary bias of the Bretton Woods system, the real gold price in 1967 amounted to approximately half of its 1913 value.

Therefore, it is not surprising that the system was characterized by a substitution of gold by dollar, ultimately leading to a dollar standard. The substitution of gold by the dollar as international reserve is illustrated in Fig. 4 which displays the development of the gold and dollar share in international reserves from 1948 to 1971. At the beginning of this period, gold was dominant in international reserves (ca. 75%), and the dollar share was tiny (ca. 5%). The share of dollar steadily increased to 25% and the jumped to 40% in 1970. The mirror image is the fall of the gold share to a value below 40% in 1971.

The data allow to test the existence of a relationship between the two series. Indeed, the development of the dollar share is strongly negatively correlated with the

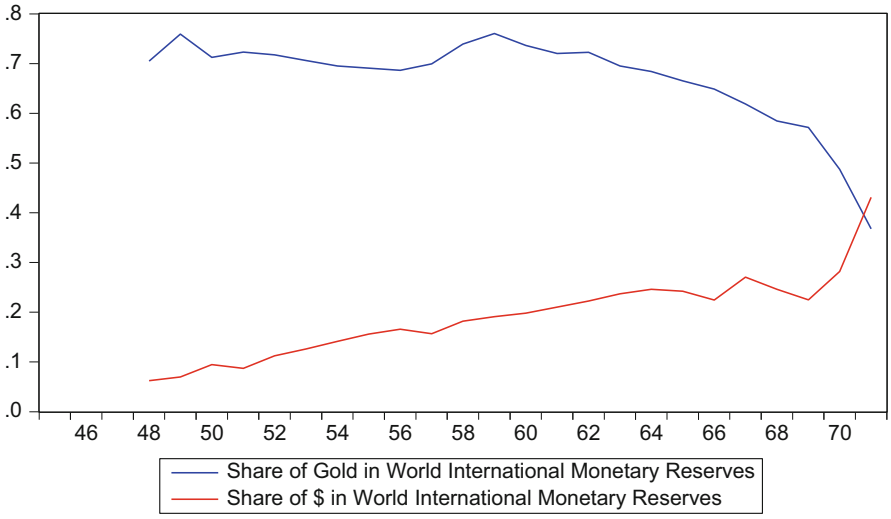


Fig. 4 Composition of international reserves 1948–1971

real dollar gold price displayed in Fig. 4. Indeed, the simple correlation coefficient is −0.89, and this correlation is not spurious, as it survives in vector autoregressive framework showing that the dollar share is influenced by the real gold price lagged 1 year even if we control for the last year's level of the dollar share (Kugler 2016). Thus, the fundamental flaw of the Bretton Woods system was its fixed link to gold without readiness to accept temporary deflations, which was a taboo since the disastrous deflation of the Great Depression. The only way to avoid this problem was a revaluation of gold in terms of the "numeraire" currency of the system, which, however, would have made the system strongly discretionary and vulnerable to the problem of time inconsistency.

Cross-References

- ▶ International Currencies in the Lens of History
- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ The Evolution of Monetary Policy (Goals and Targets) in Western Europe
- ▶ The Evolution of US Monetary Policy
- ▶ The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy

References

Bordo M (1993) The Bretton Woods international monetary system: a historical overview. In: Bordo MD, Eichengreen B (eds) A retrospective on the Bretton Woods system: lessons for international monetary reforms. University of Chicago Press, Chicago, pp 3–108

Bordo M, Eichengreen B (2013) Bretton Woods and the great inflation. In: Bordo MD, Orphanides A (eds) The great inflation: the rebirth of modern central banking. University of Chicago Press, Chicago, pp 440–489

Bordo M, Simard D, White E (1995) France and the Bretton Woods international monetary system 1960 to 1968. In: Reis J (ed) The history of international monetary arrangements. Macmillan, London, pp 153–180

Bordo M, Dib A, Schembri L (2010) Canada's pioneering experience with a flexible exchange rate in the 1950s: (hard) lessons learned for monetary policy in a small open economy. Int J Cent Bank 6(3):51–99

De Vries MG (1986) The IMF in a changing world, 1945–85. International Monetary Fund, Washington, DC

De Vries MG (1987) Balance of payments adjustment, 1945 to 1986: the IMF experience. International Monetary Fund, Washington, DC

Dominguez KME (1993) The role of international organizations in the Bretton Woods system. In: Bordo MD, Eichengreen B (eds) A retrospective on the Bretton Woods system: lessons for international monetary reforms. University of Chicago Press, Chicago, pp 357–404

Eichengreen B (1982) Did speculation destabilize the French franc in the 1920s? Explor Econ Hist 19(1):71–100

Eichengreen B (1996) Globalizing capital: a history of the international monetary system. Princeton University Press, Princeton

Eichengreen B (2007a) Global imbalances and the lessons of Bretton Woods. The MIT Press, Cambridge

Eichengreen B (2007b) The European economy since 1945: coordinated capitalism and beyond. Princeton University Press, Princeton

Flandreau M (2002) Water seeks a level: modeling bimetallic exchange rates and the bimetallic band. J Money Credit Bank 34(2):491–519

Friedman M (1990) Bimetallism revisited. J Econ Perspect 4(4):85–104

Gold J (1969) Part V: constitutional development and change. In: Keith Horsefield J (ed) The International Monetary Fund 1945–1965: twenty years of international monetary cooperation, volume II: analysis. International Monetary Fund, Washington, DC, pp 513–605

Helleiner E (2014) Forgotten foundations of Bretton Woods: international development and the making of the postwar order. Cornell University Press, Ithaca

Horsefield JK (1969a) The International Monetary Fund 1945–1965: twenty years of international monetary cooperation, volume I: chronicle. International Monetary Fund, Washington, DC

Horsefield JK (1969b) The International Monetary Fund 1945–1965: twenty years of international monetary cooperation, volume II: analysis. International Monetary Fund, Washington, DC

Horsefield JK (1969c) The International Monetary Fund 1945–1965: twenty years of international monetary cooperation, volume III: documents. International Monetary Fund, Washington, DC

James H (1996) International monetary cooperation since Bretton Woods. International Monetary Fund, Washington, DC

Kugler P (2016) The Bretton Woods system: design and operation. In: Fox D, Ernst W (eds) Money in the Western legal tradition. Oxford University Press, Oxford, pp 611–630

Machlup F, Malkiel BG (eds) (1964) International monetary arrangements: the problem of choice: report on the deliberations of an international study group of 32 economists. Princeton University International Finance Section, Princeton

Niehans J (1978) The theory of money. Johns Hopkins University Press, Baltimore

Nurkse R (1944) International currency experience: lessons of the interwar period. League of Nations, Geneva

Schenk C (2010) The decline of sterling: managing the retreat of an international currency 1945–1992. Cambridge University Press, Cambridge
Schenk C, Straumann T (2016) Central banks and the stability of the international monetary system. In: Bordo M, Eitrheim Ø, Flandreau M, Qvigstad JF (eds) Central banks at a crossroads: what can we learn from history? Cambridge University Press, Cambridge/New York, pp 319–355
Steil B (2013) The battle of Bretton Woods: John Maynard Keynes, Harry Dexter White, and the making of a new world order. Princeton University Press, Princeton
Straumann T (2010) Fixed ideas of money: small states and exchange rate regimes in twentieth-century Europe. Cambridge University Press, Cambridge/New York
Triffin R (1960) Gold and the dollar crisis: the future of convertibility. Yale University Press, New Haven
Velde FR, Weber WE (2000) A model of bimetallism. J Polit Econ 108(6):1210–1234

Currency Boards

27

Atish Ghosh, Anne-Marie Gulde, and Holger Wolf

Contents

Abstract

Currency boards have a long history, much of it in (mostly British) colonies and dependencies, typically smaller and more open economies. The historical evidence suggests that these boards succeeded in maintaining exchange rate stability and facilitating trade, at the cost of tying up reserves and restricting the scope for

A. Ghosh · A.-M. Gulde
International Monetary Fund, Washington, DC, USA
e-mail: aghosh@imf.org; aguldewolf@imf.org

H. Wolf (✉)
School of Foreign Service, Georgetown University, Washington, DC, USA
e-mail: wolfhc@georgetown.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_38

monetary policy, with more ambiguous evidence on their broader economic impact. Exits were generally to pegged regimes and proceeded smoothly.

A more recent group of currency boards was motivated by stabilization rather than trade considerations. While the sample is small, the lesson from their experience is straightforward (and unsurprising). Boards deliver short-term stabilization and an extended honeymoon period. Whether stability lasts depends on policy choices during the honeymoon period. Countries using the good times to achieve lasting fiscal balance while vigilantly counteracting real appreciation and prudently managing capital inflows stand to reap longer-term benefits at modest costs. Those that do not are likely to face an unpleasant reckoning with reality a few years down the line. The latter experience has been more common: both Argentina and the Baltics eventually faced crises reflecting external developments aggravated by domestic political choices. The Baltics, with a clear exit path to the Euro, doubled down on the boards and eventually achieved a smooth transition. Argentina, lacking a clear longer-term transition perspective, shifted to a flexible exchange rate.

Overall, the experience suggests that CBAs are neither a low-cost panacea for all monetary ailments nor an overly restrictive arrangement all but certain to end in crisis. Rather, as with all exchange rate regimes, they promise specific benefits and come with specific costs, necessitating a careful consideration of the trade-offs in the particular circumstances.

Keywords

Currency board · Stabilization · Credibility · Exit strategy · Policy constraints

Introduction

Currency boards date back to the 1800s, enjoying their heyday in the first half of the 19th century as a trade-facilitating arrangement for British colonies. Closely associated with colonialism and seeming dated in an age embracing more activist monetary policy, most boards were replaced by central banks following independence. By the 1980s, the arrangement had become almost a curiosum, with the few surviving boards concentrated among small, very open economies and dependent territories.

Argentina's embrace of a currency board (the "Convertibility Plan") to achieve stabilization in 1991 revitalized interest. In the following years, boards were routinely discussed whenever monetary instability reared its head – through eventually only five additional countries, all in Europe, opted for boards. Just as Argentina's choice of a board lifted its appeal as a currency arrangement, its troubled demise a decade later dampened it. The Argentinean experience however remains the exception. Three of the other five recent boards gave way smoothly to monetary union after some two decades as the Baltics joined the Eurozone. The same transition path is expected for Bulgaria. Bosnia-Herzegovina's board has delivered on the stability promise, with eventual EU – and Eurozone – membership providing a possible, if distant, exit option.

In the next section we discuss their essential characteristics. The following section reviews the theory underlying the decision to adopt a pegged rate, and specifically a CBA. We then turn to three case studies, looking at the colonial boards, Argentina, and the Euro boards, before concluding.

Defining Boards: Soft and Hard Variants

Theories of exchange rate regimes tend to employ neat and clear-cut distinctions between alternative arrangements. Empirical work must contend with grey areas, and currency boards are no exception in this respect (see Hanke and Schuler 1994; Tsang 1999; Ho 2002). There is nonetheless broad consensus that for a regime to qualify as a currency board, it must have three defining characteristics. First, the nominal exchange rate must be *fixed* against some anchor currency. Second, domestic central bank liabilities must be *substantially* and de jure backed by foreign reserves, and the local currency must be *freely* convertible (at least for current transactions) into foreign exchange. Third, changing the first two features must require clearing *nontrivial* legal and political hurdles. Together, the three features lend the arrangement the credibility required to achieve and maintain macroeconomic stability.

Yet each of the italicized terms admits some interpretational leeway, translating into a range of regimes that potentially qualify or fail to qualify as a currency board arrangement. The desirable degree of latitude in defining what constitutes a currency board is open to debate. Some authors prefer to hew close to an orthodox structure with minimal discretion, banishing softer currency board variants to the group of (harder) ordinary pegs. Among the cases considered below, Latvia provides an example, combining a strong political with a weak de jure claim to currency board status.

Other researchers – including us in prior work and in this essay – prefer a more expansive definition, not only because it increases the sample of cases to be studied but also because it allows for an empirical assessment of the very importance of how stringently the three characteristics of a currency board should be defined: if adherence to strict orthodoxy matters, it should be reflected in macroeconomic performance. The downside of greater leniency is an injection of some subjectivity in determining which regimes (still) qualify as a currency board.

Criteria I: The Peg and the Anchor Currency

Among the defining characteristics, the requirement of a fixed nominal exchange rate is the least controversial. Yet even here choices must be made. Does the choice of anchor currency(ies) matter? In principle, there may be a trade-off between choosing the currency of the most important trading partner and choosing a currency with strong price stability credentials. In practice, most countries adopting currency

boards have chosen from a small set of reserve currencies such as the US dollar or the DM/Euro combining broad use with high credibility.

In principle, boards could be pegged against multiple currencies (Oppers 2000) with the foreign exchange holdings matching the basket weights (though the effective weights would vary over time with cross-currency movements). Indeed, for countries with a more balanced regional trade pattern, avoiding the dangers of a sharp movement in the external value of the anchor currency may carry some advantages. But with the exception of an SDR peg in Latvia, and a brief discussion of a dual peg to the Euro and the dollar in Argentina, the basket-based currency board model has not found favor, perhaps reflecting the greater challenges in public communication and the complications of reserve management that such an arrangement would entail.

A monetary authority operating under a fixed exchange rate regime retains the option of changing the parity. In the limit, frequent parity changes would transform a de jure peg into a de facto float. To qualify as fixed in more than a notional sense, the parity must thus be durable. The concern applies a fortiori to currency boards as their credibility fundamentally depends on the promise of convertibility at the announced parity. Parity changes – other than technical changes in the anchor currency, undertaken at market exchange rates – are thus incompatible with maintaining a currency board arrangement. Examples of these technical changes include the shift from the DM to the Euro in the wake of monetary union, the shift from Sterling to the US dollar in some Caribbean countries that wanted to cater more to the American tourist and the switch from the US dollar to the Euro in Lithuania as the country became more integrated with Europe.

Criteria II: Coverage Ratio and Convertibility

In cases where the currency board is adopted as part of a macroeconomic stabilization program (rather than to facilitate trade), the credibility of the regime – and hence its success in disinflation – derives from the promise that the domestic currency will be freely convertible into the anchor currency at the fixed exchange rate and that the monetary authorities will be constrained from pursuing inflationary policies. Backing this promise is the foreign exchange coverage of the domestic monetary liabilities.

Legally mandated 100% coverage of central bank liabilities by anchor currency reserves at the stated conversion rate provides a natural threshold for *hard* currency boards (Hanke and Schuler 1994), guaranteeing full convertibility while ruling out discretion. Should it also be a maximum? Deviating to the upside – permitting excess reserves – allows for limited intervention. Used wisely, excess reserves (or an external credit line) might enhance the durability (and hence credibility) of the regime as it permits some LOLR operations that may avoid the need to choose between saving the banking system and saving the exchange rate regime. Used unwisely, it undermines the very point of establishing a stabilization-focused board, the limitation on discretionary policy.

While the 100% coverage ratio provides a natural benchmark, there is nothing magical about that number. It is unlikely that *all* of the central bank's liabilities would be presented for conversion at the same time – and if there are bank runs and a large proportion of depositors try to convert into foreign currency, the 100% coverage of central bank liabilities will be far from adequate. Nevertheless, the lower the coverage ratio of central bank liabilities, the lower the credibility of the convertibility promise. We view 50% coverage of central bank liabilities as sufficient for the regime to qualify as a currency board. Again, the caveat is somewhat academic. In practice, although some boards have occasionally toyed with low coverage ratios (including during the 1940s and 1950s (Treadgold 2005)), most retained near- or even above-full coverage.

Beyond the coverage ratio, the specific form of reserves may matter. The credibility of the conversion promise rises with the liquidity and falls with the valuation risk of the reserve asset. Yet concentration of reserves in the most liquid asset – cash in the reserve currency or direct claims on the anchor currency monetary authority – reduces returns that could be used either to augment reserves or be transferred to the treasury. As a result, boards tend to hold a modestly diversified portfolio, albeit heavily tilted toward lower-risk, more liquid external assets.

A further issue concerns the distinction between net and gross reserves, arising if the board relies on credit lines or long-term borrowing (including from the IMF) to meet its coverage requirements. While such practice may be opportune and even necessary when the currency board is first established (as was the case when Bulgaria adopted its board in 1997), excessive reliance on external borrowing makes the arrangement vulnerable to the possibility that credit lines will not be renewed (especially during times of stress), ultimately undermining the credibility of the regime.

The specific arrangements to assure free convertibility come in several varieties. In the most orthodox setting, local bank notes are exchangeable at the currency board authority (and, for practical reasons, typically some authorized financial institutions) into anchor currency bank notes. One step removed are arrangements where the convertibility right at the currency board authority itself is restricted to financial institutions, which in turn provide conversion services to the nonfinancial private sector. While the divergence is minor if the authorized institutions enjoy high credibility, in the latter case a loss of confidence in the stability of the financial system may impair credibility of the convertibility option.

Criteria III: Barriers to Change and Exit Options

Ultimately, the credibility of the currency board depends on the public's perception of policymakers' commitment to the regime (including the central bank's independence in pursuing its mandate) – and on the economic and political costs of exiting the arrangement during times of stress. Boards enshrined in the nation's constitution are harder to reverse than those established by legislation, which in turn are harder to undo than those based on simple statements of intent or de facto practice. Similarly,

boards with politically insulated leadership selected by cross-party bodies may enjoy greater credibility than those whose directors are appointed (or even removable) by the current government. In the case studies below, we document that real-world boards span the spectrum from arrangements rooted in international agreements (Bosnia-Herzegovina) to monetary authorities that choose to abide by currency board rules even though they are not formally bound by them (Latvia).

Yet exclusive reliance on formal structures risks overlooking local political realities (notably the ease of reversing legal commitments) influencing the de facto importance of the de jure arrangements, suggesting caution in interpreting cross-board institutional differences as reliable indicators of relative credibility. That said, absent a deep dive into local political economy, we would opt for a simple binary classification, requiring a constitutional anchor for a hard currency board.

While some boards – typically in very small, open, and often politically dependent territories – are conceived as semipermanent trade-facilitating arrangements, other boards were introduced as potentially or even intentionally temporary responses to specific events, from recent political independence in the Baltics to high inflation in Argentina and Bulgaria and domestic conflict in Bosnia-Herzegovina. For these CBAs, the optimal regime calculus may shift over time, raising the question of exit arrangements.

In our view, the importance of idiosyncratic circumstances prevents a general criterion for exit rules. In the case of the Baltics (as well as Bulgaria and potentially longer-term Bosnia-Herzegovina), EU membership opened up (indeed, required by treaty) a natural and well-communicated exit option to another peg. conditional on meeting the (externally assessed) Maastricht criteria had to be met – clearly did not reflect a weakening of resolve.

The exit decision is more difficult for other currency board countries. In principle, countries can shift from a CBA to an even harder regime, such as "dollarization," the unilateral adoption of the anchor currency but with no say in the setting of monetary policy and no access to central bank LOLR operations. Transition to a more flexible exchange rate regime is likely to be tricky, particularly if the monetary authorities have yet to establish their credibility – difficult to do as the very nature of the CBA restricts the scope for discretionary monetary policy – or the exit is undertaken at a moment of stress. A pre-announced exit rule – for instance, stipulating a reconsideration of the exchange rate regime after a certain number of years meeting pre-specified performance criteria (or simply unconditionally) – could be envisaged. In practice, countries have generally not formalized exit rules. Reflecting the complexity and country specificity of the issue, we do not view the existence or absence of a formal exit clause as a practical component of a classification.

Near Currency Boards

Applying these criteria to real-world exchange rate regimes yields some unambiguous currency boards, some unambiguous non-boards, and a few intermediate cases, which might be termed near-currency boards. A group of small economies with close

economic ties to a large neighbor meet the peg and the cost-of-exit criterion, boast high de facto coverage ratios and largely eschew discretion but neither self-classify as currency boards nor impose high legal minima on reserves coverage. This group includes Lesotho and Swaziland (South African Rand), Bhutan (Indian Rupee), and, to a lesser degree, tourism-dominated Caribbean islands (US dollar) both individually and, in the eastern Caribbean, as a group. Within our case studies, Latvia comes closest to the dividing line between a harder fixed exchange rate and a softer currency board; reasonable arguments can be made for either classification.

Theoretical Aspects

A country's decision to adopt a currency board – as with any other exchange rate regime – depends upon the perceived economic and political costs and benefits, and thus on its specific economic, financial, and political circumstances. Conceptually, the decision can be broken into two separate choices: a choice to adopt a fixed rather than a floating exchange rate regime and, conditional on opting for a peg, the decision in favor of a currency board rather than a softer peg. In turn, having opted for a currency board, a further consideration is the strictness with which to adhere to the defining characteristics of a CBA, as reviewed in the previous section.

Fixed Versus Floating Exchange Rates

The choice between a floating and a fixed exchange rate ultimately comes down to the trade-off between the benefits of reducing exchange rate volatility and the costs of foregoing an independent monetary policy. The vast literature exploring this trade-off identifies three central criteria: the effect of the regime on insulating the economy from real and nominal shocks, its effect on the degree of integration with partner countries, and its role in enhancing policy credibility to help disinflate and maintain low inflation.

The early postwar consensus on pegged rates began to crumble in the face of the repeated balance of payments crises during the 1960s, prompting a number of economists to revive Friedman's case for exchange rate flexibility as adjustment tool in the face of balance of payments problems. This academic debate however found little resonance in official circles, where the prevalent view remained that abandoning the Bretton Woods system of fixed but adjustable parities risked a return to the interwar chaos of competitive devaluations and beggar-thy-neighbor policies.

In the event, neither view turned out to be quite right. While floating exchange rates among the major industrialized countries appeared to be driven more by capital flows than by underlying trade imbalances – rendering them less useful for orderly adjustment than their proponents had claimed – most countries also seemed to cope reasonably well with the higher volatility of nominal and real exchange rates, contradicting the warnings of fixed exchange rate advocates.

A large body of literature spurred by the influential papers of Fleming (1962) and Mundell (1962, 1963) advanced the understanding of the insulating properties of fixed and floating regimes for national income and output. While the original models concerned the efficacy of activist macroeconomic policies under alternative regimes, they could also – through a reinterpretation of policies as shocks – be used to evaluate the passive insulating properties of regimes. Used in this manner, the models implied that under conditions of high capital mobility, fixed exchange rates provide better insulation against nominal shocks, while floating exchange provide some insulation against real demand or supply shocks (see Ghosh et al. 2002, Chap. 3, for a review of the literature).

Beyond the short-run insulating properties of alternative exchange rate regimes, a second consideration concerns their effect on cross-border trade and investment. This was probably the uppermost consideration in postwar Europe, given the agenda of greater regional integration, which took on greater urgency as the Bretton Woods system of fixed exchange rates collapsed in the early 1970s. Intuitively, it might seem that the reduction – or elimination – of exchange rate uncertainty should foster more cross-border trade and investment, especially longer-term investments whose currency risks cannot be hedged in the derivatives markets. Yet, in the presence of divergent unit labor cost trends, lower nominal exchange rate volatility under fixed rates may well be associated with greater risks of persistent real misalignment. The empirical evidence is mixed. While exchange rate volatility does not appear to be a major impediment to international trade or investment at least among industrialized country, there is evidence that more durable fixed exchange rate regimes (including currency board arrangements, monetary union, or unilateral dollarization) are associated with substantially higher bilateral trade.

A third reason for adopting a fixed exchange rate is to assist in inflation stabilization or containment, notably when monetary authorities lack credibility or a track record. The basic logic is that a peg to a strong anchor currency allows participation in the credibility of the anchor currency central bank. Not only does the fixed parity immediately lower the inflation of the imported goods component of the CPI, it also helps anchor expectations, thus contributing to lower wage inflation and, ultimately, lower inflation across the consumption basket, including of non-traded goods. As discussed below, currency boards are particularly suitable for such disinflation (and, indeed, most modern currency boards were adopted with such motivation in mind).

Soft Versus Hard Pegs

Having chosen a fixed exchange rate regime, the central bank needs to decide on how "hard" a peg to adopt. The essential trade-off is between the credibility of the regime (which, in turn, will be reflected in how quickly inflation and interest rates converge to those of the anchor currency and how much confidence those engaged in cross-border trade or investment will have in the durability of the exchange rate commitment) and the ease with which the central bank can exit the regime in order to respond to shocks with discretionary monetary policy. The very soft end of the spectrum, such as a target zone or other rules-based intervention arrangement, gives

considerable scope for monetary policy to respond to shocks that result in an output gap. The downside is that the disinflation benefits of such a regime may be quite limited – precisely because the central bank faces few constraints.

The three defining characteristics of a currency board arrangement discussed in the prior section place it near the opposite end of the spectrum – almost on par with full dollarization. First, the domestic currency parity is fixed against the anchor currency within very narrow margins. Second, the reserve coverage rule means that the central bank should always have sufficient foreign exchange reserves to back its monetary liabilities (though seldom does it have sufficient reserves to cover the full banking system liabilities). Third, the institutional, legislative, and political basis of the currency board arrangement raises the economic and political cost of exiting the regime, not least as the very credibility of the board encourages the private sector to "embed" the anchor currency in domestic transactions (e.g., foreign currency-denominated borrowing and lending), making any eventual exit highly disruptive. Politically, a forced exit from a currency board is likely to prove highly embarrassing to the government.

Wolf et al. (2008) develop a theoretical framework to formalize the case for adopting a currency board by embedding a Barro-Gordon (1993) setup in an open economy setting. As is typical in such models, once nominal wages have been set, the central bank has the incentive to generate surprise inflation to erode real wages and stimulate output. Unless the central bank has some pre-commitment device, workers will anticipate this incentive and demand large nominal wage increases, which, ex post, will be validated by expansionary monetary policy. The economy thus inherits an inflationary bias (the same logic holds if investors anticipate an incentive for the central bank to generate surprise inflation to erode the real value of government debt).

Within this framework pegging the nominal exchange rate can provide the necessary pre-commitment device and deliver a domestic inflation rate equal to that of the anchor currency's inflation rate. Moreover, by anchoring inflationary expectations and raising money demand, the peg allows greater credit (and money) expansion for a given rate of inflation. While the peg can serve as a useful pre-commitment device when the central bank lacks credibility, it also constrains monetary policy from responding to output shocks. In the face of sufficiently large shocks, therefore, the central bank will have the incentive to abandon the peg and pursue a reflationary monetary policy. Inasmuch as workers (or investors) anticipate this incentive, the usefulness of the exchange rate peg as a pre-commitment device diminishes. This is where the hardness of the currency board arrangement relative to simple pegs comes into play, as the higher exit cost translates into a higher required output shock to tilt the cost-benefit trade-off toward exit. Since inflationary expectations are a probability-weighted average of the low inflation under the peg, and high inflation if the peg were abandoned, the currency board should thus deliver lower expected – and therefore actual – inflation than a simple peg.

The model allows consideration of the optimal "exit cost" of the regime (which is isomorphic to choosing the optimal probability that the regime will be abandoned) by looking at three cases: (i) the output shock is too small to warrant exiting either the simple peg or a currency board; (ii) the shock is such that the central bank would

exit a simple peg but not a currency board (with its higher cost of exit); and (iii) the output shock is sufficiently large that the central bank would exit either a peg or a currency board. In the first case, the simple peg results in higher expected inflation but the same (low) actual inflation as the currency board; hence real wages and real interest rates are higher, and employment, output, and welfare are correspondingly lower. In the third case, the higher expected inflation under the simple peg results in higher actual inflation when the peg (and the currency board) are abandoned (since discretionary monetary policy validates expected inflation), resulting in a greater welfare loss under the simple peg because the central bank penalizes inflation in its objective function.

In the second case, welfare is higher under the simple peg than under the currency board because the shock is large enough that the central bank would have preferred to exit the regime and respond to the shock with discretionary monetary policy but cannot exit the currency board due to the high cost of exit. The bottom line is unsurprising but worth emphasizing: when the central bank enjoys high credibility, and the economy is subject to large shocks, a flexible exchange rate is best. When the central bank is somewhat lacking in credibility, and the economy is subject to potentially large shocks, a simple peg would be appropriate – providing a pre-commitment device for monetary policy while allowing for an escape clause in the event of large shocks. Finally, when credibility problems dominate – perhaps in countries with a history of high inflation and failed stabilization attempts, or in newly independent countries where the central bank has not yet established a track record, or where political fragmentation is likely to result in pressures for monetary expansion – a currency board might be an appropriate disciplining device. As these circumstances are relatively rare, so are instances in which boards are the optimal device.

Theory thus suggests that when inflation is high, and the central bank lacks credibility to break entrenched inflationary expectations, a currency board can be a valuable tool for achieving rapid disinflation. But the very constraint upon monetary policy that makes the currency board useful for disinflation also constrains the ability of the central bank to respond to output shocks or banking crises (unless the central bank has excess reserves or can tap external lines of credit) and to build up a track record for responsible use of discretion that would ease an eventual smooth exit to a softer exchange rate regime. Moreover, while the currency board can help lower inflation by instilling monetary discipline, unless the underlying political, social, and economic forces underlying the high inflation are also tackled – specifically, unless the new monetary order is accompanied by corresponding fiscal discipline – the benefits and, indeed, the very viability of the regime may be undermined.

Case Studies

Currency boards date to the mid-nineteenth century but initially enjoyed only modest appeal (material in this section draws on the comprehensive analysis in Schuler 1992). Boards left their niche existence after the turn of the century: by the

1940s, almost 60 countries operated currency boards. Yet this new popularity proved fleeting. Viewed as part of the colonial regime, criticized as both overly restrictive and tilted in favor of Britain, and out of tune with the shift toward more active monetary policy, most boards gave way to fixed rates operated by central banks during the independence process. By the 1980s, fewer than 20 boards remained, concentrated in open economies and politically dependent territories (see, e.g., Jao (1974), Kwan and Lui (1996), Tsang (2000), and Latter (2004) for the Hong Kong CBA and Hanke and Sekerke (2003) for the St. Helena CBA).

The (re)adoption of a board in Argentina – a sizable and comparatively closed economy – in the early 1990s revived the arrangement. In subsequent years, boards were seriously considered in a range of countries (Hanke and Schuler 1994) and adopted in five: Bosnia and Herzegovina, Bulgaria, Estonia, Latvia, and Lithuania. Yet while the new boards shared some of the basic characteristics with their predecessors (see Enoch and Gulde 1997 on the technical aspects of introducing CBAs), they differed fundamentally in others – notably in having stabilization rather than trade facilitation as their core motivation. In the next three sections, we take a closer look at the British colonial boards, the (second) Argentinean board, and the Euro boards.

The British Colonial Currency Boards

We begin with a look at the early currency boards, the subject of both contemporary studies (Chalmers 1893; Clauson 1944; Sidney Caine (Special Correspondent) 1948–1950; Greaves 1950, 1953; Newlyn and Rowan 1954; Loynes 1962) and more recent assessments (Schuler 1992; Schwartz 1993; Hanke and Schuler 1994; Narsey 2016). Dominating the sample – and the focus of our discussion below – were the currency boards in the colonies and dependent territories of the British Empire in the first half of the twentieth century, mostly based on the template created by the West African Currency Board (Clauson 1944; Newlyn and Rowan 1954; Hopkins 1970; Loynes 1962; Vice 1983; Schenk 1997).

But there were other notable early boards as well. The first Argentinean Board (1890–1935) was motivated by both trade facilitation and credibility concerns in the wake of the Baring crisis (Paolera and Taylor 2001). The 1927–1943 Irish board relying on private note issuers (examined in Moynihan 1975; Schuler 1992; Honohan 1997; Lau 2000; Thom and Walsh 2002) eventually gave way smoothly to a de jure central bank regime though, at least initially, monetary policy changed little. Other boards included North Russia (Hanke and Schuler 1991), the Philippines, Panama, Danzig, Italian Somaliland, Libya, Sudan, North Yemen, Swaziland, and Lesotho. Schuler (1992) provides background on all historical boards.

From Rarity to Near Default Arrangement: The West African CBA as Template

Though displaying significant variation, most pre-CBA colonial monetary systems combined domestic and foreign circulating coins and to a varying extent banknotes, including domestic issues. Domestic financial system depth varied widely (Special

Correspondent (Sidney Caine) 1948/1949/1950 and Sayers 1952 provide mid-century perspectives, a comprehensive set of references can be found in Gross et al. 2012).

International transactions, notably trade finance, were largely performed by London-based Imperial Banks (Baster 1929; Newlyn and Rowan 1954; Crick 1965) described by Baster (1929, 120) as "a compact and homogeneous group, connected among themselves or with the English domestic banks, by a comprehensive system of interlocking directorates." For the West African Currency Board, the function was fulfilled by Bank of British West Africa, initially represented on the board itself.

Economic growth and increasing trade translated into rising demand for (primarily silver) coins as well as banknote issues. While the demand for coins was satisfied by London (Special Correspondent (Sidney Caine) 1948), the potential for a flow reversal caused unease, prompting debates about alternative arrangements. In 1912, a parliamentary committee report suggested a shift to local currency systems with convertibility into Sterling. The West African Currency Board reflected the proposal and soon became a template for a proliferation of similar arrangements (Hopkins 1970; Schuler 1992). In contrast to modern boards reflecting domestic concerns, the creation of colonial boards was substantively driven by the interest of Britain, with limited domestic input in their creation and their running (Narsey 2016). Their resulting close identification with colonialism reduced their political viability post independence.

The institutional structure was quite frugal. The West African Currency Board (discussion here and below draws on Clauson 1944; Newlyn and Rowan 1954; Hopkins 1970; Loynes 1962; Schenk 1997) and most of the subsequent CBAs were based in London, typically complemented by local offices. Staffing – generally excluding representatives of the colonies – centered on a secretary assisted by generally part-time staff delegated from other institutions, primarily the Colonial Office and the Crown Agents (Abbott 1959; Sunderland 2004), contributing to similar views and approaches across boards. Given their minimalist staffing, boards relied on outside agents for operational support. Initial backing was provided by redeemed circulating monies; holders of the new notes could convert into Sterling in London and at local offices.

Boards took the overarching structure – set by the Colonial Office in consultation, though not always full agreement, with the Bank of England and the Treasury – as given. While enjoying some de jure flexibility in day-to-day operations, they usually hewed closely to a conservative strategy that emphasized full and often excess coverage, held in highly liquid assets.

While the composition of reserves constrained total returns, the bare-bones operation also limited costs, translating into flow profits. Near its end, the West African Currency Board in 1954 recorded costs of £0.08 m against income of £0.76 m, yielding a modest return on assets of 2.2% (Newlyn and Rowan 1954: 50–53). The allocation of profits emerged as a point of contention (Narsey 2016). Boards retained a sizable share of earnings to bolster coverage ratios; calls by the

colonies for a greater share mostly went unheeded. In the case of the West African Currency Board, coverage reached 114% in 1954 (ibid.).

From Routine Arrangement to Curio Cabinet

The colonial CBAs delivered on their prime objective of exchange rate stability from the 1920s through the 1950s. They did so against a global backdrop of instability during this period, including a bevy of hyperinflations. Yet few survived the following decades; by the 1980s, boards had again become a niche arrangement. The demise of the colonial CBAs was driven by both politics and changing economic fashions (Loynes 1962; Lettiche 1974; Schuler 1992; Uche 1997; Schenk 1997; Narsey 2016, inter alia). The decision to restrict local voices from the initial conception and subsequent operation of the boards, as well as perceptions that the arrangement favored Britain at the expense of the domestic economy, made them unpalatable in the postindependence political climate. As Schenk (1997: 197) argues, the "*introduction of a new national currency was an important psychological and political motive for establishing central banks since colonial currencies were discredited by allegations of colonial exploitation*."

Economic criticisms reinforced the political malus. In macroeconomics, the fashion was shifting against rule-based systems and in favor of active and broad-ranging monetary policy. Development economists highlighted the opportunity cost of maintaining full coverage in low-yielding foreign assets when domestic investment promised high returns. Operationally, boards were seen to aggravate seasonal volatility and exert deflationary pressures. Together, the association with colonialism, shifting policy views and specific operational challenges resulted in a rapid decline in the number of boards during the 1960s and 1970s as newly independent countries transitioned to central banks.

At least in the immediate aftermath of the transition, most of the new central banks retained both the fixed exchange rate and the high coverage ratio (in line with external advice (Hazlewood 1952, 1953; BIS 1953; Schenk 1994)), facilitating smooth transitions. The dissolution of the West African currency board is a case in point. Ghana was the first to leave in 1958, followed by Nigeria in 1959, Sierra Leone in 1961, and the Gambia in 1971. For members leaving the arrangement, the board exchanged tendered notes and coins for Sterling and distributed part of the excess reserves, thus creating the initial reserve basis of the new central banks. Having completed the exchange for the last member, the board closed down in 1973.

What lessons can be drawn from the operation of the boards and their eventual replacement?

Durability and Credibility

Boards maintained full and typically even excess coverage, primarily in highly liquid assets, providing assurance of convertibility. Credibility was further boosted by an expectation that the Treasury (as well as, within limits, the Imperial Banks) would – though not obligated to do so – provide support in case of monetary distress, not least to prevent contagion across the system of boards.

Cui Bono?

Were colonial boards well-suited monetary and exchange rate arrangements for economies with (in most cases) still fairly small financial systems? Or were they mainly geared to serving British interests, possibly hampering local economic and financial development? The issues were explored in a lively contemporary literature (Greaves 1950, 1953; Hazlewood 1952, 1953; Newlyn and Rowan 1954, inter alia, Treadgold 2005 for a discussion) and have recently attracted renewed attention (Schuler 1992; Schenk 1997; Treadgold 2005; Narsey 2016, inter alia).

Not surprisingly, the contemporary view from London tilted strongly toward the former assessment. Calls for a shift to central banking were seen as premature, risking instability and endangering capital flows required for sustained growth. By the middle of the twentieth century, views from the colonies and former dependencies had become more critical, highlighting both constraints on monetary policy writ large and specific implementation issues.

Reserve investment policy in particular was actively discussed (Hazlewood 1953; Newlyn and Rowan 1954; Treadgold 2005; Narsey 2016). Critics, including colonial administrations, saw the preference for highly liquid (and thus low-yielding) assets, and the retention of some of the (correspondingly modest) profits to maintain excess reserves as a safety buffer as being unnecessarily restrictive, depriving the colonial administrations of seignorage revenue. Investment choices were also argued to be influenced by a desire to support the London capital markets by using currency board reserves as a captive source of demand (Narsey 2016).

The criticism has some merit. For one, the very concentration of reserves in highly liquid assets reduced the risk of large capital losses and thus the need for excess reserves as a buffer against such losses. Yet, while boards may have erred on the side of caution, seigniorage constituted a modest share of total revenues (Schuler 1992; Treadgold 2005), placing a limit on the additional flow revenues generating by even a sizable rise in average returns. The seasonality of money demand (Schenk 1997; Treadgold 2005) furthermore necessitated a substantial portion of reserves to be held in liquid form, constraining the scope for reallocation (in the West African Currency Board, money demand rose in autumn and winter as traders paid for the cocoa and groundnut crops in cash, before declining in spring (Clauson 1944)).

Reductions in the reserve coverage ratio would have provided more significant one-time resources, at the potential cost of reduced credibility, in turn impairing capital inflows (and thus having a negative second round effect on seigniorage Treadgold (2005)). In the event, caution prevailed. An (albeit late) 1954 offer by London for a (modest) fiduciary issue was not taken up (Schenk 1997), while many new central banks opted to maintain full reserve coverage (some, including Nigeria, were mandated to retain full reserve coverage for an initial period (Schenk 1997)).

A subplot of the discussion concerned whether reserves should include assets issued by (other) colonies. The argument in favor was threefold: additional demand for colonial securities might have reduced borrowing costs, promoted financial sector development, and increased returns on reserve assets, albeit only modestly. Yet the inclusion could have also had a potential downside, as local fiscal troubles

could have led to contagion effects across boards through impaired holdings. In the event, calls for an expanded range of holdings had limited effects on reserve composition, which remained heavily tilted toward British assets, reinforcing the charge that the structure of boards reflected British rather than local interests (Narsey 2016).

A second operational criticism focused on the challenge of meeting the trend rising demand for money, leading to fears of structural deflationary pressure. The evidence on this point is mixed. In a set of case studies, Treadgold (2003, 2006) finds that money base increases stemming from balance of payment surpluses together with rising velocity and money multipliers offset the potential deflationary effects. He concludes that "[t]aken as a whole, these studies of colonial currency boards clearly failed to offer much empirical support for the deflationary bias hypothesis" (Treadgold 2003: 64), though the Philippine board may have been an exception.

Finally, critics saw currency boards not as a step toward greater financial development, but rather as a barrier (Narsey 2016), pointing to the constraints on local banking growth, limits on the issue of local securities, and the outsized role of the London-based banks. In theory, fixed exchange rate regimes provide both supportive and inhibiting effects on financial sector development. In our view, the heterogeneity of financial system depth prior to the introduction of CBAs, and the presence of other impediments, including the occasional scarcity of easily attachable collateral, cautions against any general conclusion. Boards introduced atop already well-developed financial systems may well have exerted a restraining influence. Determining whether this outweighed the benefits of close integration with the London financial markets requires careful country-specific assessments.

The Big Picture

While CBAs delivered on their primary goal of exchange rate stability and trade facilitation; this came at the cost of tying up substantial funds in reserves and in some cases may have slowed financial sector development and diversification. Would alternative exchange rate regimes have delivered better results? In our view the heterogeneity of the (at the peak more than 50) boards suggests that the question is best answered through case studies of individual boards.

An indirect approach considers relative performance under the boards and in the post-board period. If boards indeed constrained growth and more flexible regimes would have delivered comparable stability, then the transition should – ceteris paribus – be associated with an improvement in real performance but comparable nominal outcomes. Contrasting performance under the successor regimes that allowed for more discretionary monetary policy with those under currency boards, Schuler (1992) documents a pattern of higher average inflation and lower average growth, often associated with depreciation against Sterling. While the unconditional relative performance is suggestive; here again country case studies controlling for idiosyncratic features and developments promise additional insights on causal patterns (Gross et al. 2012 provide references).

Argentina

Proponents of currency boards stress the low-inflation/decent-growth experience of most currency board countries. Opponents counter with a single word: Argentina. But disentangling the drivers of Argentina's economic performance between the adoption of the Convertibility Law in 1991 and the default-cum-devaluation in early 2002 is not straightforward, featuring capital flows, commodity prices, and fiscal dynamics as well as the currency board proper, with complex cross-linkages.

Alongside most commentators, we view Argentina's arrangement (Convertibility Plan) as being at the soft end of the currency board spectrum, both in its de jure conception and in the de facto implementation (indeed, Hanke 2008 argues that the degree of monetary policy flexibility engendered by the institutional arrangements disqualifies Argentina as an (orthodox) currency board). Three features set it apart. First, eligible reserve assets initially included up to a third of FX-denominated Argentine government bonds (rather than genuine foreign reserves). Second, inflows and outflows were at times sterilized, while the central bank conducted discretionary expansions and contractions of its net domestic assets. The link between changes in net foreign reserves and changes in the monetary base thus deviated from unity, at times substantially. Third, the central bank altered banks' required reserve ratios and, on occasion, acted as lender of last resort.

Despite being relatively soft, there is little doubt that the board was instrumental in breaking Argentina's long cycle of high (at times hyper)inflation and failed stabilization efforts. Inflation, which had peaked above 3000% in 1989 and was still running at over 2000% in 1990 dropped to 170% in 1991, and less than 25% by 1992. Nor did the stabilization come at a high real cost: consistent with the general pattern of the immediate aftermath of exchange rate-based stabilizations, real GDP growth – which had been deeply negative during 1988–1990 – shot up to more than 12% in 1991 and 1992 and continued at almost 6% in both 1993 and 1994.

Yet as Argentina had successfully stabilized before only to revert to high inflation a few years later, observers could be forgiven for demanding durability: the proof of the pudding for the new regime rested in *sustaining* rather than achieving stabilization. Observers did not have to wait long for the first resilience test, arriving in the wake of the Mexican "Tequila crisis" of late 1994. Although Argentina suffered both a capital flow reversal and deposit runs on the still-fragile banking system, the currency board arrangement survived. By 1996, both capital inflows and rapid growth had resumed. The experience enhanced the credibility and appeal of the CBA both at home and abroad – serving as a model for other countries contemplating stabilization from hyperinflation, including Bulgaria and several other countries considering, but eventually not adopting, boards.

But fresh challenges – again originating abroad – were just around the corner. Contagion from the East Asian crisis in 1997 and the Russian default of 1998 sharply curtailed capital flows to emerging markets, in turn constraining growth. Argentina's terms of trade deteriorated while the appreciation of the US dollar against the Euro undermined competitiveness in a key export market. Brazil's 1999 decision to devalue aggravated these pressures.

External shocks were exacerbated by domestic policy choices. Argentina did not break decisively from its history of fiscal deficits that had been at the root of previous episodes of high inflation and failed stabilization programs (Daseking et al. 2005). The public debt-to-GDP ratio climbed steadily from 30% in 1992 to 40% in 1999, with the debt-to-export ratio eventually exceeding 500%. Yet there was little sense of alarm. Just as in the Baltic experience reviewed below, the long initial boom amid stability contributed to a degree of policy complacency even in the face of growing fragility.

When boom gave way to recession in the wake of multiple external shocks, the authorities faced a choice between pursuing internal and external devaluation. In the event, there was limited appetite for the dramatic fiscal stabilization measures and nominal wage adjustments the Baltics were to pursue after 2008. As the crisis deepened, investors' confidence in the permanence of the exchange rate and the associated convertibility promise weakened. Depositors, well aware of the dynamics from previous high inflation episodes, withdrew funds, setting in motion a vicious circle. The sudden stop in the refinancing of maturing government debt accelerated the run, ultimately resulting in the corralito (a suspension of deposit withdrawals), default on outstanding government debt, and collapse of the currency board followed by an immediate depreciation of the Argentine peso.

Lessons

What can be learned from the Argentine case? Should the currency board be assigned a star role in these developments, casting doubt on its use in larger, more closed economies? Or was it no more than a supporting actor in a drama dominated by political and in particular fiscal choices? We see several lessons.

An Ounce of Prevention Is Worth a Pound of Cure

Both the Argentinean and the Baltic experiences illustrate the dangers of being lulled by the initial good times into inaction and failing to actively monitor and counteract emerging challenges in real time. The success of the currency board in stabilizing the economy (and in surviving the Tequila crisis) prompted capital inflows that relaxed fiscal funding constraints but in so doing also raised the risks of capital flow reversals. Early intervention could have limited the extent of the exposure.

Despite a record of past stabilizations that had failed because of a lack of fiscal discipline, the honeymoon period following the adoption of the currency board dulled any sense of urgency. Indeed, the very success of the currency board in reducing inflation and surviving the first external test allowed the Argentine government to continue issuing dollar-denominated debt (De la Torre et al. 2003). These sins of omission and commission raised the cost of the eventual crisis.

The lesson is of course neither new nor unique to Argentina; indeed, it is striking that even having observed Argentina's painful crisis and exit, the Baltics tolerated the buildup of external liabilities only to eventually encounter a similar crisis – albeit it with a different ultimate outcome.

Returning to the star/supporting actor distinction, while the successful introduction of the board led to the boom, the root causes of the failure to adjust in time have

to be sought in the political economy of (non-)reform. In this regard, the board functioned as a supporting actor.

Tying Your Hands Is Costly

Argentina's experience relates to the age-old debate of rules versus discretion. Rules help provide credibility, particularly when policymakers themselves lack credibility in the face of a history of high- and even hyperinflation. Yet rules have costs. Indeed, the very feature that generates the credibility bonus for currency boards is their high exit cost, both political (devaluation is an open admission that the government has failed to keep its promises) and economic (not least the negative balance sheet effects stemming from widespread liability dollarization, in Argentina's case including both public dollar-denominated debt and FX-denominated home mortgages). Adopting a currency board implies accepting that, should it fail, the bill will be high. While currency boards may be a powerful tool for disinflation and for achieving monetary credibility, they may thus be unsuitable for economies that lack sufficient adjustment flexibility to respond to changing circumstances and countries with weaker fiscal institutions (in the case of Argentina, the central government was never able to get a firm grip on provincial finances, Perry and Servén 2003).

Indeed, once a country has a public debt problem, adjustment under a currency board (or any other fixed exchange rate regime) becomes well-nigh impossible. Adjustment without a change in the nominal exchange rate requires price deflation – but such deflation both increases the real burden of the public debt and typically widens the fiscal deficit, thus worsening the public debt dynamics. In the case of Argentina, the external shocks and sudden stop in the late 1990s necessitated external adjustment, but any gains in competitiveness through price deflation simply worsened the sustainability of the public debt problem. It was this fundamental inconsistency that doomed the country's (IMF-supported) adjustment efforts – and provides a salutary lesson to other countries contemplating hard pegs.

Do (Not!) Let Sleeping Dogs Lie

Not least, Argentina illustrates the risks of eschewing a well-specified and well-communicated exit strategy. By nature, a government introducing a board to counteract a hyperinflation caused by fiscal and monetary misdeeds faces severe credibility challenges if it were to simultaneously introduce an exit strategy involving a return to policy discretion. Nor is the successful operation of the board itself (in normal times) likely to deliver much in terms of credibility for a return to discretionary policy, after all, its very modus operandus involves the exclusion of discretionary choices and thus precludes the creation of a track record.

As long as the good times roll, discussions of an exit from the CBA based on its own success are thus intrinsically tricky, as the public has no basis to evaluate whether a relaxation of the constraint would lead to a continuation of sound policy or to a return to the monetary and fiscal challenges that prompted the introduction of the board in the first place. The credibility required to open an exit debate thus derives substantially from outside the board proper and, specifically, from the observed willingness to adopt other policies supporting the continuation of the

board (and thus the stabilization), most notably in the fiscal realm. The signal is more cheaply sent during the initial boom period – an opportunity missed by Argentina – more expensively during a crisis, successfully by the Baltics, attempted but ultimately unsuccessful by Argentina.

While the decision not to commence a formal debate on an (unforced) exit is understandable during the initial honeymoon period, it has to be recognized that fundamental uncertainty about the duration of the regime by nature impairs credibility, if only because it broadens the expected range of possible future shocks that might prompt a forced exit. The Baltics (as well as Bulgaria and, in the longer term, probably Bosnia-Herzegovina) were (are) lucky in not having to face the challenge, as the (mandated) exit to the Eurozone subject to meeting tough convergence criteria presented a natural, externally assessed exit into another hard peg. Argentina could not rely on a similar external option.

The dilemma raises the question whether an exit clause should be embedded in the original board design for countries not enjoying a natural exit option. While raising the very issue of exit may be detrimental to initial confidence, a clause spelling out tough performance criteria (including in the fiscal realm) over an extended period may ease the challenge of beginning to talk about an exit after the board has been successfully in place for some time. Indeed, specifying fiscal criteria may even counteract the temptation not to tackle fiscal consolidation during the honeymoon period, reducing risks.

There Is No Free Lunch

While the decision not to tackle the long-standing fiscal issues was facilitated by capital inflows, it ultimately reflected domestic political choices. Once the crisis struck, the board took on a more central role. The general choice between external and internal devaluation applies to all pegged regimes, yet it is accentuated for boards through the higher exit costs in both the political and – via liability dollarization – the economic realm. The Argentinean experience illustrates that the additional credibility is thus not a free good: the upside during the good times is precisely bought with the higher cost of exit in (truly) bad times. This of course links to the first lesson, emphasizing the need for early intervention to counteract emerging vulnerabilities and for pursuing consistent fiscal policies to minimize the chance of encountering circumstances forcing a costly exit. Ultimately, a currency board provides a means to administer a societal consensus to stabilize based on an agreed distribution within a feasible budget constraint (the Euro boards); it cannot compensate for a lack of such agreement – the Argentinian case (Calvo et al. 2003; Calvo and Talvi 2005).

The Euro Boards

Argentina's 1991 Convertibility Plan put currency boards back in the spotlight. Routinely discussed in countries experiencing monetary upheaval, they were eventually adopted by five European countries: Estonia in 1992, Lithuania and Latvia in

1994, and Bulgaria and Bosnia-Herzegovina in 1997. The popularity of currency boards in these European transition economies reflected the perceived need for establishing economic stability during a period of profound volatility. Though similar in many respects, the specific challenges prompting the adoption of currency boards differed across the countries: civil conflict in Bosnia-Herzegovina; stabilization in Bulgaria; new currencies and the absence of a monetary policy track record in the Baltics. Accordingly, the institutional design characteristics differed as well.

The Baltic Boards

Notwithstanding different histories and cultures, the three small, highly open Baltic Republics share significant similarities in their economic experience prior to choosing a currency board. Gaining independence in 1991, they faced the challenge of managing the transition from central planning to a market economy. The choice of monetary standard and exchange rate regime was only one element in the overall strategy, albeit an important one (Berengaut et al. 1998). As retaining the Russian ruble was not an option, either politically or economically (Schoors 2003), national currencies were established in close succession (Estonian Kroon 1992, Lithuanian Litas and Latvian Lat 1993) and pegged to an external anchor. On the following pages, we take a closer look at the individual experiences before drawing them together.

Estonia

Following independence, the Estonian economy faced severe monetary difficulties, reflecting the rapid dissolution of the ruble zone. The political imperative of introducing a national currency by the re-established Bank of Estonia (BoE) was uncontroversial; the choice of exchange rate regime was not (Knobl et al. 2002; Sepp and Randveer 2002). A debate pitting the benefits of exchange rate flexibility in a rapidly changing economic environment against the credibility advantages of a peg was eventually resolved in favor of the latter – and indeed, of a hard peg in the form of a currency board.

The decision was eased by the return of Estonian gold reserves that had been held for safekeeping since World War II at the Bank of England and elsewhere, providing backing for the new currency. Risks – including potential financial sector instability arising from foregoing the lender of last resort (LOLR) option and real exchange rate misalignment – were discussed in detail but thought to be manageable given wage flexibility and the prospect of productivity gains. The currency board introduced in June 1992 fixed the Kroon against the DM at 8:1. The board veered toward the orthodox side, enshrined in law and stipulating full reserve coverage, though it permitted accumulation of excess reserves and thus allowed for limited LOLR functions in the future (Bennett 1992).

While the inflation rate plummeted, it remained higher than that of Germany. so, however, did productivity growth, mitigating competitiveness concerns (an initial undervaluation helped). Financial sector instability associated with the transition became acute in 1992 and again in 1994 but proved manageable (Korhonen 1999). By the late 1990s, subsidiaries of foreign (mainly Scandinavian) banks dominated

the banking system. The Russian crisis of 1998 briefly interrupted, but did not derail, the steady progress in economic restructuring, culminating in the 2004 accession to the EU. Membership also opened up a viable exit option into the Eurozone, vigorously pursued through an almost immediate entry into ERM-II (Gulde et al. 2000; Lättemäe and Randveer 2004). In the aftermath, good times rolled. As FDI and capital inflows soared, the economy followed (Bakker and Gulde 2010). Strong performance was largely seen as the just reward for a decade of prudent policy; signals of potential overheating and the associated risks received limited attention.

Boom turned to bust in the wake of the 2008 financial crisis as a sudden stop of capital inflows exposed the vulnerabilities. By 2009, the economy was in full crisis, suffering a double-digit GDP decline. The sharp economic deterioration returned the spotlight to the 1992 debate between flexibility and rules. Should Estonia exit the board, allowing stimulation through depreciation but at best delay Eurozone membership? Or should Estonia retain its currency board and adjust through internal devaluation? In the event, the debate proved short-lived, reflecting a strong domestic consensus in favor of the currency board arrangement and through it, of Eurozone membership (Purfield and Rosenberg 2010). Commitment met action as a determined fiscal adjustment coupled with wage flexibility achieved the necessary internal devaluation and external adjustment. The willingness of the foreign bank parents to limit capital repatriation provided additional support. By 2010 growth returned; by 2011, it surpassed 7%. With the fiscal balance in order and inflation subdued, Estonia met the convergence criteria in 2010, entering the Eurozone in 2011.

Lithuania

While Estonia decided early on to adopt a board and stuck to the course, Lithuania and Latvia followed more convoluted paths, though eventually arriving at the same destination – Eurozone membership (Alonso-Gamo et al. 2002). Lithuania initially responded to the problems in the ruble zone by a supplementary coupon currency – the Talonas – providing a respite for the design and introduction of a national currency, the Litas, in the spring of 1993. The domestic debate centered on the trade-off between the credibility promised by a currency board and the flexibility provided by a peg allowing for some traditional central bank operations.

Ultimately, the former view prevailed, Lithuania adopted a currency board in 1994, pegging the Litas to the Dollar, at the time the dominant foreign currency in circulation and invoicing (De Haan et al. 2001). The institutional setup veered to the softer side of the spectrum. The board was anchored in a separate act rather than in the central bank law itself; the parity was to be set by the government in consultation with the Bank of Lithuania, which was granted explicit permission to use excess reserves and some money market instruments to undertake limited liquidity operations.

In the early years, Lithuania shared Estonia's experience of inflation above that in the anchor currency and banking system problems, both largely attributable to the travails of the transition to a market economy. In sharp contrast to Estonia's unwavering commitment to a board, however, Lithuania revisited its choice of exchange rate regime as the initial volatility gave way to a more tranquil

environment. This time, proponents of more flexibility prevailed. The government initiated technical preparations for a transition to a more activist central bank, envisioned for 1999. The Russian crisis of 1998, resulting in both economic recession and political uncertainty, prompted second thoughts, eventually leading to a recommitment to the currency board arrangement supported by deep fiscal adjustment. Four years later, with tranquility restored, Lithuania uneventfully switched the anchor from the Dollar to the Euro, reflecting the trade orientation of the economy, concerns about Dollar appreciation, and, not least, the approaching EU membership, attained in 2004.

Following EU accession, Lithuania participated in the Baltic boom, eventually veering into overheated territory. Concerns did not gain much traction as productivity gains and flexible economic structures appeared to allow for adjustment. The 2008 financial crisis ended the good times; by 2009 GDP was in steep decline, accompanied by banking system distress. With a Eurozone exit option in sight, Lithuania reconfirmed its commitment to the currency board, embarking on an adjustment strategy centered on sharp fiscal adjustment, supported by the willingness of foreign banks to maintain exposure. While the responses addressed the acute challenges, the legacy of the boom years and the crisis proved more lasting, delaying Lithuania's ability to meet the convergence criteria and transition to the Eurozone until 2015.

Latvia

Like Lithuania, Latvia initially introduced a coupon currency, followed by the Lats in 1993, pegged first to the SDR in order to minimize exposure to bilateral exchange rate movements, followed by a peg to the Euro in 2005. While lacking the formal legal basis of the other Euro boards, the Bank of Latvia in practice followed currency board rules including full (indeed, excess) reserve coverage (Repse 1999). Reflecting the long operation under currency board rules, most analysts treated the arrangement as a "quasi currency board" (Bakker and Gulde 2010), though it could also be seen as a voluntarily extra-hard version of a conventional peg (Twarowska 2014).

The economic trends mirrored those in Estonia and Lithuania, with initial high inflation and growth, a (milder) disruption by the Russian crisis and subsequent strong growth as well as rising capital inflows (partly into domestic bank deposits) that fueled inflation of goods and asset prices, notably of real estate. Boom gave way to sharp recession during the 2008 global financial crisis, accompanied by distress among domestically owned banks.

The severity of the decline required external support through a large joint IMF/EU program (one of the co-authors, Anne-Marie Gulde, was part of the IMF team for the 1997 Bulgaria and the 2009 Latvia programs). Despite intense debate and significant doubt voiced by external observers, the government showed unwavering commitment to the currency board and choose to embark on internal devaluation supported by fiscal adjustment, including significant public wage cuts (Aslund and Dombrovskis 2011). As in Estonia and Lithuania, the adjustment proved successful in correcting the imbalances while retaining the peg (Blanchard

et al. 2013). The economy returned to growth in 2010. Three years later, Latvia met the convergence criteria and entered the Eurozone in 2014.

Bulgaria

Bulgaria embarked on its transition to a market economy under a managed floating regime. Regulatory and governance failures during the early privatization and liberalization period contributed to banking sector problems in the mid-1990s (Poirot 2003). The costs of bank closures aggravated an already weak fiscal balance, subsequent monetary financing prompted a descent into hyperinflation. Not dissimilar to the Argentine case, rescue was sought in a 1997 exchange rate-based stabilization anchored by a currency board – initially pegged to the DM, switched to the Euro in 1999 – and supported by the IMF. The board became possible in part because hyperinflation had eroded domestic liabilities of both the government and the central bank (Gulde 1999).

The structure of the board tilted toward the soft side of the spectrum (Nenovsky and Hristov 2002). Backing was on a gross rather than on a net basis, initially assured through the reserves (including gold) and the IMF loan. For practical reasons, the board was created by a redesign of the existing central bank, with the issue department responsible for assuring full reserve coverage. Two safeguards provide limited flexibility: a fiscal reserve fund held in the issue department allows for short-term smoothing, while the banking department holds excess reserves (initially a substantial 30% of banking system assets) assuring potential liquidity support for domestic banks.

The board delivered on stabilization and withstood early challenges. By early 1998, inflation and interest rates had converged close to those of the anchor currency. Growth rebounded, only temporarily slowed by the Russian crisis and the Kosovo conflict. Culminating a long transition process, Bulgaria joined the EU in 2007, thus – like the Baltics – gaining a viable exit option. Partly reflecting its later EU membership, partly its different economic and financial structure, Bulgaria neither enjoyed the dramatic upswing before the 2008 financial crisis nor suffered as severe a decline in its aftermath, though concerns about the banking system were rekindled after a 2014 collapse of a sizable lender. Two decades after its adoption, support for the board, credited with achieving and sustaining macroeconomic stability, remains strong. An exit to the Eurozone in the intermediate future appears likely.

Bosnia–Herzegovina

Bosnia and Herzegovina emerged from bitter conflict. The 1995 Dayton Agreement set up both a jointly run government with a rotating presidency and two separate entities, the Federation of Bosna-Herzegovina and the Bosnian Serb Republic, each granted far-reaching autonomy (Coats 2007). The agreement mandated the establishment of a joint central bank (Central Bank of Bosnia and Herzegovina (CBBH)) and the creation of a currency board. While economic volatility was substantial, political considerations played the decisive role, as the rule-based structure preempted potentially divisive political choices. The desire to reduce domestic tension

also prompted the appointment of a noncitizen as governor in a selection process that included the IMF.

The CBBH began operations in August 1997. Reflecting the desire to minimize (the need for) discretionary policy choices, the board was set up in orthodox and highly transparent fashion, based on a new currency – the convertible marka – tied one-to-one to the Deutsche Mark (widely circulating in the region) and including a full coverage requirement (in practice, the CBBH operates with excess reserves) (Kovacevic 2003; Kamhi and Dehejia 2006).

The arrangement proved suitable to the complex political setting. While only mandated for an initial period, it reached its twentieth anniversary in 2017 (Central Bank of Bosnia and Hercegovina 2017). The primary objective of exchange rate stability (with the anchor now switched to the Euro) has been attained, and the arrangement has proven adept in responding to volatile capital flows, partly related to the entry of foreign banks.

Lessons from the Euro Boards

A quarter century after the initial adoption in Estonia, what can be said about the experience of the Euro boards? Evaluated against their core objective of exchange rate stability, the Euro boards have been an unqualified success (Hanson and Randveer 2013; Staehr 2015; Wolf 2016). None has abandoned the peg. The three Baltic states have successfully transitioned to another hard peg after operating boards for two decades. Bulgaria is approaching ERM-II membership en route to joining the Eurozone. Though Bosnia-Herzegovina has no short-term exit option to the Eurozone, the formal EU membership application in 2016 opens up a long-term perspective.

Political buy-in and popular support were key to the success of the boards, supported by adherence to formal rules. The buy-in was not solely rooted in economics, but to an important degree reflected political considerations: independence from the Soviet Union in the Baltics and the desire to avoid a recurrence of domestic conflict in Bosnia-Herzegovina.

Yet the experience of the Euro boards was not all sunshine. In the Baltics, the initial success led policymakers to lower their guards. While continuing to operate within the mandate of the boards, authorities did not energetically confront the buildup of vulnerabilities – notably capital inflows and domestic exuberance in financial and real estate markets – that were suddenly realized in the aftermath of the 2008 global financial crisis. The sharp 2009 recessions raised the core question endangering the credibility of boards: are there circumstances under which the government will give up the board? The Baltics faced a simple choice. Should they adopt more expansionary measures (including exchange rate devaluations) promising to mitigate the immediate crisis at the cost of exiting the board? Or should they pursue – in the near-term more painful – measures required to retain their exchange rate arrangements and their exit option to the Eurozone?

Though facing calls to embrace the former option (not least by external observers) all three instead opted for internal devaluation, putting credibility concerns to rest. In spite of pervasive skepticism, the experience bore out the expectation that internal

devaluation in highly open and flexible economies with strong policy commitment and clear exit options can work: the recessions proved sharp but short, followed by strong rebounds. Not all of the credit belongs to the domestic actors, however. The support of the Scandinavian parent banks and, in the cases of Bulgaria and Latvia, external support programs sharply mitigated one of the main downsides to hard pegs – the central bank's inability to undertake domestic lender of last resort operations.

Conclusion

Countries operating under CBAs fall into two distinct groups. The first employs CBAs as (semi)permanent arrangements aimed at facilitating trade. The evidence strongly suggests that the group has delivered on the exchange rate stability objective. Whether CBAs were (or remain) the *optimal* exchange rate regime depends on the perceived gains and costs of shifting to a different regime, the subject of continuing debate, notably regarding the effect of CBAs on financial sector growth. For countries choosing a different regime, the transition from boards to central banks has generally been smooth, not least reflecting the continued full reserve coverage in the immediate aftermath. In our view, trade-motivated boards remain useful for the subset of countries benefitting from the concurrent use of the domestic and the anchor currency, notably very small, very open tourism-focused economies and dependent territories. These are, indeed, dominating the small list of current, typically more orthodox, trade-motivated CBAs.

A second group of countries adopts boards as a temporary regime to achieve stabilization. For this group, the core lesson is straightforward and unsurprising. The introduction of boards generally delivers short-term stabilization and an extended honeymoon period. Whether the initial stability proves lasting depends on policy choices during the honeymoon period. Countries aligning fiscal policies with a fixed exchange rate and responding proactively to incipient fragilities (notably the linked trends toward real appreciation and capital inflows) can use a CBA to achieve lasting stabilization. In those that do not, the fiscal and structural troubles at the root of the very need for stabilization eventually resurface, threatening stability. While the pattern is common to exchange rate-based stabilizations, it applies a fortiori to currency boards as the cost of failure and exit is by construction higher. Alas, the experience suggests that most countries adopting CBAs choose to relearn this painful lesson: both Argentina and the Baltics allowed the emergence of fragilities. While the eventual shock leading to crisis came from abroad, domestic sins of commission and omission magnified the impact. In the event, the Baltics doubled down, accepting the strictures coming with a CBA and eventually exiting smoothly into the Eurozone. Argentina, lacking a comparable external long-term exit strategy, exited the CBA.

How much do differences in institutional arrangements matter relative to commitment? The answer is unclear. The two hardest (orthodox) modern stabilization-focused CBAs – Estonia and Bosnia and Herzegovina – achieved lasting stabilization. Yet so have two of the softer boards, Latvia (pursuing an

ambitious internal devaluation strategy even in the face of external skepticism) and Bulgaria, while a third member of the soft CBA group, Argentina, exited. While the small sample suggests caution in generalizations, the outcomes suggests that political commitment can, at least to a degree, compensate for a less stringent institutional structure.

In our view, boards remain not just viable but attractive stabilization tools for small open economies that are facing serious turmoil, are sufficiently flexible to adjust to the inevitable shocks, and enjoy sufficient domestic consensus to take the tough policy actions required by a CBA. For larger, less open economies, the trade-offs between reduced monetary policy flexibility and credibility gains are fundamentally less favorable, restricting the circumstances in which political consensus supports their adoption to severe monetary instability. The caveats are important. Currency boards are not costless; they impose real constraints in the fiscal and the financial realm. The observed rarity of their adoption – six cases in a quarter century – speaks to these constraints. It is precisely the scale of challenges preceding their adoption that can engender the degree of political support necessary to introduce boards.

Overall, the experience of countries operating currency boards suggests that they are neither a low-costs panacea for a wide range of monetary ailments nor overly restrictive arrangements doomed to result in crisis. Rather, as all exchange rate regimes, they promise specific benefits and come with specific costs. The decision whether to adopt a CBA thus entails a careful consideration of the trade-offs in the particular circumstances. If adopted, their success depends on a determined effort during the honeymoon period following stabilization to resolve the underlying fiscal challenges prompting their adoption, as well as vigilance to emerging fragilities, notably overvaluation and excessive capital inflows.

Cross-References

- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ The Historical Evolution of Monetary Policy in Latin America
- ▶ The Sterling Area 1945–1972

References

Abbott A (1959) A brief history of the crown agents and their office. Eyre and Spottiswoode, London

Alonso-Gamo P, Fabrizio S, Kramarenko V, Wang Q (2002) Lithuania: history and future of the currency board arrangement. IMF working paper 127. International Monetary Fund, Washington, DC

Aslund A, Dombrovskis V (2011) How Latvia came through the financial crisis. Peterson Institute for International Economics, Washington, DC

Bakker B, Gulde A-M (2010) The credit boom in the EU new member states: bad luck or bad policies? IMF working paper 130. International Monetary Fund, Washington, DC
Bank of International Settlements (1953) The Sterling area. Bank of International Settlements, Basle
Barro R, Gordon D (1993) Rules, discretion and reputation in a model of monetary policy. J Monet Econ 12(1):101–121
Baster A (1929) The imperial banks. P.S. King and Son, London
Bennett A (1992) The operation of the Estonian Currency Board. IMF Staff Pap 40(2):451–470
Berengaut J, Lopez-Claros A, Le Gall F, Jones D, Stern R, Westin A-M, Psalida E, Garibaldi P (1998) The Baltic countries – from economic stabilization to EU accession. IMF occasional paper no. 173. International Monetary Fund, Washington, DC
Blanchard O, Griffith M, Gruss B (2013) Boom, bust, recovery: forensics of the Latvia crisis. Brook Pap Econ Act 2:325–388
Calvo G, Talvi E (2005) Sudden stops, financial factors and economic collapse in Latin America: learning from Argentina and Chile. NBER working paper no. 11153. National Bureau of Economic Research, Cambridge, MA
Calvo G, Izquierdo A, Talvi E (2003) Sudden stops, the real exchange rate, and fiscal sustainability: Argentina's lessons. NBER working paper no. 9828. National Bureau of Economic Research, Cambridge
Central Bank of Bosnia and Hercegovina (2017) 20 years of stability. CBBH, Sarajevo
Chalmers R (1893) A history of currency in the British colonies. Eyre and Spottiswoode for HMSO, London
Clauson G (1944) The British colonial currency system. Econ J 54(213):1–25
Coats W (2007) One currency for Bosnia: creating the Central Bank of Bosnia and Herzegovina. Jameson Books, Ottawa
Crick W (1965) Commonwealth banking systems. Clarendon Press, Oxford
Daseking T, Ghosh A, Lane T, Thomas A (2005) Lessons from the crisis in Argentina. IMF occasional paper, 236. International Monetary Fund, Washington, DC
De Haan J, Berger H, Van Fraassen E (2001) How to reduce inflation: an independent Central Bank or a Currency Board? The experience of the Baltic countries. Emerg Mark Rev 2(3):218–243
De la Torre A, Yeyati EL, Schmukler S (2003) Living and dying with hard pegs: the rise and fall of Argentina's Currency Board. Economia 3:43–107
Enoch C, Gulde A-M (1997) Making a currency board operational. IMF papers on policy analysis and assessment no. 10. International Monetary Fund, Washington, DC
Fleming M (1962) Domestic financial policies under fixed and floating exchange rates. IMF Staff Pap 9:369–379
Ghosh A, Gulde A-M, Wolf H (2002) Exchange rate regimes: choices and consequences. The MIT Press, Cambridge, MA
Greaves IC (1950) Monetary systems of the colonies. HMSO, London
Greaves IC (1953) Colonial monetary conditions. HMSO, Colonial research studies, no. 10. HMSO, London
Gross T, Heft J, Rodgers D (2012) On currency boards – an updated bibliography of scholarly writings. Studies in applied economics, no.1. John Hopkins University, Baltimore
Gulde A-M (1999) The role of the currency board in Bulgaria's stabilization. Financ Dev 36(3):36–39
Gulde A-M, Kähkönen J, Keller P (2000) Pros and cons of currency board arrangements in the lead-up to EU accession and participation in the Euro Zone. IMF policy discussion paper 1. International Monetary Fund, Washington, DC
Hanke S (2008) Why Argentina did not have a currency board. Cent Bank J 18(3):56–58
Hanke S, Schuler K (1991) Keynes' Russian currency board. In: Hanke S, Walters A (eds) Capital markets and development. ICS Press, San Francisco, pp 43–63
Hanke S, Schuler K (1994) Currency boards for developing countries: a handbook. ICS Press, San Francisco
Hanke S, Sekerke M (2003) The forgotten currency board: St. Helena. Cent Bank J XIII(3):77–81

Hansson A, Randveer M (2013) Economic adjustment in the Baltic countries. Working paper 1. Eesti Panki, Talinn
Hazlewood A (1952) Sterling balances and the colonial currency system. Econ J 62:942–945
Hazlewood A (1953) Colonial external finance since the war. Rev Econ Stud 21(1):31–52
Ho C (2002) A survey of the institutional and operational aspects of modern-day currency boards. Bank for International Settlements working paper international, no. 110-1-21. Bank for International Settlements, Basle
Honohan P (1997) Currency Board or Central Bank? Lessons from the Irish Pound's link with Sterling, 1928–79. BNL Q Rev 200:39–67
Hopkins A (1970) The creation of a colonial monetary system: the origins of the West African Currency Board. Afr Hist Stud 3(1):101–132
Jao YC (1974) Banking and currency in Hong-Kong: a study of postwar financial development. Macmillan, London
Kamhi N, Dehejia VH (2006) An assessment of the currency board regime in Bosnia and Hercegovina. Emerg Market Financ Trade 42(6):46–58
Knobl A, Sutt A, Zavoico B (2002) The Estonia Currency Board: its introduction and role in the early success of Estonia's transformation to a market economy. IMF working paper 96. International Monetary Fund, Washington, DC
Korhonen I (1999) Currency boards in the Baltic countries: what have we learned? BOFIT discussion paper 6. Bank of Finland, Institute for Economies in Transition, Helsinki
Kovacevic D (2003) The currency board and monetary stability in Bosnia-Hercegovina, in Bank for International Settlements, regional currency areas and the use of foreign currencies. BIS papers no 17. Bank for International Settlements, Basle
Kwan Y, Lui F (1996) Hong Kong's Currency Board and changing monetary regimes. NBER working paper no. 5723. National Bureau of Economic Research, Cambridge
Lättemäe R, Randveer M (2004) Monetary policy and EMU enlargement: issues regarding ERM II and adoption of the Euro in Estonia. Atl Econ J 32(4):293–301
Latter T (2004) Hong Kong's exchange rat regimes in the twentieth century: the story of three regime changes. Working paper 17. Hong Kong Institute for Monetary Research, Hong Kong
Lau G (2000) Discussion on currency boards and catching up: the Irish experience. Enlargement papers 1. EU Commission, Brussels. pp 82–87
Lettiche JM (1974) Dependent monetary systems and economic development: the case of Sterling East Africa. In: Sellekaerts W (ed) Economic development and planning: essays in honor of Jan Tinbergen. Palgrave Macmillan, London, pp 186–236
Loynes JB (1962) The West African Currency Board 1912–1962. West African Currency Board, London
Moynihan M (1975) Currency and Central Banking in Ireland 1922–1960. Gill and Macmillan, Dublin
Mundell R (1962) The appropriate use of monetary and fiscal policy for internal and external stability. IMF Staff Pap 9(1):70–79
Mundell R (1963) Capital mobility and stabilization policy under fixed and flexible exchange rates. Can J Econ Polit Sci 29:475–485
Narsey W (2016) British imperialism and the making of colonial currency boards. Palgrave studies in the history of finance. Palgrave Macmillan, New York
Nenovsky N, Hristov K (2002) The new currency boards and discretion: empirical evidence from Bulgaria. Econ Syst 26(1):55–72
Newlyn W, Rowan D (1954) Money and banking in British Colonial Africa. Clarendon Press, Oxford
Oppers S (2000) Dual currency boards: a proposal for currency stability. IMF working paper, WP/00/199. International Monetary Fund, Washington, DC
Paolera GD, Taylor AM (2001) Straining at the anchor. NBER and University of Chicago Press, Chicago

Perry G, Servén L (2003) The anatomy of a multiple crisis. Policy research working paper no 3081. World Bank, Washington, DC
Poirot C (2003) Did the currency board resolve Bulgaria's financial crisis of 1996–97? J Post Keynesian Econ 26(1):27–55
Purfield C, Rosenberg C (2010) Adjustment under a currency peg Estonia, Latvia and Lithuania during the global financial crisis 2008–09. IMF working paper WP/10/213. International Monetary Fund, Washington, DC
Repse E (1999) Mr. Repse gives a short speech on monetary policy in Latvia and exchange rate strategies. BIS Rev 125/199
Sayers R (1952) Banking in the British commonwealth. Oxford University Press, London
Schenk C (1994) Britain and the Sterling area: from devaluation to convertibility in the 1950s. Routledge, London
Schenk C (1997) Monetary institutions in newly independent countries: the experience of Malaya, Ghana and Nigeria in the 1950s. Finan Hist Rev 4:181–198
Schoors K (2003) The effect of Soviet monetary disintegration on the collapse of trade between members of the commonwealth of independent states. Econ Syst 27:1–26
Schuler K (1992) Currency boards. PhD Dissertation, George Mason University
Schwartz A (1993) Currency boards: their past, present and possible future role. Carn-Roch Conf Ser Public Policy 39:147–187
Sepp U, Randveer M (2002) Aspects of the sustainability of the Estonian Currency Board arrangement. Working papers of Eesti Pank, 5. Eesti Pank, Talinn
Special Correspondent (Sidney Caine) (1948/49/50) Monetary systems in the colonies. The Banker, various issues, 87–89
Staehr K (2015) Exchange rate policies in the Baltic States: from extreme inflation to Euro membership. CESifo Forum 16(4):9–18
Sunderland D (2004) Managing the British Empire. The Royal Historical Society, The Boydell Press, Woodbridge
Thom R, Walsh B (2002) The effect of a currency union on trade: lessons from the Irish experience. Eur Econ Rev 46(6):1111–1123
Treadgold M (2003) The Philippine Currency Board arrangement, 1945–48: a case of deflationary bias? Finan Hist Rev 10:57–74
Treadgold M (2005) Colonial currency boards: the seigniorage issue. Hist Econ Rev 41:126–141
Treadgold M (2006) Factors inhibiting deflationary bias in currency board economies: evidence from the colonial era. Aust Econ Hist Rev 46:130
Tsang S-k (1999) Legal frameworks of currency board regimes. Q Bull, Hong Kong Monetary Authority 8:50–63
Tsang S-k (2000) The evolution and prospects of the Hong Kong CBA. Mimeo. Eesti Pank, Talinn
Twarowska K (2014) The role of the currency board in the Latvian economy. Ann Univ Mariae Curie-Sklodowska, Section H XLVIII(2):111–122
Uche CU (1997) Bank of England vs the IBRD. Explor Econ Hist 34:220–241
Vice D (1983) The coinage of British West Africa and St. Helena 1684–1958. Format Publications, Birmingham
Wolf H (2016) Currency boards as a path toward the Eurozone: lessons from the Baltics. Int Econ Econ Policy 13(1):45–57
Wolf H, Ghosh A, Berger H, Gulde A-M (2008) Currency boards in retrospect and prospect. The MIT Press, Cambridge, MA

Part VIII

Monetary Integration

The Evolution of the Modern US Monetary and Payments System

28

David F. Weiman and John A. James

Contents

Abstract

In this chapter, we trace out the long, uneven evolution of the modern US monetary and payments systems, from the Early Republic Era (beginning in the 1790s) to the formative years of the Federal Reserve System (in the 1920s). Our narrative is divided into four periods (and sections), demarcated by essential elements constituting a truly common national monetary-payments union: a common unit of account, currency, and finally bank deposit money. At the bookends, the monetary-payments system was coordinated administratively by a central federal authority which realized to a lesser and greater extent the criteria of a "more perfect" monetary-payments union. In the interim, the balance of power over the monetary-banking system would shift between the states and the

John A. James: Deceased.

D. F. Weiman (✉)
Barnard College-Columbia University, New York, NY, USA
e-mail: dweiman@barnard.edu

J. A. James
University of Virginia, Charlottesville, VA, USA

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_21

federal government over two distinct but inter-related regimes of competing, "multiple" circulating currencies and a common national currency. Yet, despite their differences, these two distinct phases were connected by a common thread. Instead of formal hierarchies, the banking-monetary system over this long nineteenth century rested on an ever-expanding complex of networks that connected banks within and between cities and coordinated the flows of good funds that literally "greased the wheels of commerce." Often initiated by government policy, the growth and articulation of these networks, we show, ultimately depended on the basic logic of network economics, which in fact policy makers recognized in the design of the National Banking and the Federal Reserve Systems.

Keywords
Banks · Monetary union · Correspondent banking system · National Banking System · Federal Reserve

Introduction

In this chapter, we trace out the long, uneven evolution of the modern US monetary and payments systems, from the Early Republic Era (from the 1790s to the eve of the Civil War) to the formative years of the Federal Reserve System (from the eve of WWI to the early 1920s). Our narrative is divided into four periods (and sections), demarcated conceptually by essential elements constituting a truly common national monetary-payments union: a common (dollar) unit of account, a common (national later Federal Reserve banknote) currency, and finally an integrated banking system with a common national bank (deposit) money. At the bookends, the monetary-payments system was coordinated administratively by a central federal authority – the "National Bank" or Bank of the United States (from 1791 to 1811 and 1816 to 1836) and the Fed (from 1914 on) – which realized to a lesser and greater extent the criteria of a "more perfect" monetary-payments union.

In the interim the balance of power over the monetary-banking system would shift between the states and the federal government over two distinct but inter-related moments. In the first (from the early 1830s to the Civil War) "states' rights" over the banking-monetary system would prevail and usher in a regime of competing, "multiple" circulating currencies (Hammond 1957, p. 573; Dowd 1994; Selgin and White 1994). The National Banking Acts (1863–1865), instituting a common national currency, restored federal authority but only temporarily (Curry 1968; Bensel 1990). The rapid diffusion of deposit banking in 1880s reignited the jurisdictional competition over the banking-monetary system, but in this case over the formation and regulation of deposit not note-issuing banks.

Yet, despite their differences, these two distinct phases were connected by a common thread. Instead of formal hierarchies, the banking-monetary system over this long nineteenth century rested on an ever-expanding complex of networks that connected banks within and between cities and coordinated the flows of good funds that literally "greased the wheels of commerce." Often initiated by government

policy, the growth and articulation of these networks, we show, ultimately depended on the basic logic of network economics, which in fact policy makers recognized in the structural design of the National Banking and the Federal Reserve Systems.

In telling this tale we highlight two themes. Appropriately modifying Lerner's thesis, we show that money is "a creature of the state" but not entirely (Lerner 1947; Cohen 1998; Helleiner 2003). Illustrating the first part our initial claim, our history shows how government interventions, often in response to war-time exigencies, would constitute and reconstitute the essential properties of an American money – as a unit of account and means of payment for final settlement in domestic transactions. They also, importantly, shaped the evolution of the banking and hence payments system, the vital infrastructure for the circulation of this money between buyers and sellers, debtors and creditors, etc. As a corollary to this claim, our account clearly shows that when it comes to the monetary-banking system, "history matters."

At the same time, we observe, the vast majority of transactions in the USA were (and still are) denominated and settled in the convertible liabilities of payments intermediaries, most importantly commercial banks (Garber and Weisbrod 1990; McAndrews and Roberds 1995; Kahn and Roberds 2002, 2009). These private moneys are generally acceptable, because the corresponding banks are integrated into a hierarchical network that can efficiently and reliably transfer the good funds to finalize the payment. Within this network, the de facto good funds are the liabilities of the center bank(s), whether official money or not – that is, the liabilities of a central monetary authority or (e.g.) an association of private banks. As the US case clearly illustrates, a central monetary authority was not a necessary condition for an efficient interbank payments network, though without government sanction center banks could not insure that all bank deposits are the same in economic value nor the efficient delivery of good funds between banks in bad times as well as in good.

The Antebellum Payments System

Bearer notes, rather than deposits, were the principal form of bank liability used as means of payment in antebellum America. In 1819, for example, notes in circulation issued by state-chartered banks were 74% greater than their deposits (Van Fenstermaker 1965, p. 76; Carter et al. 2006, Table Cj7–21; by 1859 bank deposits would exceed banknotes by 13%, though the public still evinced a greater preference for holding currency including specie coin over deposits). Notes circulated at par locally, as long as the issuing bank was solvent and had not "suspended payments" or restricted their convertibility into specie. After all, if the market value of notes issued by a local bank ever fell below par, the bearer could simply present them at its counter for redemption in specie (Gorton 1996, p. 353, 1999).

Banks in larger urban centers were the exception, as they relied more heavily on deposits rather than on note issues. According to the earliest available systematic data, the value of deposits exceeded that of notes in circulation on the balance sheets of banks in Boston (beginning in 1803), Philadelphia (in 1814), Baltimore (in 1824),

and New York (in 1831) (Van Fenstermaker 1965, p. 41). Checks drawn on individual deposits were largely confined to local transactions, however, because they required timely clearing and settlement (Klein 1911, p. 607; Kahn and Roberds 1999). Unlike notes, they were dual claims on a bank and the payer's account. The recipient of a check was, therefore, subject to two risks – that the payer's account did not have sufficient funds to cover the check payment and that the bank on which it was drawn could not redeem the check in specie at par.

To minimize these potential costs (to their customers and themselves), urban banks developed elaborate procedures to expedite check redemptions. In the first half of the nineteenth century, banks cleared and settled check payments bilaterally by dispatching messengers to deliver the check claims to each counterparty and then return to pay the balance (Myers 1931, pp. 94–95). Following the model of London, New York banks organized the Clearing House Association in 1853. It established a central location (originally at 14 Wall Street) where exchanges among member banks would occur. On the morning of each business day, member banks would deliver to the Clearing House checks for collection. After totaling the items due from and due to each other, they would settle the difference in gold (and later legal-tender notes) often via debits and credits to their deposit accounts at the Clearing House. Such multilateral clearing and settlement was clearly more efficient than the bilateral system which it replaced. Clearing houses were organized in Boston and Philadelphia later in the 1850s and subsequently spread rapidly to other cities. By 1907 there were clearing houses in 106 cities (Cannon 1910).

Unlike deposits or specie, banknotes were used in making nonlocal payments, as they were cheaper and more convenient to ship than hard money. Indeed, specie was rarely used in nonlocal, large value transactions (Colwell 1860, pp. 135, 190, 262, 447). There is no direct information on the extent to which notes circulated across states or regions (e.g., Atherton 1971, p. 139; Knodell 1988; Gorton 1996). Gorton (1996, p. 354), relying on qualitative evidence, concludes that these long-distance note flows were significant and followed the natural course of interregional trade to larger commercial centers. There, periodicals routinely published information on the accumulated stocks of foreign notes (e.g., their value if "current," the description of counterfeits, etc.) in "daily exchange" reports. Moreover, in the largest centers with the densest note markets, specialized periodicals such as *Van Court's Bank Note Reporter* (Philadelphia) and *Thompson's Bank Note Reporter* (New York City) comprehensively surveyed the terrain. Between 1839 and 1858, for example, *Van Court's* quoted discounts on notes from individual banks in virtually every state and territory of the country (save Iowa, Minnesota, Missouri, and Texas).

Additionally, specialized intermediaries in commercial centers – note brokers or "shavers" – would aggregate and remit these notes to the issuing bank for redemption or buy and sell them in local markets (see Hammond 1957, pp. 702–703). In the latter case, "foreign" notes typically commanded a discount relative to their par or face value. Discounts varied systematically with the location and reputation of the issuing bank (as well as the state government regulatory regime), which determined the costs and risks of returning the notes for redemption in specie or its value equivalent (Gorton 1996, 1999; Jaremski 2011).

To enhance the value and hence acceptability of their notes in commercial centers, "foreign" banks negotiated convenient, reliable redemption arrangements with a correspondent, typically a chartered or private bank. In exchange for a reserve deposit, the correspondent agreed to purchase the bank's notes for specie or its own notes, at or near par value (see Bodenhorn 2000, pp. 192–193, 197). The Suffolk system, lasting from the mid-1820s to the late 1850s, was the first, if not most (in) famous, example of such an arrangement (Bodenhorn 2002; Rolnick et al. 1998; Calomiris and Kahn 1996). It maintained par banknote redemption and clearing operations which covered virtually all of New England. New York state's banking reform law of 1840 (later amended in 1851) institutionalized these arrangements by requiring all note-issuing "free" banks to main a reserve account in the state capital of Albany or New York City (Myers 1931, p. 105; Weiman and James 2018). By the 1840s western banks in Ohio and Indiana similarly had established redemption accounts with Philadelphia and New York correspondents (Knodell 1988, p. 297; Redenius 2004).

As an alternative to shipping currency in whatever form, agents could also transact in bills of exchange. Though the primary instrument in foreign trade since the early colonial era, dealings in domestic or inland bills in the early decades of the republic were sparse, the province of assorted note brokers and a few state banks (on the legal foundations of these negotiable instruments, see Freyer 1976). Under the leadership of Nicholas Biddle, however, the Second Bank of the United States (BUS) came to dominate the system of interregional payments in the 1820s and early 1830s and essentially created the domestic bill of exchange market (Catterall 1903, pp. 138–143; Bodenhorn 1992, 2000, pp. 168–177; Knodell 2003). Branches in interior cities were instructed to replace, whenever possible, local promissory notes with domestic bills of exchange. This policy would tie "note issue at those branches to the means of redeeming them at eastern branches and [preserve] the liquidity of the Bank as a whole" (Knodell 1998, p. 715).

As a result, BUS purchases of domestic exchange rose from less than $6 million in 1820 to almost $70 million by 1833, more than enough to finance the total volume of trade passing through New Orleans from the Midwest (Bodenhorn 1992, p. 595). Biddle exploited the economies of centralization to lower the exchange costs of transferring funds from one location to another. Through mutual clearing of its branch accounts on the books at its Philadelphia headquarters, the BUS economized on currency shipments in settling interregional payments imbalances (Knodell 1998, p. 716, 2003, p. 13). Moreover, to promote interregional trade, the Bank acted as a market maker in domestic exchange, reducing and stabilizing exchange costs. This centralization of the interregional payments system, however, was short lived.

Andrew Jackson's infamous bank war ended the BUS's unique privilege of interstate branching in 1836 and ushered in a period of financial disintermediation in interregional payments. Note brokers and private bankers, which until then had been overshadowed by BUS operations, became active participants in the now more decentralized system of domestic exchanges (Knodell 1998, pp. 717–719; Bodenhorn 2000, pp. 177–185; Knodell 2010). Individuals making payments in New York, for example, could buy bills payable there through a broker. Remittance

and collection of domestic bills were again handled through interstate networks of private bankers and exchange brokers. Knodell (1998, p. 717) finds the growth in domestic exchange facilities in Cincinnati and Cleveland during this period was largely concentrated in this "unregulated" sector. Sylla (1976) estimates that by 1860 private banks accounted for almost one-third of all banks and one-quarter total bank capital, and domestic exchange operations was one of their principal activities (see also Knodell 2010).

Mediating out-of-town payments was one of the major functions of antebellum state-chartered banks as well (Hammond 1957, p. 700). Because of the legal constraint on interstate (and even within-state) branching, banks forged an alternative payments network through their correspondent relationships. As interior banks discounted their customers' bills of exchange drawn on their (e.g.) New York correspondent, they accumulated balances in the city. Instead of repatriating these funds through costly specie shipments, they adopted a more profitable option, selling these funds or "New York exchange" to other local customers who needed to make payments there. The growth of the domestic bill of exchange market therefore complemented the development of a local market for drafts on money centers, providing a ready supply of funds for local bank customers who needed to make long-distance payments (Catterall 1903, pp. 141–142; Hammond 1957, p. 318).

Through their role as payments intermediaries issuing banknotes or discounting bills, interior banks naturally accumulated interbank balances in money centers – either for purposes of note redemption or as a source of domestic exchange. These correspondent bank relationships between smaller country banks holding assets called "due from other banks" and larger city banks holding liabilities labeled "due to other banks" proliferated after the 1820s (Bodenhorn 2000, pp. 192–198; Redlich 1968, p. 51). Warren Weber (2003) describes the pattern of such antebellum interbank relationships using a data source which disaggregated the amounts due from other banks for Pennsylvania creditor banks over the 1850s. He finds that country banks dealt almost entirely with banks in financial centers (Philadelphia, New York, Baltimore, or Pittsburgh), the choices of which were determined by trade patterns. In addition, their relationships with a single bank in a particular city were highly stable over time.

As the preeminent commercial or jobbing center, New York banks soon dominated the correspondent banking market (Burrows and Wallace 1999, p. 436). As early as 1835 net bankers' balances (defined as due to less due from other banks) held in New York amounted to $4.40 million as compared with $2.93 million in Philadelphia and $1.03 million in Boston; by 1850 the New York total had risen to $12.51 million versus $2.45 million in Philadelphia and $4.17 million in Boston (Bodenhorn 2000, p. 196). By 1850 almost 600 out of 700 incorporated US banks maintained New York accounts, with brokers and out-of-town individuals holding about an equal amount (Myers 1931, pp. 103–125). A. B. Johnson, president of the Ontario Branch Bank in Utica, New York, observed in 1857: "The selling of drafts on New York becomes . . . one of the regular sources of profit to country banks, as well as of convenience to men of business; and every country bank keeps funds there. . ." (Johnson 1857, p. 26; quoted in Weber 2003, p. 471).

The Civil War Era: A Second (Monetary) Revolution

In his reports on the National Bank and the mint, Hamilton envisioned a truly American monetary union, one built on dual federal-state foundations (Hamilton et al. 1821). Exercising its explicit constitutional authority, the federal government would institute a national monetary standard – the bimetallic dollar unit of account – and "coin" legal tender currency pegged 1-for-1 to the dollar standard. State governments, at the same time, were delegated the authority to charter private banks that could also issue circulating currency in the form of bearer notes. In Hamilton's scheme, the "National Bank" or BUS would anchor this dual monetary system by enforcing the par convertibility requirement on state banknotes. Through its dominant position in regional markets for inland bills, the Bank also pegged their value relative to the dollar through a scheme of administered exchange rates (Catterall 1903, pp. 138–139; Knodell 2017, pp. 75–77). Together, these policies knit the territorial United States into a "tighter" monetary union and effectively transformed BUS notes and deposits into a common national money, on par with the Treasury's legal tender currency (Fraas 1974; Knodell 2017, esp. Chap. 4).

Jackson's successful "bank war" clearly demonstrates the pivotal position of the BUS in – or perhaps more accurately the tenuous political economic foundations of – this fledgling American monetary union. As Sylla (2006) observes, without a central authority of some kind to integrate banks into a tighter payments network, the monetary union could unravel at least within limits determined literally by the cost and risks of shipping official specie money. The recent literature on the ensuing "free" banking period vividly illustrates the resulting monetary fragmentation. Knodell (1988) emphasizes the dramatic shift from a regime of stable, administered to more volatile, floating rates of exchange between private bank and official Treasury money, which increased the cost and uncertainty of interregional payments and trade (see also Weiman and James 2007).

In this context, Shambaugh (2006) observes, state governments could wield their chartering-regulatory authority to conduct independent monetary policy, evident in their divergent money supply growth, price inflation, and interest rates. This credit-monetary expansion was greater in peripheral areas, where banks could exploit their relative spatial economic isolation, and their financing of local economic development could potentially yield large rents for their shareholders and chartering governments. The more decentralized banking-monetary system even weakened the common dollar standard (Redenius 2007). As a contemporary accounting text explained, intermediaries such as banks confronted the very real and "most difficult" problem of selecting the appropriate unit for reckoning their transactions, precisely because they routinely traded in a variety of bank moneys of varying value (Bryant et al. 1863, pp. 256–57).

Hamilton's vision of a quasi-public "National Bank" represented only one institutional solution to the formation of a "more perfect" monetary-payments system. As an alternative model of monetary integration, the adoption of a truly common currency would institutionalize the economic link between the monetary standard

and money supply. At the very least a common currency would represent an enduring (or "irreversible") commitment to monetary integration and so foster stronger, self-reinforcing economic linkages between member regions (Mundell 2002, p. 127; Rose 2000; Rose and van Wincoop 2001). In particular, it would mitigate the exchange rate-unit of account uncertainties that increased the cost and risk of interregional transactions in late antebellum America, such as the "assorting" of heterogeneous currencies in settling interbank claims (Weiman and James 2007).

The adoption of a common currency, however, confronted an insurmountable political economic hurdle during the decades before the Civil War. The territorial USA at the time was not an obviously "optimal" currency region, and consequently legislating monetary integration would likely face credible veto threats – notably from Southern Democrats who favored a more decentralized (or states' rights) banking-monetary system (Hammond 1970, pp. 138–141 and Rockoff 2003; see also Knodell 2006 and Jaremski and Rousseau 2013, pp. 15–18). Exploiting the opportunity afforded by Southern secession, a Republican congress in 1863 barely passed the National Currency Act, first in a series of laws now known as the National Banking Acts. The NCA was modeled on earlier free banking laws notably that of New York (for a fuller analysis of this episode, see Bodenhorn 2003, Chap. 8 and Weiman and James 2018). It established a new form of bank organization, national banks with a federal charter. National banks, in turn, could issue banknotes, secured by US government bonds.

The NCA was only a first step towards making national banknotes a truly common currency, one that circulated at par throughout the country. After all, they were still the liabilities of individual banks, though secured by US government bond collateral and deemed a legal tender for private and most public transactions. Moreover, the law only mandated their par redemption at the office of the issuing bank. Consequently, their value could vary spatially depending on "pure exchange" arising from regional payment imbalances, seasonal or chronic, within bounds set by the transaction costs and risks of shipping notes and official money (for alternative views on this matter, see Redenius 2007, p. 220 vs. Bensel 1990, pp. 263–264). The Legal Tender Act of 1862 did substantially lower these costs and risks transforming zero-interest Treasury notes into a common legal tender money that could substitute for specie in settling interbank payments.

Recognizing this lapse, proponents of the NCA debated opposing note redemption schemes that called for a single national or multiple regional reserve centers. In the end they settled on an ingenious compromise incorporated into the 1864 National Banking Act. It established a more integrated interbank payments network through a tiered hierarchy of redemption agents, corresponding to federally chartered banks located in an officially designated "central reserve" city (at the time only New York) and 16 regional "reserve" centers spanning the continent from Boston and Philadelphia in the Northeast to Charleston and New Orleans in the South to Chicago, St. Louis and San Francisco in the Midwest and Pacific coast. (It also contained a provision to add two more centers in the South.) "Country" banks located outside of these official centers were required to maintain sufficient note redemption reserves with an approved agent in an economically proximate reserve

center where their notes tended to accumulate. Likewise, reserve city banks were mandated to hold redemption reserves with an approved agent in New York.

This "pyramiding" of reserves, as Redenius (2007, pp. 225–226) concludes, realized the ends of a more centralized plan with a single reserve center, but still afforded country banks on the periphery the convenience of a regional reserve agent. Reserve city banks could directly orchestrate the clearing and par settlement of note payments within their region. For the longer distance transactions of hinterland banks, they could efficiently bundle and ship in bulk notes and then settle the resulting payments imbalances, typically via draft transactions on their New York reserve accounts instead of cash shipments. The centralization of settlement in New York not only economized on agents' excess reserves, but effectively pegged the value of all national banks notes to a single monetary standard, New York reserve deposits, at a parity of 1-to-1. In other words, it transformed New York deposits into a de facto national bank money; akin to BUS monetary liabilities, they constituted a less costly and risky substitute for currency in making large-value payments over longer distances.

This "network effect," ironically, greatly enhanced the value of a New York correspondent and reserve account (on network effects in payments systems, see Economides 1993 and McAndrews 1997). The essential condition was the formation of the intermediate tier of regional reserve centers in the national banking hierarchy, which were directly connected to the New York banking sector and so to each other. Through a New York correspondent, then, any bank throughout the country could furnish its customers with direct monetary access to these pivotal commercial-financial-manufacturing centers and in turn a critical mass of their distant trading partners. Within this novel payments network, New York drafts replaced banknotes as a generally accepted national payments instrument, which buyers and sellers regardless of their location could readily acquire and clear and settle through their local bank. Like reserve agents, country banks could in turn realize significant economies by holding their official and excess clearing reserves in a single New York account.

Indirect evidence of this unintended network effect comes from two complementary sources (again for details, see Weiman and James 2018). By 1870 New York was the dominant note redemption center in the national banking system, with a market share of 60% far greater than the combined shares of Boston, Philadelphia, and Chicago. Moreover, New York banks secured a significant if not dominant presence in the country bank market in every region of the country. At the same time nearly 90% of all banks – national, state-chartered and private – secured a New York correspondent.

The alternative index compares New York banks' correspondent balances before and after the Civil War. These "due to" items on their balance sheets (adjusted for war-time inflation) jumped by more than 50% over the period, from an average of $27.2 before the war to $41.4 million afterwards (for details, see James and Weiman 2011). The increasing centralization of bankers' balances in New York after the Civil War, moreover, does not simply reflect the tiered reserve structure under the NBS, as country banks and eventually reserve banks held substantial excess or clearing

reserves in their correspondent accounts. From these results, we infer that interior banks allocated significant proportions of their correspondent balances to domestic exchange transactions mediated by New York banks.

In other words, the National Banking Acts reconstituted a more centralized, national long-distance payments-banking network akin to the Second BUS's hub-and-branch hierarchy. Within this more tightly integrated monetary-banking union, sight drafts drawn on New York banks became a generally accepted national payments instrument. As striking evidence, we observe the rapid demise of the New York domestic exchange market by 1868, as New York agents could now substitute "local" drafts drawn on their New York banks for "foreign" drafts drawn on banks in other commercial centers, which they had previously acquired through market transactions.

Unlike the BUS, New York correspondents could not administer domestic exchange markets in regional centers, where they competed for the business of local banks against each other and against banks in other "eastern" money centers. Nonetheless, the adoption of a common currency (both legal tender greenbacks and national banknotes) and the centralization of correspondent relations in New York approximated this more stable exchange rate system. Consider, for example, the purchase of $1000 of New York funds in Chicago. Though the par value of these funds was $1000 in each city, they sold in Chicago for a premium of 1.5% on the eve of the Civil War. By the end of the war, exchange rates (measured as a deviation from a par value of one) had fallen precipitously to only 0.1% where they remained throughout the decade.

More important, the variation in and hence uncertainty over rates also declined sharply over this period. Again taking the case of Chicago, the standard deviation in exchange rates fell from 51 to 5 basis points, or by 90%, between the late 1850s and late 1860s. These magnitudes, in fact, overstate agents' uncertainty, as exchange rates continued to fluctuate seasonally, albeit within lower bounds, due to the influence of the regional agricultural sector on the balance of payments. Thus, despite the residual variability in rates, merchants could more accurately predict their movements and so adjust their prices accordingly.

A number of alternative but related hypotheses have been advanced to explain this sharp decline in domestic exchange rates and their variability, though none are consistent with evidence on the timing of these changes (Garbade and Silber 1979; Phillips and Swamy 1998). Instead, we emphasize two related monetary policy innovations during the Civil War era that dramatically lowered the costs and risks of interbank settlement. The first was the diffusion of a common settlement medium, beginning with Treasury issues of transferrable notes in 1861 and followed by its larger issues of legal tender fiat greenbacks in 1862. The other critical ingredient was the hierarchical organization, and hence, greater administrative coordination of payments flows through central reserve or correspondent banks. As a result, New York exchange or funds became the standard means of payment for buyers and sellers regardless of their location, as long as their banks maintained a New York correspondent which most did. And because of the increasing centralization of correspondent relations and balances among a handful of large New York banks,

they could more effectively coordinate the vast flows of payments instruments and insure the greater liquidity of bankers' balances (James and Weiman 2010).

The Postbellum Era: A "Maturing" National Banking-Monetary System

From Banknote to Deposit Money

Over the latter half of the nineteenth century, banks became more integral and specialized intermediaries in the US payments system. Prior to the formation of the Fed, banks continued to supply both circulating currency and transferrable demand deposits (for the money supply data, see Carter et al. 2006, Table Cj42–48). The share of bank money steadily increased over the period, from approximately 65 to over 90%, but only because of the dramatic shift in the public's holdings of deposits relative to currency. Following a nearly two-decade decline, the share of deposits in the money supply more than tripled between 1850 and 1914, from 26.3% to 88.4% (for contrasting explanations of these trends, see Bordo and Jonung 1987 and Bordo et al. 1997 vs. Friedman and Schwartz 1963, p. 122). Over the same period banks' share of the circulating currency declined from nearly 50% in the late 1840s to only 20% in 1914. The issue of greenbacks or legal tender fiat currency during the Civil War of course had some influence, but private banknotes rebounded sharply after 1864. Moreover, the most dramatic decline came in the 1880s after the USA returned to the gold standard. Despite the revival of national banknote issues after 1900, their share of the total currency supply remained well below late antebellum levels.

The diminishing circulation of private banknotes is clearly tied to the expansion of silver-backed government currency issues. Under the provisions of the Bland-Allison Act in 1878, the Treasury began purchasing around $2 million of silver per month, which was minted into coins or served as backing for paper silver certificates (Friedman and Schwartz 1963, pp. 132–133; see also Hepburn 1903, p. 312 and Dewey and Shugrue 1922, pp. 22–24). The public seemed to be quite willing to hold government paper currency in place of private banknotes, as evidenced by their divergent trends between 1878 and 1892; the volume of silver certificates grew from $7 to $327 million, while that of national banknotes fell by nearly 50% from $311.7 to $167.2 million. (The "conundrum of the low issue of national banknotes can only partly explain this decline, see James 1976a and more recently Champ et al. 1992; Selgin and White 1994; Calomiris and Mason 2008).

The changing composition of banks' convertible liabilities, the ratio of deposits to notes, tracks the timing and extent of their greater specialization in supplying deposit-based payments services. The longest balance sheet series, dating from the late antebellum period, covers the entire banking system, state-chartered banks prior to 1864 and state and national banks thereafter, and measures total deposits relative to notes. Beginning in 1896, published balance sheet data discriminate between individual and bankers' deposit balances and so provide a finer picture of their

payment services to individual rather than bank customers. Though paralleling the trends in the composition of the aggregate money supply, these data clearly show a dramatic structural break in the 1880s, when the ratio increases more than five times. The nearly 50% decline in the denominator, national banknotes in circulation, certainly played a role. Still, the more important factor was the rapid growth in deposits. Even after 1900, the ratio still hovered around $23 ($13) of total (individual) deposits per banknote, as compared to the 2-to-1 ratio in the Civil War era.

The relative growth of bank deposits can be partially explained by the peculiarly American "dual" banking system, which fueled competition between federal and state regulatory authorities over bank charter requirements (James 1978, Chap. 1; White 1982, 1983, Chaps. 1 and 3). An initial foray in 1864, the revision to the National Banking Act, imposed a prohibitive 10% tax on state banknote issues and so effectively confined state-chartered banks to the deposit business (Jaremski 2013). Beginning in the 1880s increasing numbers of more rural states adopted liberal bank incorporation laws that reduced minimum capital requirements to a fraction of the federal thresholds and so spurred the formation of state-chartered banks in smaller cities and towns (Cooke 1897; James 1978, pp. 29–39). Between 1880 and 1895 the number of state banks multiplied by more than 4.8 times and the total capital investment by 3.8 times.

Despite their rapid growth, state banks were on average smaller than their national counterparts and accounted for only a slightly larger share of total deposits over the period, from one-half in 1880 to just under 60% in 1900. Focusing only on national banks, we see the same dramatic shift in the composition of their convertible liabilities beginning in the mid-1880s. In this case we can estimate the magnitude of individual and bankers' deposit balances from 1880 on and so show the nearly parallel trend of individual deposits to notes. Regardless of the measure, both indicate a two- to three-fold increase in the deposit-note ratio over the late postbellum period.

With the more disaggregated national bank balance sheet data, we can further decompose these shifts according to the geographic-political location of national banks (see Weber 2000). Between 1881–1882 and 1909–1910, the total deposit-note ratio for reserve city banks increased by only 20%. The increase, moreover, derived mainly from the growth of correspondent balances in reserve cities (see below). By contrast, the deposit-note ratio for country banks, whether measured by total or individual deposits, nearly doubled over the same period. According to these data, by1880 deposit banking had virtually penetrated large commercial centers and was spreading rapidly down the urban hierarchy. In turn, we witness a parallel trend in the diffusion of bank clearinghouses to smaller cities, especially in the 1890s. By 1915 banks in 229 cities had formed urban clearinghouses, nearly 75% after 1890 (Cannon 1910; Rand McNally 1915).

A series of surveys conducted by the Comptroller of the Currency and for the National Monetary Commission provide additional quantitative evidence on the spread of deposit money in larger value wholesale trade between 1881 and 1909 (see Kinley 1910, pp. 20–31). The initial surveys covered only national banks and provide data on the composition of their total deposit receipts. "Credit instruments," meaning drafts and checks, accounted for more than 90% of these bank transactions.

Moreover, the use of deposit money followed the hierarchy of centers under the National Banking system. Deposit receipts were highly concentrated among national banks in reserve centers. New York banks alone accounted for 40% of the total, and the other reserve banks, an almost equal share. Also, checks and drafts comprised a larger share of receipts for banks in reserve centers relative to country banks, 93% versus 85% respectively.

Subsequent surveys detail the economic as well as spatial variations in the diffusion of deposit banking. As the data for 1896 and 1910 clearly show, deposit money was *the* means of payment in larger value, wholesale trade. In both years bank credit instruments accounted for at least 95% of the deposits made by wholesale merchants. In New York, for example, drafts and checks were almost universally employed in these exchanges, as well as in the residual category which recorded deposit receipts of noncommercial businesses. Even in smaller cities and towns, we can detect the rapid diffusion of deposit money in larger value transactions.

While these data point to the displacement of banknotes by deposit money in interregional transactions after the Civil War and especially after 1880, they are silent on the relative importance of drafts versus checks in nonlocal payments. In the absence of systematic quantitative evidence, we rely instead on more qualitative sources. A careful reading of *Bankers' Magazine*, for example, reveals an increase in bankers' concerns about customers' use of individual checks in nonlocal transactions in the early 1880s and again in the late 1890s. In the initial episode, the articles focused on check transactions within metropolitan regions of centers such as Boston, New York, and Chicago. By the turn of the century, however, check transactions had clearly transcended regional boundaries; according to one estimate by the New York Clearing House Association or NYCHA (1899); Shreve (1898), its members conducted about 30% of their check collection business with banks in the Midwest (15%) and the southern and border states (14%). By the time of his final survey in 1909, Kinley (1910, p. 123) could confidently assert that checks had become the "preonderatingly (sic) method of payment" in local and long-distance wholesale trade.

The Evolution of a National (Correspondent) Banking System

The spread of deposit banking after the Civil War increased the scale and geographic scope of draft transactions and in turn the demand for local correspondents, especially in centers near or in areas of recent settlement (Conzen 1977). Exploiting their geographic proximity, correspondents in interior transport hubs and regional commercial centers mediated the increasing shipments of drafts to and from banks in their trade area (for a fuller discussion of this development, see James and Weiman 2010, Quinn and Roberds 2008, and James (2016); see also Hallock 1903, pp. 43–46; Spahr 1926, pp. 84–130; Gilbert 2000, pp. 123–128; and Chang et al. 2008). Additionally, in the larger commercial centers of more developed, diversified economic regions (such as Boston and Chicago), correspondents settled local draft payments and so attracted increasing reserve balances from hinterland banks.

Despite the gradual erosion in their market share, New York correspondents remained the linchpin of this more decentralized, geographically expanding interbank network. Following Smith's adage about the division of labor and the extent of the market, New York correspondents increasingly specialized in the settlement of interbank draft payments, especially between banks in regional commercial centers such as reserve cities. Even when reserve banks (typically in larger centers) directly cleared their reciprocal draft payments, they tended to settle their accounts by drawing drafts on their New York correspondent. In other words, New York correspondents integrated regional correspondents into a national settlement system akin to the Gold Settlement Fund of the early Federal Reserve System, and their balances constituted a de facto reserve base for bank deposits and deposit transactions (Federal Reserve System Board of Governors 1918; Goldenweiser 1925; Spahr 1926, Chap. 8).

Evidence on the evolving structure and functioning of the postbellum correspondent banking system is both limited and indirect. Balance sheet data published by federal and state bank regulators gauge the changing concentration of bankers' balances among the largest correspondent banks and banking centers (Watkins 1929; James 1978). The *Rand McNally Bankers' Directory* lists the names and locations of banks' correspondents, which can be mapped to show the shifting trade areas of existing correspondent centers and the emergence of new ones (Conzen 1977; Odell and Weiman 1998; Redenius 2002; Jaremski and Wheelock 2017). More recently, Calomiris and Carlson (2017) have found in the Office of the Comptroller of the Currency call reports for select states detailed data on the location and amounts of national banks' correspondent balances.

An alternative source fills this empirical gap but only for national banks over years 1890–1892. It documents the actual flows of draft payments in settling accounts between country and reserve city banks and between reserve city banks, respectively (U.S. Comptroller of the Currency 1890, pp. 14–22, 1891, pp. 16–24, 220–233, 1892, pp. 24–31). They illustrate, in other words, the inner workings of the correspondent banking system in transferring funds across space and thereby in forging a national payments system. As this evidence clearly shows, New York was the only truly national "money" center (on the use of these transactional data to identify "center" banks, see Bech et al. 2010). Virtually all national banks – over 90% of those filing reports – drew drafts on a New York correspondent during the year, and the New York market share of 61.3% exceeded the cumulative shares of the next four largest correspondent centers (equal to only 27.6%). Moreover, New York correspondents accounted for the vast majority (74.3%) of settlement transactions between reserve city banks and thereby knitted together regional reserve centers (and their market areas) into a truly naftional interbank payments system (for parallel results based on other Comptroller of the Currency data, see Calomiris and Carlson 2017).

The tiered structure of interstate cash flows to and from banks in select reserve cities from 1905 to 1908 provides further evidence in support this generalization and shows the role of regional reserve centers in this hierarchical network (for the source of these data, see Kemmerer 1910, pp. 276–357). The substantial cash flows to and

from regional financial centers such as Boston, Philadelphia, Cincinnati, and San Francisco were almost always just with their hinterland and with New York banks. Other interregional flows, such as between Boston and the southern states or between Cincinnati and western states, were negligible at most. These interregional payments therefore must have been overwhelmingly channeled through New York banks rather than through correspondent accounts in other regional financial centers. The latter did mediate the financial flows of smaller cities in their hinterlands (Providence and Boston, Wilmington and Philadelphia, Little Rock and St. Louis), which had limited direct ties to New York.

Through the centralization of bankers' balances and settlement transactions in New York, correspondents there increasingly assumed essential central bank functions (see, for example, Goodhart 1988; Norman et al. 2011; Quinn and Roberds 2007; Kahn et al. 2016). Because of their general accessibility via the correspondent system, New York funds effectively became an alternative national means of payment, as good as gold in settling long-distance transactions at least during normal times. As the evidence above shows, banks could meet their mutual obligations by remitting a New York draft, as their counterparties could readily clear and settle the items via their own New York correspondent connections. In turn, ledger entries on the books of the NYCHA substituted for more costly, risky cash shipments in mediating the flows of funds between banks and ultimately their customers.

In modern parlance, New York correspondents constituted the central node in the domestic large-value payments network – "the clearing house of the country" in contemporary terms (see Sprague 1977 [1910], p. 126; and also James and Weiman 2005, pp. 123–28, Green and Todd 2001, Bank for International Settlements, Committee on Payments and Settlement Systems 2003). As a result, they realized additional benefits, reinforcing their nodal position. After the Great Depression of 1893, for example, banks sharply increased their excess reserve balances in New York relative to their holdings of vault cash (Gendreau 1983; James and Weiman 2010). They were responding in part to the diffusion of individual checks in long-distance payments which increased the uncertainty of withdrawals from their correspondent accounts. Additionally, they were exploiting economies of scope in reserve management, as local clearinghouses increasingly allowed members to settle their clearing balances by drawing a draft on their New York correspondent, rather than by transferring vault cash.

Compared to regional money center banks, New York correspondents also reaped greater liquidity externalities because of the vaster scale and scope of their market area (Gilbert 1983; McAndrews and Rajan 2000; Kashyap et al. 2002). By pooling the reserve transactions of banks from across the country, they maximally benefitted from the netting of customers' offsetting transitory and especially seasonal cash-out and in-flows. Reinforcing these individual level economies, New York banks constituted a liquid funds market at least during normal times through their transactions in overnight (call) broker loans (Garber and Weisbrod 1990; Goodfriend 1990). With the published data on reserve national banks' balance sheets, we can only gauge these liquidity economies at the aggregate or market level – in terms of the seasonal variation in banks' net "due to" positions. Despite the regular seasonal swings in net

cash flows into and out of New York, bankers' balances there varied significantly less over the course of the year than did balances in the other central reserve cities of Chicago and St. Louis and in regional reserve centers, by margins of 20% and 68%, respectively (on the seasonality of these flows, see Kemmerer 1910; Miron 1986; Champ et al. 1996).

With less seasonally variable bankers' balances, New York correspondents could economize on their holding of excess reserves and in turn expand their lending capacity. The evidence suggests that they used these resources to underwrite their greater "interbank lending" after the 1893 panic (Lockhart 1921). According to qualitative sources, New York correspondents provided regional bank customers with the equivalent of overdraft facilities by routinely extending the collection period on their outstanding checks until they could accumulate sufficient offsetting deposits. More distant customers were furnished with seasonal credit often disguised as rediscounts and bills payable on their balance sheets. Banks secured this credit line by maintaining a compensating balance in their correspondent account (usually set at 20% of their credit limit) and paid a customary rate of 6% interest on the actual loan.

Akin to Federal Reserve credit programs, these lending channels furnished banks with yet another source of clearing reserves to meet their customers' long-distance payments demands. Moreover, New York banks could supply these funds more elastically throughout the year and on better (albeit administered) terms than they could find in local domestic exchange markets. In other words, New York reserve agents became "market makers" in domestic exchange, implicitly quoting ask and bid prices in the form of interest on correspondent loans and balances (for the latter see James 1976b and Gendreau 1983).

Though we lack direct evidence on these credit flows, their indirect impact is evident in the data on New York exchange rates in regional markets. By lowering the effective cost of transferring funds across space – in this case through a net flow of funds borrowed from New York banks – this financial innovation further diminished the variability of domestic exchange rates beginning in the mid-1890s and especially after 1900. Consequently, by the advent of the Federal Reserve System, New York exchange during normal times traded at virtually par in Boston, Philadelphia, and even Chicago and commanded a premium of merely 0.05 to 0.1% in San Francisco.

Conclusion: What Role the Fed?

By the eve of the WWI, the foundations of the modern US monetary and payments system had been firmly established. The correspondent banking system reached from coast to coast (and even beyond), and its tiered hierarchy of intermediaries efficiently transmitted payments instruments and the good funds to settle them between banks regardless of their location. At the center of this system were New York correspondents and the NYCHA, which effected the final step of settlement through ledger entries to banks' correspondent accounts (for a sharply divergent view on this "system," see Laughlin 1912, p. 12). Thus, even without official government

intervention, New York bank deposits constituted a de facto common (bank) money, generally accepted at nearly par value even in the most remote locations.

Yet, this system was not without its critics, then and today. Contemporaries, in particular, complained about a number of risky practices deployed by correspondents operating in highly competitive markets (New York Clearinghouse Association 1873; James and Weiman 2011). To attract bankers' balances, correspondents paid interest (typically 2%) on these deposits, and though modest, it induced country banks to allocate larger shares of their interest-earning assets to correspondent accounts rather than to loans for local customers (Sylla 1969; Redenius and Weiman 2011). They also extended overdraft facilities to their customers by certifying checks in advance of the deposits to cover them and by treating their check deposits as cash even though they had not been cleared and settled (NYCHA 1873; Richardson 2007). Correspondents in turn placed these more seasonally elastic bankers' balances in call loans, which at least during normal times yielded high interest rates and were highly liquid. Meanwhile, country banks dispersed their reserve holdings across a number of correspondents to meet their customers' increasingly wide-ranging transactions but also to exploit these income-earning opportunities (Laughlin 1912, pp. 46–47).

A second frequent criticism, in this case directed mainly at country banks, was the practice of "nonpar clearing" (Jessup 1967; for a more in-depth analysis of this controversy, see Stevens 1998 and James and Weiman 2014). When redeeming their customers' checks delivered through the mails (not at the teller window), banks could and often did charge "exchange," meaning in this case presentment fees, usually a percentage of the check's face value (though often with a minimum of say 10¢ per item). Nonpar banks justified the policy to compensate for loss of their draft business and the additional costs of the check redemption system including the compensating balances to secure a credit line from their correspondent. To avoid these costs, collecting banks could remit checks indirectly through a succession of respondents, until they could be presented in person at the paying bank office.

Reformers such as Hallock (1903) and Cannon (1910) highlighted the inefficiencies in this circuitous routing of checks, though recent evidence suggests that their most flagrant examples were more the exception than the rule (Chang et al. 2008; James and Weiman 2010). An alternative explanation sees the conflict in terms of the allocation of the fixed costs of expanding the check-deposit payment network to rural areas with sparser, seasonal banking markets (Baxter 1983; Weinberg 1997; Lacker et al. 1999). In these environments, banks tended to rely more heavily on revenues from payment services, and because of the divergent competitive conditions at the central hubs versus the more distant locations along the spokes, money center correspondents ultimately bore the cost (McAndrews 1998).

The most serious defect of this interbank network, however, was institutional, what today would be termed a regulatory lag. The prevailing federal legislation – the National Banking Acts of 1863 through 1865 – was designed to create a common national currency, immune from the risks and costs of the multiple banknote moneys of the pre-Civil War era. They were relatively lax, however, when it came to deposit money. The Acts did impose steep deposit reserve requirements ranging from 15% to

25% and restricted banks' lending (to insiders and on real estate) to curb potential default risks. Still, they allowed country and reserve city banks to "pyramid" some of their reserves in interest-earning deposits with higher order agents (for the recurrent refrain against this practice, see Sprague 1977 [1910], pp. 235–236; Myers 1931, pp. 238–239; and most recently Bordo and Wheelock 2013, pp. 63–66). And they did not equip the Office of the Comptroller of the Currency with the authority to more tightly oversee the pivotal reserve agents that anchored the expanding deposit monetary system. Moreover, the OCC had no jurisdiction over the "dual" banking system of depository institutions chartered and more loosely regulated by state governments. As White (1982, 1983, esp. Chaps. 1 and 3) observes, the charter competition between the federal and state governments constrained the power of the OCC to more tightly regulate national banks.

As a consequence, the deposit monetary system was highly vulnerable to the recurrent panics that uniquely afflicted the American banking system during this era (Bordo 1985, p. 73; see also Sprague 1977 [1910], Miron 1986, and Calomiris and Gorton 1991; Wicker 2000; Jalil 2015). And the correspondent banking system was at least partially implicated. At the first signs of a crisis, panicky country banks rushed to withdraw their correspondent balances and so reinforced the liquidity pressures in money centers (Andrew 1908; Smith 1991; Donaldson 1992; Hoag 2005; Carlson 2005; James et al. 2014; for contrary evidence, see Dwyer and Hasan 2007). These fears were often self-fulfilling, as money center banks in extreme cases would resort to "suspending payments," that is, restricting their customers' access to deposit funds. Their decision, of course, simply transmitted these liquidity pressures through their correspondent networks (Richardson 2007; Carlson et al. 2007; Mitchener and Richardson 2016; Calomiris and Carlson 2017; Dupont 2017). The 1907 case is illustrative. The panic originated in the New York financial-banking sector, but the decision by the NYCHA to suspend payments immediately – that is within weeks – precipitated suspension crises in regional reserve centers throughout the country. Put another way, commensurate with their spatial-economic scope, the New York banks nationalized what was a local banking crisis.

From the very onset banks in correspondent centers sought to redress these purported defects, but all of their initiatives fell short. Reformers in the NYCHA, as early as 1857, moved to ban outright the payment of interest on deposits, a precursor to Regulation Q. Because of the Association's unanimity rule, a handful of members, most likely the large correspondent banks, successfully vetoed the repeated attempts (James and Weiman 2011). At the very end of the century, the Boston Clearing House revived a venerable institution, a regional clearinghouse along the lines of the Suffolk Bank and Bank of Mutual Redemption. This innovation was successful in creating a par clearing zone covering nearly all of New England and was successfully adopted in other regional centers such as Kansas City and Atlanta.

Around the same time, NYCHA proposed an ambitious *national* plan to administer the expanding scale and scope of its members' check collection business. Instead of "par" clearing on the one hand and discretionary presentment fees on the other, it mandated that members charge customers for collection services

according to a schedule of standardized rates varying by region or zone (largely a function of distance from New York). Competition over the correspondent business, however, foiled the plan, as respondents could route checks to New York through intermediaries in Philadelphia which were located in New York' par zone and so exempt from these fees.

Local clearinghouses also evolved a lender of last resort monetary policy, usually as a first resort to assisting members during crisis periods. Lacking the authority to create official or banknote money, clearinghouse members instead authorized the issue of a truly inside money, loan certificates (Timberlake 1984; Gorton 1985; Kroszner 2000; Moen and Tallman 2000; Tallman and Moen 2012; Hoag 2011, 2016). A collateralized loan between member banks but co-insured by the entire clearinghouse association, this inside money enabled illiquid (not insolvent) borrowers to settle their daily clearinghouse balances and to devote their excess vault cash to meet customers' reasonable withdrawal demands. During the 1907 suspension crisis, clearinghouses outside of the Northeast went one step further and issued small denomination clearinghouse certificates (that is, local currency) which members could pay out to their customers in lieu of official or national banknote currency (Andrew 1908; James et al. 2014).

Like the other clearinghouse reforms, these private, collective monetary policies were insufficient to stem the most serious banking crises, notably the 1907 panic. Despite unprecedented monetary injections by clearinghouses around the country and the US Treasury, the suspension crisis lingered on for 3 months rather than the usual one (on the role of the Treasury, see Andrew 1907; Heckleman and Wood 2005). Moreover, by fragmenting the national banking and currency payments systems, this prolonged suspension crisis took its toll on the real economy, reinforcing the depths and length of the contraction and ultimately costing an estimated 10–20% in lower real output (James et al. 2013).

The formation of the Federal Reserve System and its early evolution through the 1920s brings to a close this chapter in the history of the US monetary and payments system, literally and figuratively. Like its predecessor the National Banking Acts, the Federal Reserve Act would largely institutionalize the salient features of the mature postbellum monetary-banking system. The two most familiar reforms were the formation of the Gold Settlement Fund and the discount lending window. The former effectively nationalized *the* essential role of New York correspondents and the NYCHA in the interbank settlement system. Unlike the NYCHA, however, the Fed effectively adopted a policy of real-time gross rather than net collective settlement, which assures finality in payments but at the expense of possible gridlock during the settlement period (Emmons 1997; Stevens 1998; McAndrews and Rajan 2000).

The Fed's discount window too supplanted the role of the clearinghouse, in this case its loan committee which authorized and brokered the issue of "emergency" clearinghouse money. As lender of last resort, of course, the Fed had clear advantages, notably its capacity to create official emergency money either in the form of circulating currency or bank reserves (see for example Myron 1986 as well as Richardson and Troost 2009 and Carlson et al. 2011; for a more general analysis

of the Fed's role, see Flannery 1996). While credited with accommodating the seasonal liquidity pressures that afflicted the postbellum banking system, the Fed's discount window too had its limitations. In particular, access was restricted to member, and only those with eligible commercial paper, which tended to exclude vulnerable unit banks operating in rural areas (though they did have indirect access through their correspondents, see Anderson et al. 2018).

The correspondent banking system also left its imprint on the very structure of the Federal Reserve System. In siting the twelve Reserve banks and drawing their district boundaries, the Organizing Committee largely followed the hub-and-spoke structure of the correspondent banking hierarchy (Odell and Weiman 1998; McAvoy 2006; and Jaremski and Wheelock 2017). For example, despite some compromises to geographic proximity, they located the Reserve banks in regional correspondent centers, precisely because of their accessibility to member banks throughout their districts.

By centralizing banks' clearing and settlement transactions through Reserve banks (and their branch offices), the Fed could further rationalize the check collection system. Notably, they substituted telegraphic transmission on secure leased lines (the original "Fedwire" payments system) for postal delivery to effect the last settlement leg of the journey. The resulting reduction in transit times from the Fed's organizational and technological innovations nearly halved the average check collection period and enabled banks to economize on the holdings of excess reserves (Gilbert 2000; see also James 1998 and Carlson and Wheelock 2016; Calomiris et al. 2015).

Finally, the Fed extended its mandate to pursue an agenda of "par" clearing and settlement of check payments throughout the country (for recent reviews of this history, see Stevens 1988 and James and Weiman 2014). Like clearinghouses in regional centers, it sought to make bank deposits a truly common money, one that would circulate at par throughout the country not just in a region. The Fed could and did mandate par clearing for all member banks which sent checks for collection through its clearinghouses. After the formation of the Fedwire, it underwrote the residual exchange costs and ended its own practice of charging banks for collection. Still, this subsidy was not sufficient to entice rural banks, especially in the South and Plains States to join the Fed or at least its par list. As in earlier times, they tended to rely more on payments services than lending for income than their counterparts in more developed areas. The Fed's more aggressive efforts to enforce their compliance were ultimately rebuffed by the Supreme Court, which imposed a political limit to the formation of a truly "perfect" US monetary-banking union.

The "par clearing" controversy is another reminder of one of the central lessons of banking-monetary history, that in the end the formation of a "more perfect" or truly common *national* monetary-banking union is more a matter of politics than economics (Capie 1998). For the USA Rockoff (2003) elaborates the reverse case and explains the recurrent contentious political conflicts in the USA over monetary and banking policy – from Jackson's war against the National Bank to the American Bankers Association's staunch opposition to the Fed's par clearing policy – by the division of the country into distinct monetary regions. Viewed in this light, the

evolution of correspondent banking system actually delineates the regional boundaries of a more purely economic-banking union, for it ultimately depends on the decisions of individual banks and local banking associations whether or not to join the network and on what terms. As we have shown in this chapter, this private or more economic union virtually spanned the continental USA by 1914, but would not be complete until 1980 with the passage of the Monetary Control Act.

Cross-References

- ▶ Privately Issued Money in the United States
- ▶ The Evolution of US Monetary Policy

References

Anderson H, Calomiris CW, Jaremski M, Richardson G (2018) Liquidity risk, bank networks, and the value of joining the Federal Reserve System. J Money Credit Bank 50(1):173–201

Andrew AP (1907) The treasury and the banks under Secretary Shaw. Q J Econ 21(4):519–568

Andrew AP (1908) Substitutes for cash in the panic of 1907. Q J Econ 22(4):497–516

Atherton L (1971) The frontier merchant in Mid-America. University of Missouri Press, Columbia

Bank for International Settlements S (2003) The role of central bank money in payments systems. Bank for International Settlements, Basel

Baxter WF (1983) Bank interchange of transactional paper: legal and economic perspectives. J Law Econ 26(3):541–588

Bech ML, Chapman JTE, Garratt RJ (2010) Which bank is the 'central' bank? J Monet Econ 57(3):352–363

Bensel RF (1990) Yankee leviathan: the origins of central state authority in America, 1859–1877. Cambridge University Press, New York

Board of Governors, Federal Reserve System (1918) Federal reserve inter-district collection system; banks upon which items will be received by federal reserve banks for collection and credit. Jan. 1, 1918. Government Printing Office, Washington, DC

Bodenhorn H (1992) Capital mobility and financial integration in Antebellum America. J Econ Hist 52(3):585–610

Bodenhorn H (2000) A history of banking in antebellum american: financial markets and economic development in an era of nation-building. Cambridge University Press, New York

Bodenhorn H (2002) Making the little guy pay: payments-system networks, cross-subsidization, and the collapse of the Suffolk system. J Econ Hist 62(1):147–169

Bodenhorn H (2003) State banking in early America: a new economic history. Oxford University Press, New York

Bordo MD (1985) The impact and international transmission of financial crises: some historical evidence, 1870–1933. Riv Stor Econ 2(2):41–78

Bordo MD, Jonung L (1987) The long-run behavior of the velocity of circulation: the international evidence. Cambridge University Press, New York

Bordo MD, Jonung L, Siklos PL (1997) Institutional change and the velocity of money: a century of evidence. Econ Inq 35(4):710–724

Bordo MD, Wheelock DC (2013) The promise and performance of the federal reserve as a lender of last resort. In: Bordo MD, Roberds W (eds) The origins, history, and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, New York, pp 59–98

Bryant HB, Stratton HD, Packard SS (1863) Bryant & Stratton's counting house book-keeping. Ivison, Blakeman, Taylor & Company, New York
Burrows EG, Wallace M (1999) Gotham: a history of New York City to 1898. Oxford University Press, New York
Calomiris CW, Mason JR (2008) Resolving the puzzle of the underissuance of National Bank notes. Explor Econ Hist 45(4):327–355
Calomiris CW, Carlson M (2017) Interbank networks in the national banking era: their purpose and their role in the panic of 1893. J Financ Econ 125(3):434–453
Calomiris CW, Gorton G (1991) The origins of banking panics: models, facts, and bank regulation. In: Hubbard RG (ed) Financial markets and financial crises. University of Chicago Press, Chicago/London, pp 109–174
Calomiris CW, Kahn CM (1996) The efficiency of self-regulated payments systems: learning from the Suffolk system. J Money Credit Bank 28(4, Part 2):766–797
Calomiris CW, Jaremski M, Park H, Richardson G (2015) Liquidity risk, bank networks, and the value of joining the federal reserve system. National Bureau of Economic Research, Working paper series, 21684. National Bureau of Economic Research, Cambridge, MA
Carlson M (2005) Causes of bank suspensions in the panic of 1893. Explor Econ Hist 42(1):56–80
Cannon JG (1910) Clearing houses. Government Printing Office, Washington, DC
Capie F (1998) Monetary unions in historical perspective: what future for the euro in the international financial system. Open Econ Rev 9(Suppl 1):447–465
Carlson M, Wheelock DC (2016) Did the founding of the Federal Reserve affect the vulnerability of the interbank system to contagion risk? Bank for International Settlements, Basel
Carlson M, Mitchener KJ, Richardson G (2011) Arresting banking panics: Federal Reserve liquidity provision and the forgotten panic of 1929. J Polit Econ 119(5):889–924
Carter SB, Gartner SS, Haines MR, Olmstead AL, Sutch R, Wright G (eds) (2006) The historical statistics of the United States: millennial edition. Cambridge University Press, New York
Catterall RCH (1903) The second panic states. University of Chicago Press, Chicago
Champ B, Wallace N, Weber WE (1992) Resolving the national bank note paradox. Fed Reserve Bank Minneap Q Rev 16(2):13–21
Champ B, Smith BD, Williamson S (1996) Currency elasticity and banking panics: theory and evidence. Can J Econ 29(4):828–864
Chang HH, Danilevsky M, Evans DS, Garcia-Swartz DD (2008) The economics of market coordination for the pre-Fed check-clearing system: a peek into the Bloomington (IL) Node. Explor Econ Hist 45(4):445–461
Cohen BJ (1998) The geography of money. Cornell University Press, Ithaca
Colwell S (1860) The ways and means of payment: a full analysis of the credit system, with its various modes of adjustment, 2nd edn. J.B. Lippincott & Co., Philadelphia
Conzen MP (1977) The maturing urban system in the United States, 1840–1910. Ann Assoc Am Geogr 67(1):88–108
Cooke T (1897) Distribution of small banks in the west. Q J Econ 12(1):70–72, 105–118
Curry LP (1968) Blueprint for modern America; non-military legislation of the first Civil War Congress. Vanderbilt University Press, Nashville
Dewey DR, Shugrue MJ (1922) Banking and credit; a textbook for colleges and schools of business administration. The Ronald Press Company, New York
Donaldson RG (1992) Sources of panics: evidence from the weekly data. J Monet Econ 30 (2):277–305
Dowd K (1994) Competitive banking, bankers' clubs, and bank regulation. J Money Credit Bank 26(2):289–308
Dupont B (2017) Bank networks and suspensions in the 1893 panic: evidence from the state banks and their correspondents in Kansas. Financ Hist Rev 24(3):265–282
Dwyer GP, Hasan I (2007) Suspension of payments, bank failures, and the nonbank public's losses. J Monet Econ 54(2):565–580

Economides N (1993) Network economics with application to finance. Fin Mkts, Inst Instrum 2 (5):89–97
Emmons WR (1997) Recent developments in wholesale payments system. Fed Reserve Bank St Louis Rev 79(6):23–43
Flannery MJ (1996) Financial crises, payment system problems, and discount window lending. J Money Credit Bank 28(4, Part 2):804–824
Fraas A (1974) The second bank of the United States: an instrument for an interregional monetary union. J Econ Hist 34(2):447–467
Freyer TA (1976) Negotiable instruments and the federal courts in antebellum American business. Bus Hist Rev 50(4):435–455
Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1867–1960. Princeton University Press, Princeton
Garbade K, Silber WL (1979) The payment system and domestic exchange rates: technological versus institutional change. J Monet Econ 5(1):1–22
Garber P, Weisbrod S (1990) Banks in the market for liquidity. National Bureau of Economics Research, Working paper series, 3381. NBER, Cambridge, MA
Gendreau BC (1983) The implicit return on bankers' balances. J Money, Credit, Bank 15(4):411–424
Gilbert RA (1983) Economies of scale in correspondent banking: note. J Money Credit Bank 15(4):483–488
Gilbert RA (2000) The advent of the federal reserve and the efficiency of the payments system: the collection of checks, 1915–1930. Explor Econ Hist 37(2):121–148
Goldenweiser EA (1925) Federal Reserve system in operation, 1st edn. MacGraw-Hill, New York
Goodfriend MS (1990) Money, credit, banking, and payment system policy. In: Humphrey DB (ed) The U.S. payment system: efficiency, risk and the role of the Federal Reserve. Kluwer, Boston, pp 247–277
Goodhart C (1988) The evolution of central banks. The MIT Press, Cambridge, MA
Gorton G (1985) Clearinghouses and the origin of central banking in the United States. J Econ Hist 45(2):277–283
Gorton G (1996) Reputation formation in early bank note markets. J Polit Econ 104(2):346–397
Gorton G (1999) Pricing free bank notes. J Monet Econ 44(1):33–64
Green EJ, Todd RM (2001) Thoughts on the Fed's Role in the Payments System. Fed Reserve Bank Minneap Q Rev 25(1):12–27
Hallock JC (1903) Clearing out-of-town checks in England and the United States. Private, St. Louis
Hamilton A, United States, Department of the Treasury (1821) Official reports on Publick credit, a National Bank, manufactures, and a mint. William McKean, Philadelphia. https://fraser.stlouisfed.org/title/3647. Accessed on 26 Apr 2018
Hammond B (1957) Banks and politics in America from the Revolution to the Civil War. Princeton University Press, Princeton
Hammond B (1970) Sovereignty and an empty purse: banks and politics in the Civil War. Princeton University Press, Princeton
Heckelman JC, Wood JH (2005) Political monetary cycles under alternative institutions: the Independent Treasury and the Federal Reserve. Econ Polit 17(3):331–350
Helleiner E (2003) The making of national money: territorial currencies in historical perspective. Cornell University Press, Ithaca
Hepburn AB (1903) History of coinage and currency in the United States and the perennial contest for sound money. The Macmillan, New York
Hoag C (2005) Deposit drains on "interest-paying" banks before financial crises. Explor Econ Hist 42(4):567–585
Hoag C (2011) Clearinghouse membership and deposit contraction during the Panic of 1893. Cliometrica 5(2):187–203
Hoag C (2016) Clearinghouse loan certificates as interbank loans in the United States, 1860–1913. Financ Hist Rev 23(3):303–324

Jalil AJ (2015) A new history of banking panics in the United States, 1825–1929: construction and implications. Am Econ J Macroecon 7(3):295–330
James JA (1976a) The conundrum of the low issue of national bank notes. J Polit Econ 84(2):359–367
James JA (1976b) A note on interest paid on New York bankers' balances in the postbellum period. Bus Hist Rev 50(2):198–202
James JA (1978) Money and capital markets in Postbellum America. Princeton University Press, Princeton
James JA (1998) Did the fed's founding improve the efficiency of the U.S. payments system? Commentary. Fed Reserve Bank St Louis Rev 80(3):143–150
James JA (2016) Payment Systems. In: Claude D, Haupert M (eds) Handbook of Cliometrics. Springer, Berlin, pp 353–373
James JA, Weiman DF (2005) Financial clearing systems. In: Nelson RR (ed) The limits of market organization. Russell Sage Foundation, New York, pp 114–155
James JA, Weiman DF (2010) From drafts to checks: the evolution of correspondent banking networks and the formation of the modern U.S. Payments System, 1850–1914. J Money, Credit, Bank 42(2–3):237–265
James JA, Weiman DF (2011) The National Banking Acts and the transformation of New York City banking during the Civil War era. J Econ Hist 71(2):338–362
James JA, Weiman DF (2014) Political economic limits to the fed's goal of a common national bank money: the par clearing controversy revisited. Res Econ Hist 30:91–134
James J, McAndrews J, Weiman D (2013) Wall Street and Main Street: the macroeconomic consequences of New York bank suspensions, 1866–1914. Cliometrica 7(2):99–130
James J, McAndrews J, Weiman D (2014) Banking panics, the 'derangement' of the domestic exchanges, and the origins of central banking in the United States, 1893 to 1914. Unpublished mss., Barnard College
Jaremski M (2011) Bank-specific default risk in the pricing of bank note discounts. J Econ Hist 71(4):950–975
Jaremski M (2013) State banks and the National Banking Acts: measuring the response to increased financial regulation, 1860–1870. J Money, Credit, Bank 45(2–3):379–399
Jaremski M, Rousseau PL (2013) Banks, free banks, and U.S. economic growth. Econ Inq 51 (2):1603–1621
Jaremski M, Wheelock DC (2017) Banker preferences, interbank connections, and the enduring structure of the Federal Reserve System. Explor Econ Hist 66(1):21–43
Jessup PF (1967) The theory and practice of nonpar banking. Northwestern University Press, Evanston Ill
Kahn CM, Roberds W (1999) Demandable debts as means of payment: banknotes versus checks. J Money Credit Bank 31(3, Part 2):500–525
Kahn CM, Roberds W (2002) The economics of payment finality. Fed Reserve Bank Atlanta Econ Rev 87(2):1–12
Kahn CM, Roberds W (2009) Why pay? An introduction to payments economics. J Financ Intermed 18(1):1–23
Kahn CM, Quinn S, Roberds W (2016) Central banks and payment systems: the evolving tradeoff between cost and risk. In: Bordo MD, Eitrheim O, Flandreau M, Qvigstad JF (eds) Central Banks at a Crossroads: what can we learn from history? Studies in macroeconomic history. Cambridge University Press, New York, pp 563–609
Kashyap AK, Rajan R, Stein JC (2002) Banks as liquidity providers: an explanation for the coexistence of lending and deposit-taking. J Financ 57(1):33–73
Kemmerer EW (1910) Seasonal variations in the relative demand for money and capital in the United States: a statistical study. Government Printing Office, Washington, DC
Kinley D (1910) The use of credit instruments in the United States. Government Printing Office, Washington, DC

Klein JJ (1911) The development of mercantile instruments of credit in the United States. J Account (pre-1986) 12(8):594–607
Knodell J (1988) Interregional financial market integration and the bank note market, 1815–1845. J Econ Hist 48(2):287–298
Knodell J (1998) The demise of central banking and the domestic exchanges: evidence from antebellum Ohio. J Econ Hist 58(3):714–730
Knodell J (2003) Profit and duty in the Second Bank of the United States' exchange operations. Financ Hist Rev 10(1):5–30
Knodell J (2006) Rethinking the Jacksonian economy: the impact of the 1832 bank veto on commercial banking. J Econ Hist 66(3):541–574
Knodell J (2010) The role of private bankers in the US payments system, 1835–1865. Financ Hist Rev 17(2):239–262
Knodell JE (2017) The Second Bank of the United States: "Central" banker in an era of nation-building, 1816–1836. Routledge, New York
Kroszner RS (2000) Lessons from financial crises: the role of clearinghouses. J Financ Serv Res 18 (2/3):151–171
Lacker JM, Walker JD, Weinberg JA (1999) The Fed's entry into check clearing reconsidered. Fed Reserve Bank Richmond Econ Q 85(2):1–32
Laughlin JL (ed) (1912) Banking reform. The National Citizens' League, Chicago
Lerner AP (1947) Money as a creature of the state. Am Econ Rev 37(2):312–317
Lockhart OC (1921) The development of interbank borrowing in the National Banking System, 1869–1914, Parts I and II. J Polit Econ 29:138–160 222–140
McAndrews J (1997) Network issues and payment systems. Fed Reserve Bank Phila Bus Rev:15–25
McAndrews J (1998) Direct presentment regulation in payments. Res Econ 52(3):311–326
McAndrews J, Rajan S (2000) The timing and funding of fedwire funds transfers. Fed Reserve Bank N Y Econ Policy Rev 6(2):17–31
McAndrews JJ, Roberds W (1995) Banks, payments, and coordination. J Financ Intermed 4(4):305–327
McAvoy MR (2006) How were the Federal Reserve bank locations selected? Explor Econ Hist 43 (3):505–526
Miron JA (1986) Financial panics, the seasonality of the nominal interest rate, and the founding of the Fed. Am Econ Rev 76(1):125–140
Mitchener KJ, Richardson G (2016) Network contagion and interbank amplification during the great depression. National Bureau of Economic Research
Moen JR, Tallman EW (2000) Clearinghouse membership and deposit contraction during the panic of 1907. J Econ Hist 60(1):145–163
Mundell RA (2002) Monetary unions and the problem of sovereignty. Ann Am Acad Pol Soc Sci 579(1):123–152
Myers MG (1931) The New York money market, volume 1: origins and development. Columbia University Press, New York
New York Clearing House Association (1873) Report to the New York Clearing House Association of a committee upon reforms in the banking business. W. H. Arthur & Co., New York
Norman B, Shaw R, Speight G (2011) The history of interbank settlement arrangements: exploring central banks' role in the payment system. Bank of England working paper. Bank of England, London
Odell KA, Weiman DF (1998) Metropolitan development, regional financial centers, and the founding of the fed in the Lower South. J Econ Hist 58(1):103–125
Phillips RJ, Swamy PAVB (1998) Par clearance in the domestic exchanges: the impact of national bank notes. Res Econ Hist 18:121–144
Quinn S, Roberds W (2007) The Bank of Amsterdam and the leap to central bank money. Am Econ Rev 97(2):262–265

Quinn S, Roberds W (2008) The evolution of the check as a means of payment: a historical survey. Fed Reserve Bank Atlanta Econ Rev 93(4):1–28
Rand McNally and Company (1915) Rand McNally bankers' directory. Rand McNally and Company, Chicago
Redenius SA (2002) Hubs and spokes: network effects and the Formation of Regional Banking Centers. Unpublished mss. Bryn Mawr College
Redenius SA (2004) Banknote redemption networks under the National Banking Act, 1863–1874. Bryn Mawr College, Bryn Mawr
Redenius SA (2007) Designing a national currency: antebellum payment networks and the structure of the National Banking System. Financ Hist Rev 14(2):207–227
Redenius SA, Weiman DF (2011) Banking on the periphery: the Cotton South, systemic seasonality, and the limits of national banking reform. In: Rhode PW, Rosenbloom JL, Weiman DF (eds) Economic evolution and revolution in historical time. Stanford University Press, Stanford, pp 214–242
Redlich F (1968 [1947]) The molding of American banking: men and ideas. Johnson Reprint Corporation, New York
Richardson G (2007) The check is in the mail: correspondent clearing and the collapse of the banking system, 1930 to 1933. J Econ Hist 67(3):643–671
Richardson G, Troost W (2009) Monetary intervention mitigated banking panics during the great depression: quasi-experimental evidence from a Federal Reserve district border, 1929–1933. J Polit Econ 117(6):1031–1073
Rockoff H (2003) How long did it take the United States to become an optimal currency area? In: Capie FH, Wood GE (eds) Monetary unions: theory, history, public choice. Routledge, New York, pp 76–103
Rolnick AJ, Smith BD, Weber WE (1998) Lessons from a laissez-faire payments system: the Suffolk Banking System (1825–58). Fed Reserve Bank St Louis Rev 80(3):105–116
Rose AK (2000) One money, one market: the effect of common currencies on trade. Econ Policy 15 (30):7–33
Rose AK, Engel C (2002) Currency unions and international integration. J Money, Credit, Bank 34(4):1067–1089
Rose AK, van Wincoop E (2001) National money as a barrier to international trade: the real case for currency union. Am Econ Rev 91(2):386–390
Selgin GA, White LH (1994) How would the invisible hand handle money. J Econ Lit 32(4):1718–1749
Shambaugh JC (2006) An experiment with multiple currencies: The American monetary system from 1838–60. Explor Econ Hist 43(4):609–645
Shreve BJ (1898) Country checks and country bank accounts. Bank Mag 56(2):221–231
Smith BD (1991) Bank panics, suspensions, and geography: some notes on the "contagion of fear" in banking. Econ Inq 29(2):230–248
Spahr WE (1926) The clearing and collection of checks. The Bankers Publishing Co., New York
Sprague OMW (1977 [1910]) History of crises under the National Banking System. Augustus M. Kelley Publishers, Fairfield
Stevens E (1998) Non-par banking: competition and monopoly in markets for payments services. Federal Reserve Bank of Cleveland working paper series. Federal Reserve Bank of Cleveland, Cleveland
Sylla RE (1969) Federal policy, banking market structure, and capital mobilization in the United States, 1863–1913. J Econ Hist 24:657–686
Sylla R (1976) Forgotten men of money: private bankers in early U.S. history. J Econ Hist 36(1):173–188
Sylla RE (2006) The transition to a monetary union in the United States, 1787–1795. Financ Hist Rev 13(1):73–95
Tallman EW, Moen JR (2012) Liquidity creation without a central bank: clearing house loan certificates in the banking panic of 1907. J Financ Stab 8(4):277–291

Timberlake RH (1984) The central bank role of clearinghouse associations. J Money, Credit, Bank 16(1):1–15
U.S. Comptroller of the Currency (1890–1892) Annual report of the comptroller of the currency. Government Printing Office, Washington, DC
Watkins LL (1929) Bankers' balances: a study of the effects of the Federal Reserve System on banking relationships. A.W. Shaw, Chicago
Weber WE (2000) Disaggregated call reports for U.S. national banks, 1880–1910. Research Department, Federal Reserve Bank of Minneapolis, Minneapolis. http://cdm16030.contentdm.oclc.org/utils/getdownloaditem/collection/p16030coll4/id/6/filename/7.xls
Weber WE (2003) Interbank payments relationships in the antebellum United States: evidence from Pennsylvania. J Monet Econ 50(2):455–474
Weiman DF, James JA (2007) The political economy of the US monetary union: the Civil War era as a watershed. Am Econ Rev 97(2):271–275
Weiman DF, James JA (2018) Towards a more perfect monetary-payments union: the Civil War as a second American Revolution. Unpublished mss., Barnard College
Weinberg JA (1997) The organization of private payment networks. Fed Reserve Bank Richmond Econ Q 83(2):25–43
White EN (1982) The political economy of banking regulation, 1864–1933. J Econ Hist 42(1):33–40
White EN (1983) The regulation and reform of the American banking system, 1900–1929. Princeton University Press, Princeton
Wicker E (2000) Banking panics of the gilded age. Cambridge University Press, New York
Van Fenstermaker J (1965) The development of American commercial banking, 1782–1837. Bureau of Economic and Business Research. Kent State Univerity, Kent

Currency Unions

29

Anders Ögren

Contents

Abstract

Currency unions have been a recurring phenomenon in monetary history. The most basic definition of a currency union is when two or more sovereign nations share a common currency. Even though a currency union shares many characteristics with international monetary regimes based on fixed exchange rates (such as the Bretton Wood system), it is the most extreme case of a fixed exchange rate as it also implies sharing a common unit of account between the participating countries. While the economic rationale for currency unions focuses on gains in monetary efficiency from decreasing transaction costs, history shows that

A. Ögren (✉)
Department of Economic History, Lund University School of Economics and Management, Lund, Sweden
e-mail: anders.ogren@ekh.lu.se

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_28

political reasons also played a key role. This chapter discusses the theoretical foundations of monetary unions (the so-called optimum currency area theory) and analyzes the most important historical cases of currency unification, both domestic (the USA, Germany) and international. Finally this chapters discusses the key lessons that can be drawn from the history of currency unions and their implications for the Economic and Monetary Union.

Keywords
Central banks · Fiscal systems · Monetary theory · Monetary policy · Optimum currency areas

Introduction

The most basic definition of a currency union is when two or more sovereign nations adopt a common currency. Even though a currency union shares many characteristics with international monetary regimes based on fixed exchange rates (such as the Bretton Wood system), a currency union is an extreme case as it also implies sharing a common unit of account between the participating countries. A currency union is also different from other extreme cases of fixed exchange rates such as *dollarization* or *currency boards* – monetary regimes usually adopted to eradicate high inflation from the domestic economy. *Dollarization* is when a nation decides to import the currency of another nation (historically, the US Dollar in most cases) and use it as legal tender instead of a domestic currency. A *currency board* is when a nation ties its domestic currency to an international high powered currency (as the USD) and guarantees a full reserve coverage of the domestic currency issued. That is that for each unit of national currency that is issued by the nation's monetary authority, it has to hold the same amount of the international high powered money (for instance USD) in its reserves. At difference from dollarization and currency boards. While *dollarization* and *currency boards* are based on a unilateral commitment by the government of the adopting nation, a currency union rests on the joint commitment of the governments of all participating nations.

The economic rationale behind currency unification has been worked out by the so-called optimum currency area (OCA) theory. This on one hand emphasizes the advantages of a currency union in terms of monetary efficiency (through decreasing transaction costs), on the other hand discusses its possible costs caused by the fact that governments had to give up a potentially important instrument – the nominal exchange rate – which could be used for macroeconomic stabilization. However, there exist also political reasons for pursuing currency unification. In fact, most national currencies are not only tied together by an economic rationale, such as the integration of the domestic market, but are based also on common fiscal systems.

In this chapter, I first discuss the basic tenets of the OCA theory. The second section focuses on historical cases of currency unifications, both national (the USA and Germany) and international (the Latin Monetary Union and the Scandinavian Monetary Union). Finally, the lessons from those historical experiences provide a

background for the analysis of the Economic and Monetary Union (EMU) adopted since 1999 by an increasing number of European countries.

Currency Unions in Theory

Optimum Currency Areas (OCA)

The theory of optimum currency areas (OCA) is the most widely used analytical framework to assess the benefits and costs of a currency union. In short, the theory states that the following criteria are necessary for a group of economies (regions or countries) to qualify for an OCA:

> (i) regions should be exposed to similar sources of economic disturbance (common shocks); (ii) the relative importance of these shocks across regions should be similar (symmetric shocks); (iii) regions should have similar responses to common shocks (common responses); and (iv) if regions are subject to region specific economic disturbances (idiosyncratic shocks), they need to be capable of quick adjustment. The basic idea is that regions satisfying (i)-(iv) will have similar business cycles, so a common monetary policy response would be optimal. (Kouparitsas 2001, p. 1)

The economic rationale for monetary unions draws on theories of exchange rate regimes. In the case of OCA, it emerged from discussions on the pros and cons of a floating versus a fixed exchange rate in the USA after World War II. Rockoff (2000, p. 4) suggests that Milton Friedman's *The Case for Flexible Exchange Raters* (1953) and Leland Yeager's *Exchange Rates Within a Common Market* (1959) were the seminal contribution that triggered the debate. Both Friedman and Yeager (among others) argued for floating exchange rates as they would allow a country to run an independent monetary policy to achieve stability in prices and employment. In their view, however, they allowed for exceptions as far as small open economies were concerned, since they were considered not large enough to profit from flexible exchange rates.

It was this debate that led Robert Mundell to formulate his *Theory of Optimum Currency Areas* (1961). What was new in Mundell's work was that, instead of assuming a national currency to represent a given currency area, he asked under what conditions would a single currency be economically most efficient. Of course, what is evaluated as economic efficiency depends on the objective of the currency and here Mundell emphasized not only transaction costs but also stabilization policy, i.e., the ability of a government to pursue full employment. In fact, if the focus was only on gains in monetary efficiency, Mundell pointed out, the logical conclusion would be that the entire world should share a common currency (Mundell 1961, p. 662). However, adding macroeconomic stabilization implied that, in order to qualify as an optimum currency area, regions or countries should be characterized by complete factor mobility. Thus, limited international factor mobility provides a rational for national currencies. In the words of Mundell:

> *If the world can be divided into regions within each of which there is factor mobility and between which there is factor immobility, then each of these regions should have a separate currency which fluctuates relative to all other currencies.* (Mundell 1961, p. 663)

This implies that national currencies (a currency area that perfectly overlaps with a nation's borders) could be theoretically justified only in case of perfect factor mobility within national economies. For factor immobility within a currency area would create a problem of imbalances:

> *The argument works best if each nation (and currency) has internal factor mobility and external factor immobility. But if labor and capital are insufficiently mobile within a country then flexibility of the external price of the national currency cannot be expected to perform the stabilization function attributed to it, and one could expect varying rates of unemployment or inflation in the different regions.* (Mundell 1961, p. 664)

Taking over from Mundell, the OCA theory was further developed by McKinnon (1963), who focused on the relevance of trade openness (the ratio of tradable to nontradable goods) for the analysis. The next important step came with the work of Peter Kenen (1969) who argued for the importance of fiscal redistribution for an OCA. His main argument was that, as Mundell's factor mobility was in fact labor mobility, this would require homogenous labor forces and products, which implied that OCAs were likely to be quite small. By adding product diversification to the model, he highlighted the problem of asymmetric shocks. A symmetric shock is no problem for a currency area as it hits all parts of the economy equally and thus affects prices and wages to the same extent across the union. However, in a diversified economy that uses the same currency, a shock will affect different sectors with different amplitude and thus put different stress on prices and wages. An asymmetric shock could be absorbed only if labor could move across sectors with minimum costs. However, should this not be the case, fiscal transfers would be necessary to mitigate its effects. It should be noted that Mundell's goal was not to argue for currency unions, quite the contrary; in fact, based on his OCA model, he contended that a floating exchange rate would be a logical choice: *"This carries the argument for flexible exchange rates to its logical conclusion."* (Mundell 1961, p. 663).

In short, the OCA theory highlights two important economic factors that are necessary for a smooth working of any currency area (including individual nations): (1) high factor (labor) mobility within the area and (2) fiscal transfers to even out the effects of asymmetric shocks (see also Krugman 2013, pp. 441–3). These features require also a high degree of political integration, which explains why in practice currency areas as a rule overlap with nations. In mainstream monetary theory, as in the OCA literature, money is viewed as an exogenous neutral item that serves the sole purpose of minimizing transaction costs. This is clearly illustrated by Mundell:

> *Mill, like Bagehot and others, was concerned with the costs of valuation and money-changing, not stabilization policy, and it is readily seen that these costs tend to increase with the number of currencies. Any given money qua numeraire or unit of account fulfills this function less adequately if the prices of foreign goods are expressed in terms of foreign*

> *currency and must then be translated into domestic currency prices. Similarly, money in its role of medium of exchange is less useful if there are many currencies; although the costs of currency conversion are always present, they loom exceptionally large under inconvertibility or flexible exchange rates.* (Mundell 1961, p. 662)

That currency (or money) is viewed as a neutral exogenous device implies that money should be created at the crossroad of the cost and benefits of trade in relation to transaction costs. This is a strong assumption which is not empirically supported. If currencies (money) were created from the need to minimize transaction costs in trade, the link between currencies and nations would not be as solid as it is in fact. Empirical studies suggest that even a deeply integrated economy, such as the USA today, would not meet the OCA criteria in the absence of interstate fiscal transfers operated by the federal government (see, for instance, Kouparitsas 2001; Rockoff 2000; Sala-i-Martin and Sachs 1991). However, this notion has led economists not to reject the OCA theory, but rather to conclude that it is the world which either has too many currencies, or too few.

> *Traditionally, each country had its own currency, and only one currency circulated in each country. Monetary unions were rare, and, therefore, the surge in the number of countries in the post-war period generated a large increase in the number of currencies circulating in the world. In 1947 there were 76 countries in the world, today there are 193, and, with few exceptions, each country has its own currency. Unless one believes that a country is, by definition, an "optimal currency area," either there were too few currencies in 1947 or there are too many today. In fact, the increasing integration of international markets implies that the optimal number of currencies would tend to decrease, rather than almost triple as it has.* (Alesina and Barro 2000, p. 1)

Despite its inconsistencies, the OCA framework has been widely adopted in the literature on the Economic and Monetary Union (EMU), which in fact spawned an increased interest in the theory since the 1990s (see for instance Mongelli 2002, 2008). At the same time, this new stream of research often took the existence of national currencies as a starting point. An example is Frankel and Rose (1998), who used the OCA theory to test under what conditions countries would be more or less suitable for entering into a currency union. On the base of a panel of 20 countries over a time of 30 years, they showed that countries with more intense trade also tended to share business cycles – and thus were more able to form a currency union. Of course such closer links could also be generated by a higher degree of political integration.

Criticism of the OCA Theory: The Role of the State and the Importance of Long-Term Capital Flows

The OCA theory is widely accepted as the benchmark framework within which monetary unions such as the EMU should be studied. However, it is not exempt from criticism. In fact, some contend that its basic assumptions as to the nature of money and the process of money creation are seriously flawed. One important critique has

been formulated by Charles Goodhart (1998), who emphasizes the importance of the State and its fiscal power in order to understand the process of monetary creation. His main criticism is directed to the monetary theory that underpins the OCA approach, which he labels "M-theory." Goodhart argues that its microeconomic foundations, such as rational, individual, and atomistic agents pursuing transaction cost minimization, fails to understand the origins and the function of money. He explains:

> Much of the economic analysis of moving to EMU has been undertaken within the context of the Optimal Currency Area paradigm. This is the spatial/geographic counterpart of the currently dominating model of the nature and evolution of money, here termed M theory, whereby money is viewed as having developed from a private sector cost minimization process to facilitate trading. Here, I argue, first, that there is a second, cartalist, or C theory alternative, which is empirically more compelling. Second, I claim that this approach can predict observed relationships between sovereign countries and their currencies better than the OCA model. (Goodhart 1998, p. 407)

While economists generally subscribe to the M-theory (with the exception of post-Keynesians), the cartalist approach – which underlines the central role of the State (or alternative sovereign authorities) as issuer of the money – is very popular in disciplines that study monetary phenomena from a more empirical perspective, such as economic historians, historians, political scientists, and anthropologists. Goodhart (1998, p. 409) explains this polarization with the normative preference of mainstream economists for private market solutions, as opposed to the complexity of political factors. His main critique is that economic theories based only on optimizing rational agents fail to recognize the role of the State in providing the necessary requirements for well-functioning markets, such as the legal framework and enforcement mechanisms.

> *While it is, of course, the relationship between taxation and the demand for money that the C-form theory emphasizes, it should also be remembered that it is the maintenance of law and order, the form and enforcement of contracts, and the whole infrastructure of regulation within society, that allows the epiphenomena of (organized) (private sector) markets to occur at all.* (Goodhart 1998, p. 418)

In a monetary system in which the pivotal role of the State is absent (as in the M-theory), there should be no link between sovereign States and currencies. Any State should be able to have any number of currencies and any currencies should circulate in any number of States. The fact that the OCA literature always starts out from the assumption of "one currency, one sovereign Government" shows its very limited explanatory power. It is thus ironic that economists tend to analyze currency unions, such as the EMU, in an OCA perspective (Goodhart 1998, p. 420).

While the historical record supports the C-theory by consistently illustrating the strong correlation between strong States and strong currencies (including successful currency reforms), as well as the fact that currencies tend to breakup as States collapse, the M-theory has nothing to say about the rise and fall of currencies. For it, once a currency has been established as a result of optimization and cost minimization, there can be only a stable equilibrium (Goodhart 1998, p. 414).

In short, the key premises of the OCA approach based on the M-theory miss out the complexity of organized society and political authority, and the link between political sovereignty, the fiscal system, and money creation that is a critical dimension for the analysis of currency unions. In fact, the OCA framework has been gradually expanded to include new factors – fiscal transfers among them, as mentioned earlier – in order to improve its explanatory power (see, for instance, Broz 2005; Dellas and Tavlas 2009). In any case, its supporters emphasize that the theory is less positive than normative; it does not aim to explain how a currency union actually works but to shed light on the necessary conditions for its functioning in a purely economic perspective. For Goodhart, however, the overwhelming historical evidence in support of the link between political sovereignty and money creation makes the cost-benefit analysis derived from the OCA-model an element of second-order importance compared to the overall political factors that are embedded in a currency (Goodhart 1998, p. 424). And the augmentation of the OCA theory with fiscal transfers is in itself an admission that currencies are not merely the result of agents' desire to minimize transaction costs. The fact that recent macroeconomic analysis connects the issue of money to domestic public debt (Reinhardt and Rogoff 2009) is a further step towards recognizing the key role of the State in the monetary process and the need to include it in the analysis of monetary unions.

Neglecting the link between currencies and capital flows unrelated to trade is a second serious weakness of the OCA approach and of the empirical research on currency unions based on it. This is somehow surprising, since most of capital flows in modern economies are not related to trade but to investments of all kinds, including sovereign debt. As a consequence, the external value of a currency is not set by spot market transactions in relation to trade (as they would be in accordance to the OCA model) but rather by the futures market. This reflects the complex interaction between many possible determinants such as expected economic growth, government debt, expected monetary and fiscal policies, or political risk – all variables that are missing in the OCA model. In a similar fashion, the OCA approach does not capture the tensions generated by long-term imbalances between surplus and deficit regions or countries. Historically, in the process of real convergence, catching-up developing countries have been net capital importers. While this might not be a problem as long as capital exporting countries are willing to lend, "sudden stops" in the access to foreign capital markets experienced by capital importing countries (Calvo 1998; Edwards 2004) may generate strong tensions within currency unions (Merler and Pisani-Ferry 2012). In this perspective, the problem is not an asymmetric shock but rather an asymmetric trend (Saint-Paul 2010). The example of the Gold Standard is illustrative. In that period, as they caught up with more advanced economies, Scandinavian countries recorded structural trade deficits for decades but managed to remain on a convergence path thanks to their continuous access to long-term international capital (Ögren 2009). In case of sudden stops, countries could go off gold (as Portugal did, by example). No such flexibility nor escape clause exists in currency unions such as the EMU (Ögren and Øksendal 2012).

Reasons for Establishing Currency Unions

If we look at the establishment of monetary unions empirically, there are several historical cases that provide insight into the possible reasons for forming currency unions. The degree of political integration seems to be a common factor for successful experiments of monetary integration, as well as a common reason for pursuing this objective. Some monetary unions were part of nation building; as a consequence, they are no longer viewed as experiments of monetary unions today. This is a mistake as monetary unification in some cases preceded complete political integration. Bordo and Jonung (1999, p. 6) argue that a distinction can be made between "national" and "multinational" currency unions. In a "national" union, currency unification is a step towards uniting political and monetary sovereignty, such as was the case for Germany, Italy, Switzerland, and the USA. In the "multinational" case, the currency union consists of several independent nations that use the same currency without a common monetary authority. Historical examples of such currency unions are the Latin Monetary Union (LMU) and the Scandinavian Currency Union (SCU). Bordo and Jonung (1999, p. 7) also claim that "multinational" unions will not be able to survive in the long run without such political unification and that the future of the EMU will depend on its ability to "closely resemble a national monetary union." Again this underlines the problem of evaluating a monetary union from a theory based only on minimizing transaction costs in relation to trade.

In his analysis of the Latin Monetary Union, Flandreau (2000) stresses that the economic rationale for forming the union was more important than it had been hitherto recognized. But this economic rationale was also more complex than just striving to minimize transaction costs in relation to trade. At least equally important were foreign debt issues and the ties linking capital exporters and importers. Usually foreign debt is not denominated in the debtor's currency (unless it is a big economy with an internationally trusted currency as in the case of the US), which means that for the debtor a falling exchange rate increases the debt and the cost to service the debt. For the creditor country, this may increase the risk of debt default. Thus there are economic incentives for debtor as well as creditor countries to share a common unit of account and form a currency union.

In mainstream economics, however, the main economic rationale for currency unification is strongly related to trade costs. The standard approach to test for the effect of currency unions on trade is the so-called "gravity model." This is a variant of Newton's theory of gravity that assumed that the force in-between two objects is a function of their masses and the square of their distance. In economics, "gravity models" have been used since the 1960s to understand bilateral trade and migrations (Smith 2002). By adding control variables to the original variables of mass (GDP as a measure of the economy's size) and distance, such as borders, shared language, common or different political systems or a common currency, the gravity model contributes to explain observed patterns of bilateral trade or factor flows.

The "gravity model" has been widely used to assess the impact of the EMU on trade with mixed results. The seminal paper of this extensive literature was Rose

(2000a), whose opening sentence was: "*What is the effect of a common currency on international trade? Answer: Large.*" Rose used a dataset of bilateral trade flows for 186 countries in the period 1970–1990. His gravity model augmented with a currency union (or common currency) dummy found a strong positive effect on international trade. Similar results were obtained in companion papers that used an expanded dataset (Rose 2000b; Rose and Engel 2002; Glick and Rose 2002).

Rose's findings stirred a lively debate. His critics raised concerns that his results could be a consequence of the choice of econometric techniques than an actual fact, i.e., that the result was not robust in statistical sense. By instance, Torsten Persson (2001, p. 434) argued:

> *The impact of a common currency on trade can be grossly mismeasured if countries that belong to currency unions are systematically different from those that do not, and if the relationship between trade and its observable determinants are complex. I argue that such complications are plausible and likely to distort the empirical result . . .. Using techniques designed to be robust in this situation, I find that the effects of common currency on international trade are considerably less dramatic and much less precisely estimated.*

This debate continued in the following years (see, for instance, Nitsch 2002; Pakko and Wall 2001; Rose 2001a, b; Smith 2002). It should be noted that Rose initially was careful about drawing any conclusions from his research in the specific case of the EMU (Rose 2000a, 2001). Still the implicit argument was that national currencies were, as predicted by the OCA-theory, impeding international trade and that currency unions would boost trade by removing these frictions. This line of reasoning was visible in the title of Rose and Wincoop (2001): "National Money as a Barrier to International Trade: The Real Case for Currency Union."

More recent research is much less sanguine about the trade effects of currency unions, EMU included. By instance, results of Glick and Rose's (2015a, b) "Currency Unions and Trade: A Post-EMU Mea Culpa" were rather gloomy about the EMU's effect on trade. More importantly, results were found to be very sensitive to different econometric methodologies, so that "no substantive reliable and robust effect of currency union on trade" could be found. The published version (Glick and Rose 2016) was more optimistic as EMU was found to have boosted exports by 50%, and the results seemed less sensitive to empirical methods and more reliable.

Yet questions about whether currency unions significantly foster trade or to what extent a gravity approach is the optimal tool to pursue this issue remain open. However, it is out of question that currency unions do remove the transaction costs that stem from having to deal with different currencies, as well as exchange rate risk. Historically, it has been observed that in regions with large trade relationships with regions of another country, agents are more inclined to adopt the same currency simply because it is more practical and less risky (Flandreau 1996).

Flandreau also ran a gravity test for the Latin Monetary Union (LMU) and the Scandinavian Currency Union (SCU), and found no direct effects on trade in both cases (Flandreau 2000, pp. 29–31). It should be noted that certain features in the design of the LMU may have introduced frictions on cross-country coin circulation and thereby to some extent hamper the union's effect on trade (Timini 2018). Also,

as explained above, the inclusion of capital flows not related to trade confirmed the existence of other economic rationales for the establishment of the LMU. Nontrade-related capital flows are not captured by traditional gravity models, in spite of their utmost relevance for modern economies. By instance, the historical case of the SCU shows that the inclusion of banknotes and drawings of central bank bills (beyond coins) fostered immediately financial market integration (Øksendal 2007).

To sum up, this extensive empirical literature suggests a number of critical issues. First, the creation of currency unions can be motivated by, and affect, not only trade-related financial flows but also other equally (or more) important capital flows related directly to the international capital market. Second, the effects of currency unions may also depend critically on how unions are designed. This implies that trade gravity models augmented with simple membership dummies might not be adequate to capture the complex effects of monetary integration.

Third, political reasons might be important for the success (or failure) of monetary integration. The importance of the currency union for political integration can be seen in the historical case of the USA. At the same time, political integration was essential for the success of the US currency union (see, for instance, Grubb 2006; Michener and Wright 2006; Rockoff 2000; Rousseau 2006; Sylla 2006). Political ambitions can, however, be challenged by the economic reality, as we can appreciate in the present case of the EMU. The next section provides an overview of historical cases of currency unions to understand this complex interplay of economics and politics.

"National" Currency Unions

The USA

The USA offers a challenging view on currency unions as it is in fact the world's most successful currency. It can of course be debated to what extent the USA really is a currency union, but at the outset, the USA certainly faced many of the problems that the EMU (and other currency unions) are facing today (see for instance Eichengreen 1992). It is thus worth to include a lengthier discussion on the US monetary union here.

First of all, despite the high level of political integration of the US states, the development of the US currency union was a long process that went through several stages. This, if anything, points to the fact that monetary unification is a far more complex process than usually acknowledged. The breakup of the monetary union during the American civil war also shows that monetary unions are by no means irreversible. At large, as will be seen in this section, currency unions have historically been closely linked to political integration (and disintegration).

The literature on the US monetary union discusses different stages, all related to the changing historical circumstances: the adoption of a common unit of account and a common currency in relation to the independence; the adoption of a common legal tender currency pegged to the unit of account after the Civil War; a common

payment system for clearing and interbank transactions with the establishment of the Federal Reserve in 1913; and finally in the 1930s – in relation to the Great Depression – the implementation of institutional changes that enhanced fiscal transfers between surplus and deficits regions (see, for instance, Michener and Wright 2006, p. 19; Rockoff 2000; Weiman 2006, p. 13).

Starting with the independence, an important point is that it was by no means given that monetary issuance in the USA should be centralized. Prior to the independence and the creation of the US Dollar (USD), each colonial government issued their own currency in the form of credit notes. These currencies were usually denominated after British model and issued with a floating exchange rate as the notes were backed by the colonies' future tax revenues. To what extent the colonial currencies were successful or not, or more precisely whether the decision to tie the USD to a bimetallic standard was a key for its success, are matters of debate (Grubb 2006; Sylla 2006). But the consensus is that the colonial system guaranteed a stable monetary value as long as there was no extraordinary pressure on the monetary system, such as war finance.

The first attempt to issue a common US currency was the so-called Continental currency that was issued by the Continental Congress during the American Revolution (1760–1791). The Continental currency can be regarded as a failure as it was overissued and quickly fell in value until it became more or less worthless in 1780. The reason for this was, however, that the Continental currency was issued as a source of funds for the war of independence (during the American Revolutionary War 1775–1783). Moreover, the Continental Congress had no autonomous fiscal capacity at the time as the power to tax belonged to individual States. The latter in turn simultaneously issued their own currencies to fund their expenses during the war.

The Continental currency and the State currencies were formally based on a fixed exchange rate during the war. Each note was redeemable at a fixed value into Spanish Dollars or their equal value in gold or silver. To what extent the promise to redeem the notes in silver or gold was credible is hard to tell, but in practice it quickly became impossible (Sylla 2006, pp. 76–79). It is even possible that this experience served to underpin the need for a stable common currency, as stated by Rousseau.

> *At that time unbacked issues of fiat money that helped to finance the American victory in the Revolutionary War gave way to an inflationary spiral, debt depreciation and a scarcity of real money balances. The need to unify the nation's currency and to restore the public's confidence in it weighed heavily in the minds of the forefathers as they drafted a constitution that forbade emissions of paper money by individual states in favour of committing to a securely backed transactions asset.* (Rousseau 2006, p. 97)

The first experience of a common US currency was thus not a success. It pointed to the fact that fixing its value in terms of specie (or any other currency) failed to discipline its issuance, especially in times of war. Yet a common currency, the USD, was established in 1792 and became one of the cornerstones of the US financial and economic success.

According to Michener and Wright, the American monetary union was originally not designed to deliver a uniform medium of exchange – as is the case in the modern meaning of a currency union. Instead, the objective was to instigate a common unit of account that also served as a standard of deferred payment. Part of the reason for this was that during colonial times agents expressed monetary values in their own colonies' unit of account regardless of what was de facto used in the transaction (Michener and Wright 2006, pp. 20, 24). Thus the main issue was to ensure that the unit of account for contracts and transactions had a uniform meaning. In order to achieve this result, it was necessary to define the unit of account as a standard of weights, i.e., a standardized metallic coin. The constitution in article 1, section 8 stated that (among other things) the federal government ought: *"To Coin Money, regulate the Value thereof, and of foreign Coin, and fix the Standard of Weights and Measures;. . .*". As to the States, article 1, section 10 stated (among other things) that: *"No State shall. . . coin money; emit Bills of Credit; make any Thing but gold and silver Coin a Tender in Payment of Debts; . . ."* (Michener and Wright, pp. 34–35).

To settle a uniform standard of account was even more important as the most important part of the money supply was not composed of legal tender cash, such as specie or specie coins, but consisted of bills of exchange and bank money (notes and deposits). As such these were privately issued, not recognized as legal tender but still possible to redeem for specie at the issuing banks (Michener and Wright 2006, p. 38).

The USD was thus adopted at a fixed exchange rate based on a bimetallic standard where the gold-to-silver parity was set at 15-to-1. The individual States gave up on their right to issue money; the question is why they were willing to do so. With the Constitution and the USD, the federal governments took over the existing stock of debt accumulated by individual states during the Revolutionary War, as well as the fiscal burden of servicing it. According to Sylla (2006, p. 85), this made individual States less inclined to issue money, raise taxes or compete for the tax base with the Federal government, now entrusted with the power to tax.

A fundamental aspect of any currency system is its relation with the fiscal system. So the fact that the Federal State enjoyed exclusive rights to tax imports and shared the tax base with the individual States further strengthened the monetary union. Scholars agree that the loss of income suffered by individual States was offset by the relief on their debt burden and the possibility to charter banks. In addition, the individual States did not need to fund their defense and gained access to a common market with one single unit of account (Grubb 2006; Rousseau 2006; Sylla 2006).

It should also be noted that after independence, the USA also established a kind of central bank with the creation of the First Bank of the Unites States (in Philadelphia), chartered for 20 years by the US Congress in 1791. Today, the presence of central banks is seen as pivotal for any monetary system, but this was not the case in the late eighteenth century. Political resistance towards too strong federal power at the expense of the autonomy of the states made opponents criticize the charter of the

Bank from its outset. At the time of the renewal of its charter in 1811, opponents managed to block it and the bank had to close. In 1816, the Second Bank of the USA was chartered (also in Philadelphia) for another 20 years and it took over the role of the prior First Bank – but as its predecessor it was denied renewed charter in 1836 and closed in 1841.

Of course a currency union, as any political economy project, also affects the basic political and economic structure in terms of winners and losers. Even more so perhaps in the case of the US, as the monetary union was part of the entire political project of the US Constitution. Echoing the view of Charles Beard, an influential American historian, in his *Economic Interpretation of the Constitution of the United States* (published in 1913), Weiman (2006, p. 12) emphasizes how "*the shifting balance of monetary power between the states and national governments . . . pitted the interests of farm households on the periphery of the bourgeoning American market system, who were often in debt or at least short of specie, against those of creditors and merchants at its core*."

With the Civil War (1861–1865), the US monetary union was shattered. California (which politically stayed within the Union) kept the old system whereas both the North States and the South created their own respective unit of accounts and media of exchange. The pressing needs for war finance meant that both the North and the South issued large amounts of non-backed credit bills, which made the old monetary system de facto untenable. After the war, the federal government reintroduced a bimetallic standard, a common unit of account and a standard of deferred payment, thus returning to the old monetary union. As the economy became more monetized and the use of dollar-denominated US coins and uniform dollar-denominated notes of national banks became widespread, the union developed toward a more uniform system of transaction media – a trend that ultimately led in the twentieth century to the emergence of the modern US system based on a uniform medium of exchange denominated in dollars (Michener and Wright 2006, p. 40).

In a way the establishment of the FED, with its mandate to issue the currency (notes and coins) of the USA and to manage the monetary system, can be seen as the moment when the US monetary union became a modern monetary union (Sylla 2006, pp. 72–73).

Another important lesson of the US case points to the need for fiscal integration (see, for instance, Rockoff 2000; Sala-i-Martin and Sachs 1991). Rockoff showed that this stage of the US monetary union was achieved during the Great Depression. Against the background of frequent region-specific shocks, the institutions adopted in the 1930s – "a system of inter-regional fiscal transfers and some form of deposit insurance, or regionally sensitive lender-of-last-resort facilities" – prevented the unfolding of banking crises and the ensuing monetary contraction, thus limiting the scope for political battles over the reform of the overall monetary system (Rockoff 2000, pp. 36–37). The lesson that "multinational" projects of currency unions should draw from the US experience is clear: in the absence of a system of fiscal transfers, they will have small chances of surviving in the long run (Capie 1998; Bordo and Jonung 1999).

Germany

Three countries that resemble the US experience – a currency union as part of a scheme aiming at a larger political union – are Germany, Italy, and Switzerland.

The process of monetary unification of the hundred of German independent states that existed at the beginnings of the nineteenth century roughly lasted between 1834 and 1871. As the interest in the German historical experience was revived by the launching of the EMU in the 1990s, some contended that the political integration of Germany forestalled its economic integration. This was not the case, however; on the contrary in the German case economic integration preceded political unification (Capie 1998, p. 82; Holtfrerich 1993, p. 518). The movement towards monetary unification began with the establishment of a common free trade area, the so-called *Zollverein* (custom union) in 1834. It was followed by the Munich Coin Treaty in 1837 and the Dresden Coinage Convention in 1838, after which the states had to choose between the Thaler or the Gulden as their monetary unit. In relation to this, a "union coin" (*Vereinsmunz*) was minted, which equaled in value 2 Thalers and 3.5 Gulden. In practice, this meant the establishment of common unit of accounts and a fixed exchange rate between all the states in the customs union. With the Vienna Coinage Treaty in 1857, coinage was standardized and paper money was made convertible into specie at a fixed rate. However, it was only after the formation of the German Reich in 1871, in the aftermath of the Franco Prussian war, followed by the act of 1873 and the establishment of the Reichsbank (the German central bank) in 1876 that a common currency in terms of a common medium of exchange was adopted. It should be noted, however, that the fiscal system was badly designed as the central government lacked the power to tax and obtained limited funding from individual states, thus making it incapable of reacting to crises (Capie 1998, pp. 82–83).

"Multinational" Currency Unions

The Quest for Intranational Currency Unions in Nineteenth Century Europe

The idea of a common currency at transnational level became widely accepted in the second half of the nineteenth century. A common currency was seen then as an answer to the problem of large transaction costs stemming from the need to operate with several currencies in the context of a globalizing economy. Historically, the problem was related to the notion that money represented goods and was therefore a measure of their value. The challenge was to understand the relationship between different currencies subject to continuous changes in value. With the consolidation of nation states and global economic integration, the idea of moving from national currencies to supranational currencies was logically based on the quest to lower these transaction costs. Perhaps more important in practice was the fact that governments also had to fight the problem of deviations between coins' face value and their

metallic contents – a situation in which efficient arbitrage could trigger massive exports of coins, thus leading to circulation shortages.

The first modern proposals to create a common currency for Western Europe and the USA were advanced in various international statistical conferences that took place in the 1850s (London 1851, Brussels 1853 and Paris 1855). The focus was on currency as a measure that, like others, needed to be harmonized by the adoption of a common unit of account. To achieve this objective, however, it was also necessary to harmonize the content of the coins, i.e., to adopt a common money for circulation. Otherwise the pressing problem of monetary arbitrage and shrinking circulation of coins would not be solved. In the 1855 Paris meeting, delegates from all countries signed a declaration in support of conformity of coinage. This recommendation was renewed at the statistical conferences in Vienna 1859 and in Berlin 1863. In the latter, the delegates prompted governments to study the basis for a common monetary system. Consensus in Western Europe and the USA on the desirability of a common monetary system was rising. The British proposal to use the Pound Sterling as reference, but also to adopt the French metric system, was adjourned by the US Congress, not because the idea of a common currency system was criticized but because the proposal included the continuation of the bimetallic standard in the USA and continental Europe. At the time, falling gold prices made silver coins disappear from circulation and an increasing number of countries, including the USA, opted for a pure gold standard (Russell 1898, pp. 18–22).

The Latin Monetary Union

It was in this context that the Latin Monetary Union (LMU) was formed in 1865. It aimed to harmonize the coins of Belgium, France, Italy, and Switzerland. They were all neighboring countries and important trading partners to France, but they all had coins with different silver contents. The LMU was expanded to include the Papal State in 1866 and Greece and Romania in 1867. It should be noted that even before the formal decision to adopt a currency union, many neighboring countries had already adopted the French 10 Franc gold coin as legal tender (Piedmont in 1816, Belgium in 1832, Switzerland in 1850, and Italy in 1862), so that the switch into a formal currency union with France was not farfetched.

The LMU was however a looser arrangement than what we may think of as a currency union today. First, it only concerned full bodied gold and silver coins that were holding the same value and metallic content in all countries. Second, national banks of issue were all private institutions, and private agents were not obliged to accept foreign coins. As far as common overarching rules, these were limited to the standard of minted coins and a limit on over issuance of token coins in silver (bronze, copper coins, or notes were not included in the LMU) (Einaudi 2000, p. 287).

Meanwhile, the discussions on an international currency for most of Europe and the USA continued in the famous meetings in Paris in 1865 and 1867. At the meeting during the International exhibition in Paris 1865, the outspoken aim was again to establish a universal monetary system for the USA and Europe through a universal

coinage. To reach an agreement on a common currency, two issues had to be addressed: (1) How to divide the denominations of coins, i.e., between the main unit of account and its fractions. (2) How to choose a monetary standard and a universal coin with a common metallic content (value). On the first issue, the French pushed for the metric system which already had been adopted by many countries as a result of the meetings in the 1850s.

However, the debate focused mainly on the choice of the monetary standard. Britain was on a gold standard since 1819 whereas France and the other countries in the LMU were on a bimetallic standard. Prussia and the Scandinavian countries were on a silver standard. As gold findings in the USA had made gold cheaper relative to silver, silver coins were disappearing from circulation, and most countries viewed the gold standard as the feasible alternative. The problem was how to harmonize the switch from silver to gold on an international scale.

This would also affect the decision about which "main coin" should be used as the standard of such an international currency union. Here the choice was between the British gold sovereign and the French 10 Franc gold coin. Whereas the British sovereign was the main coin of the world's most important reserve currency and was used throughout the British Empire, the French 10 Franc coin was the preferred coin on the European continent. However close a universal coinage might have been at that time, the project was shattered by the Franco-Prussian war of 1870–1871.

Some regard the LMU as a failure, since it failed to prevent some member countries (such as Italy) from "free riding" on others. The "free riders" budgets and government debts were "not kept in control" and they "imposed costs" on other LMU members by issuing small denomination coins and notes far from their face value. Hence the argument is that the LMU broke down because the "free riding problem" converted it into a "noncredible" monetary arrangement (Bae and Bailey 2011). There may be a grain of truth in this assessment, but controlling the fiscal behavior of member sovereign nations was never part of the system. In fact, the LMU was built on the same assumptions that the OCA theory would emphasize one century later, that is, to adopt a common currency to reduce transaction costs and promote trade. The LMU was, as pointed out by Marc Flandreau, a truly liberal project, based on voluntary participation and with no specific mechanism to constrain the behavior of member countries. In its original design, its rules were few and simple: governments should limit the issue of debased silver coins and share information to monitor compliance (Einaudi 2001). As the insufficiency of these rules became clear very soon, additional rules were negotiated, such as the "liquidation clause" introduced in the Treaty's revision of 1885 – a reform that, in the view of contemporary observers (Willis 1901, p. 236) led to a de facto abrogation of the Union and its substitution with a "new monetary league." However, the liberal design was preserved and it represented "one of the most important stumbling blocks" of the Union, which dissuaded the French government from pursuing its further extension (Flandreau 2000, p. 42).

When assessing historical monetary arrangements, some tend to assume that there is a "one size fits all" solution to monetary problems – that is, there exists one monetary system that invariably provides the optimal outcome under different

economic and political conditions. By doing so, they fail to acknowledge that any monetary regime is – to quote Barry Eichengreen's assessment of the Gold Standard – "*a socially constructed institution whose viability hinge on the context in which it operate*" (Eichengreen 2008, p. 29).

In spite of its structural difficulties and declining relevance (mainly due to a diminishing role of gold and silver coins in member countries' monetary base), the LMU was a resilient institution. It survived the Franco-Italian trade war of the 1880s and a Greek default, allowed some degree of flexibility in exchange rate management of weak currency countries, and lasted until the outbreak of WWI in 1914 (although it survived *de jure* until 1927), which is longer than many other monetary regimes.

The Scandinavian Currency Union

The Scandinavian Monetary Union was also an offspring of the international monetary meetings in the 1850s and 1860s. The Scandinavian countries were on a silver standard and experienced the same problem of arbitrage and coins going out of circulation in relation to their internal exchange rates, as some coins were valued more in silver than others. The solution, in line with the idea of an international monetary standard, was to harmonize the unit of accounts as well as the specie content of coins in the Scandinavian countries. Representatives of the Scandinavian countries regularly attended the international conferences, and the idea of participating in an international monetary system was generally accepted.

When the possibility of an international monetary union evaporated, the Scandinavian countries quickly followed Germany, who had adopted *de jure* the gold standard in 1873. In the same year, Denmark and Sweden formed the Scandinavian Currency Union(SCU), switched to gold, adopted a common unit of account (*Krone* (DKK) and *Krona* (SEK)), and minted full bodied coins with the same metallic content and the status of legal tender in both countries. Norway joined the union in 1875. As was the case with the LMU, the union initially dealt only with coinage and coin circulation, and the three national banks of issue continued to be independent. However, the scope of the union was expanded in 1885 with the adoption of a common clearing system that allowed national banks to draw drafts on each other to settle international balances. As mentioned above, this step was fundamental for the full integration of the Scandinavian financial markets (Øksendal 2007). In 1901, the union was extended to include also the circulation of central bank notes.

Just like the LMU, the SCU ended in practice with the outbreak of WWI and the suspension of the Gold Standard, although *de jure* it lasted until 1924. Also in the case of the SCU, there has been some discussions about "free riding" within the union in relation to over issuance of bank notes and token coins. But the union did work well until WWI, surviving the breakup of the political union in 1905 that had been forced upon Norway by Sweden. After WWI, when notes no longer were redeemable for gold, small token coins continued to pose a problem as they were subject to different valuations in the different countries but could be exchanged at

par in accordance with the rules of the union. This problem was eventually solved by the complete dismantling of the union in 1924. Clearly, the union helped to increase economic and political integration, not least because governors of each national bank started to meet regularly (Talia 2004).

As seen above, the rationale for establishing the SCU was both economic and political. From the late eighteenth century, the idea of Nordic unity started to circulate in small intellectual circles. The movement gained momentum during the nineteenth century and from the 1860s "Scandinavianism" (*Skandinavismen*) became very influential with increased cooperation and exchanges in different areas between the three countries. In fact this movement has been seen as a factor that contributed to the establishment of the SCU.

The Economic and Monetary Union (EMU) and the Lessons from History

As it is well known, the macroeconomic and institutional conditions that countries had to meet in order to participate in the Economic and Monetary Union (EMU) were the subject of harsh negotiations and were officially formulated in the Maastricht Treaty signed in February 1992 (James 2012, pp. 265–323). One of the key aspects of the Treaty was the so-called "no bail out" clause, under which member states should not be liable for, nor assume, the commitments or debts of any other. The logic behind the inclusion of this clause in a currency union was based on fears of "free riding" and the risk of moral hazard: knowing beforehand that they could be rescued by the rest of member countries, governments would have less incentives to abide by the rules and act prudently. This in turn would seriously question the credibility of the EMU. Thus, the clause was intended as a firewall against the risk of debt mutualization and its potential spillovers over the credibility of the entire project. Accordingly, the so-called "convergence criteria" were designed so that they would clearly signal a commitment of participating governments to the principles of noninflationary policies and "sound" finance.

The five criteria were: (1) price stability (inflation should not exceed by more than 1.5% that of the three best performing countries), (2) sound public finances (the government budget should not exceed 3% of GDP), (3) sustainable public finances (the government debt should not exceed 60% of GDP), (4) durability of convergence (the long-term interest rate should not exceed by more than 2 percentage points that of the three best performing countries), and (5) exchange rate stability (participation in the narrow band of the exchange rate mechanism (ERM I) of the European Monetary System for at least 2 years with no strong deviations from central rate and no devaluation; for countries joining the EMU at a later stage, the same criterion applies with respect to a peg to the Euro under ERM II) (European Commission 2018).

Interest in past examples of monetary unions was revived in the run-up to the EMU, and even more so after the emergence of the Euro Area crisis in 2008. Today, it is widely accepted that the largest problems in the design of the EMU is the lack of

fiscal integration, the absence of a mechanism to support deficit countries, and the fact that the European Central Bank lacks a clear mandate to act as lender of last resort in times of crisis. In some way, the design of the EMU eliminated the stabilizers that existed at national level (see De Grauwe 2012, 2013; De Grauwe and Ji 2015; Krugman 2013). It is thus reasonable to ask to what extent the policy makers behind the EMU project had learnt anything from history.

Looking at the final result, perhaps the only lesson drawn from past currency unions such as the LMU and the SCU was that a single currency could not be managed effectively by multiple independent central banks. The option of fiscal transfers (in the form of "solidarity funds") from surplus to deficit countries, as well as the possibility for the ECB to act as a lender of last resort, were considered during preparatory works but were not implemented in the end.

The road to Maastricht had been long (see for instance Eichengreen 1993). In the 1960s, the main idea behind plans for monetary integration was that it was the necessary complement of the common market. Without monetary integration it would have been too easy to enhance exports at the expense of other participants by manipulating the exchange rate. This in turn would question the legitimacy of the whole project of European economic integration. Another recurrent economic argument was the benefits of creating a European reserve currency to compete with, or to complement, the dominant international currencies of the time – i.e., gold, the US dollar, and the British pound (European Commission 1962, pp. 87–88).

On the contrary, the crises of the 1970s raised awareness of potential problems. The Keynesian paradigm was still guiding the economic thinking and the legacy of the Bretton Wood system – which had been built from the ashes of the policy failures of the interwar period – was fresh memory. New ideas emerged, such as an Exchange Stabilization Fund as well as a more influential role for the European Monetary Cooperation Fund to deal with imbalances within the union (see European Commission 1966). The idea that a devolution not only of monetary but also fiscal prerogatives by national governments in favor of the EEC institutions also gained momentum (European Commission 1975). As we know, this idea was abandoned in the plans of the 1980s in favor of an approach in which each country was supposed to pursue convergence criteria to defend its fixed exchange rate with respect to the common currency (European Commission 1989, 1990).

There are several reasons for this choice in the design of the EMU, but arguably the two most important factors were political resistance and a change in the paradigm of economic theory. Political resistance against the EMU was (and is) based on the fear of losing influence – which would be the case if fiscal authority was transferred from national governments to the EU. Add to that the above-mentioned fear of "free riding," which means that surplus countries would have to transfer resources to deficit countries. By eliminating institutions that could be deployed as emergency funds by deficit countries (such as an Exchange Stabilization Fund), the idea of a currency union became more attractive for surplus countries.

The idea that the economic discipline of each country independently would be sufficient to uphold the EMU also started to gain support in economic theories from the late 1970s. Or more precisely that adhering to the EMU would impose an

external constraint on participating countries that would discipline their policy. This idea is evident in the so-called Optica Reports (Basevi et al. 1976, 1977; see also Thygesen 1978) published in 1976 and 1977. The theoretical background was the so-called "monetary approach to the balance of payments" and the idea that "purchasing power parity" holds. Money was again seen as only a nominal item not related to the real economy. In this new approach, joining a currency union was simply to commit to a common inflation rate. As long as a government wanted to adhere to the common currency, it had to use monetary policy to fight inflation in order to maintain a fixed exchange rate. The "credibility" of this commitment to the peg was crucial to make a currency union work without inducing market forces from putting pressure on the currency. This "credibility" in turn was based on clear and transparent rules that the participating nations adhered to. The fact that countries subject themselves to a fixed exchange rate, in this case a common currency, thus meant that they were forced to discipline their economic policy to align their inflation rate to the one of the other countries or opt out. Thus the "credibility" of the EMU was reached by the fact that each country accepted to "tie their hands" (Giavazzi and Pagano 1988).

As discussed in the previous sections, however, many economists and economic historians were not convinced that surrendering monetary prerogative would be enough to make the EMU work (Bordo and Jonung 1999; Capie 1998; Goodhart 1998). A quote from the seminal paper by Forrest Capie summarizes these concerns:

> *The monetary criteria are not problematic. In a world where inflation has been the devil and has been attacked remorselessly almost to the point of temporary defeat there is less to worry about. Most countries in and outside of Europe have low inflation rates. It is the fiscal criteria that pose, and are likely to continue to pose, problems.* (Capie 1998, p. 80)

It is also well known that in reality, fiscal convergence criteria have been repeatedly violated by most members of the EMU with relative impunity. By example, already in 2002, France and Germany, the two biggest economies in the EMU, violated the second criteria of "sound public finances."

However, history suggests that the making of a successful and enduring currency union is a long process. The economic and political cost of leaving a currency union as the EMU should also not be underestimated (Eichengreen 2007). If there is enough political will to sustain the EMU, it will survive. Thus, the political circumstances are of more importance for the EMU than the economic circumstances, even if it is much easier to manage a currency union during expansionary cycles than in recessions. It is in times of crisis that this political will is tested. Again, the lesson that Europe can draw from US history is clear:

> *It is far from clear… that the United States was an optimal currency area. This pattern held until the 1930s when institutional changes, such as increased federal fiscal transfers (which pumped high-powered money into regions that were losing reserves) and bank deposit insurance, addressed the problem of regional banking shocks. Political considerations, of course, ruled out separate regional currencies in the United States.* (Rockoff 2000)

Concluding Remarks

History suggest that currency unions and political (fiscal) integration go hand in hand. It is possible to form currency unions that are less politically integrated, but such unions are more resembling fixed exchange rate systems managed by independent central banks. The EMU today is in some ways a hybrid since monetary policy is centralized in the hands of an independent central bank (ECB), but there is a very limited degree of fiscal integration and each country is supposed to fend for itself in the struggle to remain in the Union.

It should be noted that more or less all "national" currencies work as politically integrated currency unions supplemented by fiscal transfers from surplus to deficit regions. These can be found both in centralized and federal polities, and are especially necessary during crises that affect asymmetrically different economic sectors and regions in a country. The same rule applies for the sustainability of a "multinational" currency union.

References

Alesina A, Barro RJ (2000) Currency unions. NBER working paper no 7927. https://www.nber.org/papers/w7927

Bae K-H, Bailey W (2011) The Latin monetary union: some evidence on Europe's failed common currency. Rev Dev Finance 1:131–149

Basevi G. et al (1976) The Optica Report 1975. Doc. II/909/75 final. Directorate General for Economic and Financial Affairs [online]. Brussels: European Commission, 16.01.1976 [accessed 29 May 2012]. Available on: http://ec.europa.eu/economy_finance/emu_history/documentation/chapter8/19760116en61opticareport1975.pdf

Basevi G. et al (1977) Inflation and Exchange Rates: Evidence and Policy Guidelines for the European Community. Optica Report 1976. II/855/76, 10 February 1977. [EU Commission – Working Document]

Bordo MD, Jonung L (1999) The future of the EMU: what does the history of monetary unions tell us. NBER working paper no 7365. https://www.nber.org/papers/w7365

Broz T (2005) The theory of optimum currency areas: a literature review. Privredna kretanja i ekonomska politika 104:53–78

Calvo G (1998) Capital flows and capital market crises: the simple economics of sudden stops. J Appl Econ 1(1):35–54

Capie F (1998) Monetary unions in historical perspective: what future for the euro in the international financial system. Open Econ Rev 9:447–465

De Grauwe P (2012) The governance of a fragile Eurozone. Aust Econ Rev 45:255–268

De Grauwe P (2013) Design failures in the eurozone. Can they be fixed? European Commission Economic papers 491. http://ec.europa.eu/economy_finance/publications/economic_paper/2013/ecp491_en.htm

De Grauwe P, Ji Y (2015) Correcting for the Eurozone design failures: the role of the ECB. J Eur Integr 37:739–754

Dellas H, Tavlas GS (2009) An optimum-currency-area odyssey. J Int Money Financ 28:1117–1137

Edwards S (2004) Thirty years of current account imbalances, current account reversals, and sudden stops. IMF Staff Pap 51(suppl. 1):1–49

Eichengreen B (1992) One Money for Europe? Lessons from the US Currency Union Economic Policy, Europe 1992, 5(10):117–187

Eichengreen B (1993) European monetary unification. J Econ Lit 31:1321–1357

Eichengreen B (2007) The breakup of the euro area. NBER working paper no 13393. https://www.nber.org/papers/w13393
Eichengreen B (2008) Globalizing capital. A history of the international monetary system. Princeton University Press, Princeton
Einaudi LL (2000) From the franc to the 'Europe': the attempted transformation of the Latin monetary union into a European monetary union, 1865–1873. Econ Hist Rev 53:284–308
Einaudi L (2001) Money and politics: European monetary unification and the international gold standard (1865–1873). Oxford University Press, New York
European Commission (1962) Mémorandum de la Commission sur le programme d'action de la Communauté pendant la deuxième étape. Bruxelles: Service des publications des Communautés européennes, p 107. http://www.cvce.eu/obj/memorandum_de_la_commission_24_octobre_1962-fr-4bf24e3a-80ca-4886-b8dd-1a4d0a92d411.html
European Commission (1966) The development of a European capital market. Report of a group of experts appointed by the EEC Commission, Brussels
European Commission (1975) Report of the study group, "Economic and monetary union 1980" 8 March 1975 European Commission, Brussels http://www.cvce.eu/obj/the_marjolin_report_brussels_march_1975-en-93d25b61-6148-453d-9fa7-9e220e874dc5.html
European Commission (1989) Report on economic and monetary union in the European Community. Presented April 17, 1989 (commonly called the Delors Plan or Report) By Committee for the Study of Economic and Monetary Union, Brussels
European Commission (1990) One market, one money. An evaluation of the potential benefits and costs of forming an economic and monetary union. Eur Econ 44, Oct 1999, Brussels
European Commission (2018) Convergence criteria for joining. Homepage https://ec.europa.eu/info/business-economy-euro/euro-area/enlargement-euro-area/convergence-criteria-joining_en
Flandreau M (1996) The French crime of 1873: an essay on the emergence of the international gold standard, 1870–1880. J Econ Hist 56:862–897
Flandreau M (2000) The economics and politics of monetary unions: a reassessment of the Latin monetary union, 1865–71. Financ Hist Rev 7:25–43
Frankel JA, Rose AK (1998) The endogeneity of the optimum currency area. Econ J 108:1009–1025
Giavazzi F, Pagano M (1988) The advantage of tying one's hands: EMS discipline and central bank credibility. Eur Econ Rev 32(5):1055–1075
Giavazzi F, Pagano M (1988) The advantage of tying one's hands: EMS discipline and Central Bank credibility. European Economic Review 32(5)1055–1075
Glick R Rose AK (2001) Does a currency union affect trade? The time-series evidence. CEPR discussion paper no 2891
Glick R, Rose AK (2002) Does a currency union affect trade? The time-series evidence. Eur Econ Rev 46:1125–1151
Glick R, Rose AK (2015a) Currency unions and trade: a post-EMU Mea Culpa. CEPR discussion paper no 10615. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=10615
Glick R, Rose AK (2015b) Currency unions and trade: a post-EMU Mea Culpa. NBER working paper no 21535. https://www.nber.org/papers/w21535
Glick R, Rose AK (2016) Currency unions and trade: a post-EMU reassessment. Eur Econ Rev 87:78–91
Goodhart CAE (1998) The two concepts of money: implications for the analysis of optimal currency areas. Eur J Polit Econ 14:407–432
Grubb F (2006) The US constitution and monetary powers: an analysis of the 1787 constitutional convention and the constitutional transformation of the US monetary system. Financ Hist Rev 13:43–71
Holtfrerich C-L (1993) Did monetary unification precede or follow political unification of Germany in the 19th century? Eur Econ Rev 37:518–524
James H (2012) Making the European monetary union. Harvard University Press, Cambridge

Kenen P (1969) The theory of optimum currency areas: an eclectic view. In: Mundell R, Swoboda A (eds) Monetary Problems of the International Economy, pp 41–60. The University of Chicago Press, Chicago
Kouparitsas MA (2001) Is the United States an optimum currency area? An empirical analysis of regional business cycles. Federal Reserve Bank of Chicago WP No 2001-22. https://fraser.stlouisfed.org/title/5285/item/533613
Krugman P (2013) Revenge of the optimum currency area. NBER Macroecon Annu 27(1):439–448
McKinnon RI (1963) Optimum currency areas. Am Econ Rev 53(4):717–725
Merler S, Pisani-Ferry J (2012) Sudden stops in the euro area. Rev Econ Inst 3(3):1–23
Michener R, Wright RE (2006) Development of the US monetary union. Financ Hist Rev 13 (1):19–41
Mongelli FP (2002) 'New' views on the optimum currency area theory: what is EMU telling us? ECB working paper no 138. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp138.pdf
Mongelli FP (2008) European economic and monetary integration, and the optimum currency area theory. European Commission Economic papers 302. http://ec.europa.eu/economy_finance/publications/pages/publication_summary12083_en.htm
Mundell RA (1961) A theory of optimum currency areas. Am Econ Rev 51:657–665
Nitsch V (2002) Honey, I shrunk the currency union effect on trade. World Econ 25:457–474
Ögren A (2009) Financial revolution and economic modernisation in Sweden. Financ Hist Rev 16(1):47–71
Ögren A, Øksendal LF (2012) The euro and the gold: what are the lessons? In: Ögren A, Øksendal L-F (eds) The gold standard peripheries – monetary policy, adjustment and flexibility in a global setting. Palgrave Macmillan, London
Øksendal LF (2007) The impact of the Scandinavian monetary union on financial market integration. Financ Hist Rev 14:125–148
Pakko MR, Wall HJ (2001) Reconsidering the Trade-Creating Effects of a Currency Union. Federal Reserve bank of St. Louis Review, Sep.-Oct., pp 37–45
Persson T (2001) Currency unions and trade: how large is the treatment effect? Econ Policy 16 (33):433–448
Reinhardt C, Rogoff K (2009) This time is different: eight centuries of financial folly. Princeton University Press, Princeton
Rockoff H (2000) How long did it take the United States to become and optimal currency area? NBER historical paper no 124. https://www.nber.org/papers/h0124
Rose AK (2000a) One money, one market: estimating the effect of common currencies on trade. Econ Policy 15:7–33
Rose AK (2000b) Does a currency union boost international trade? Calif Manag Rev 42:52–62
Rose AK (2001a) Honey, the currency union effect on trade hasn't blown up. World Econ 25:475–479
Rose AK (2001b) Currency unions and trade: the effect is large. Econ Policy 16:449–461
Rose A, Engel C (2002) Currency unions and international integration. J Money Credit Bank 34(4):1067–1089
Rose AK, van Wincoop E (2001) National money as a barrier to international trade: the real case for currency union. Am Econ Rev 91:386–395
Rousseau PL (2006) A common currency: early US monetary policy and the transition to the dollar. Financ Hist Rev 13(1):97–122
Russell HB (1898) International monetary conferences. Their purposes, character and results with a study of the conditions of currency and finance in Europe and America during intervening periods, and in their relations to international action. Harper & Brothers Publishers, New York/London
Saint-Paul G (2010) Is the euro a failure? http://www.voxeu.org/index.php?q=node/4999
Sala-i-Martin X, Sachs J (1991) Fiscal federalism and optimum currency areas: evidence for Europe from the United States. NBER working paper no 3855. https://www.nber.org/papers/w3855

Smith C (2002) Currency unions and gravity models revisited. Reserve Bank of New Zealand discussion paper no 2002/07. https://www.rbnz.govt.nz/research-and-publications/discussion-papers/2002/dp2002-07

Sylla R (2006) The transition to a monetary union in the United States, 1787–1795. Financ Hist Rev 13(1):73–95

Talia K (2004) The Scandinavian currency union 1873–1924: studies in monetary integration and disintegration, Dissertation, Stockholm School of Economics

Thygesen N (1978) Inflation and exchange rates: Evidence and policy guidelines for the European community. J Int Econ Rev 8(2):301–317

Timini J (2018) Currency unions and heterogeneous trade effects: the case of the Latin Monetary Union. Eur Rev Econ Hist 22:322–348

Weiman D (2006) Introduction to the special issue on the formation of an American monetary union. Financ Hist Rev 13(1):11–17

Willis H (1901) History of the Latin monetary union. Economic Studies of the University of Chicago, Chicago

The Sterling Area 1945–1972

30

Catherine Schenk

Contents

Abstract

The sterling area was an international monetary system that operated for almost 30 years after the end of the Second World War. Born from wartime exchange controls, it was initially a short-term response to global imbalances in the wake of the war and the failure of the new Bretton Woods institutions to support multilateral trade and payments. From 1945 to 1972, members of the sterling area agreed to maintain fixed exchange rates with sterling, to hold the bulk of their foreign exchange reserves in sterling, and to impose exchange control in common with Britain to protect against possible flight from sterling to other currencies. In return, members enjoyed freer trade with Britain and freer access to British capital than other countries. In the early years, it was defined by Britain's war debts, but through the 1950s these were retired and replaced by fresh accumulations of sterling by other members. But by the 1960s, a weaker pound and waning enthusiasm for monetary cooperation among its members undermined the system, and it became part of the crumbling of the wider Bretton Woods system.

C. Schenk (✉)
Faculty of History, University of Oxford, Oxford, UK
e-mail: catherine.schenk@history.ox.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_29

A concerted multilateral effort supported the gradual retreat from sterling until it was finally abandoned without fanfare in 1972.

Keywords
International currency · Bretton Woods · Sterling · Sterling area · Devaluation

Introduction

In the planning for the peace after the Second World War, the question of how the international monetary system would be organized was a central concern. The promise of a trade-led recovery that would create the foundations for a lasting peace required convertibility and multilateral payments, and this was the focus of US and UK planning from the early 1940s (Schenk 2011). The famous Bretton Woods Conference of 1944 created the institutional setting to try to ensure that countries were able to liberalize their international payments through convertibility on current account while retaining stable exchange rates. This included the International Monetary Fund, which would provide short-term liquidity to allow members to smooth over imbalances in their external payments without resorting to exchange controls or adjusting their exchange rates. The system, however, never came to function as it was designed. The demands of capital for reconstruction and recovery were too great, and the imbalances in the global economy were too prolonged for governments to abandon their exchange controls in a pegged rate system. Instead, groups of countries sought interim solutions to maximize payments and free up trade while protecting themselves against dollar convertibility. This was the context in which the sterling area emerged from wartime exchange and trade controls; it was not a monetary union or a currency union since each member operated its own currency. The sterling area operated alongside the other main regional payments system, the European Payments Union (EPU), which operated for 15 years until the end of 1958. The sterling area, in contrast, was less formalized, but it continued in amended form until the British government turned to floating exchange rates in June 1972. Because the sterling area was so prolonged, it had a profound impact on the operation and development of the Bretton Woods system as it came to operate in the 1950s and 1960s. This chapter reviews the origins and evolution of the sterling area after 1945 and its effects on the global payments and reserves system.

The Sterling Area: Origins

The link between British trade networks, Empire, and sterling stretched back to the nineteenth century classic gold standard. The emergence of the UK as an imperial power was driven by, and in turn reinforced by, the economic power of Britain, so it is no surprise that along with the close commercial relations, which defined the Empire, came close monetary links. The British Empire was never a formal

monetary union, but political ties were accompanied by close monetary relations between Britain and overseas possessions. The enforced commercial integration between Britain and its colonies encouraged monetary dependence as well as political dependence. Starting from the middle of the nineteenth century, the British state established colonial monetary systems in many parts of the Empire, which linked local currency issue to sterling assets held in currency boards (for a critical view of colonial currency boards, see Narsey 2016). The backing for local note and coin issues was invested mainly in British government securities, and the funds were managed from London. During the heyday of the gold standard at the end of the nineteenth century, the predominance of sterling as the currency for international transactions and reserves made it natural for countries within and outwith the Empire to peg their exchange rates tightly to sterling and to use sterling to denominate their international transactions. Beyond the imperial link, other countries in Europe and the Middle East also chose to hold their reserves primarily in sterling and to peg their exchange rates to sterling. In the nineteenth century gold standard, sterling was used as a means of exchange and store of value because of the confidence in the gold value of the British pound and the unrivalled commercial and financial services available in the City of London. After the First World War, however, this was to change.

As Eichengreen and Flandreau (2009) have shown, the weakening of Britain's economic leadership in the interwar period (1919–1939) was marked by competition for international currency status, from the US dollar in particular. Supported by the US authorities, the dollar briefly claimed the largest share of global reserves for countries not associated with the British Empire. Within the British Empire, however, the role of sterling was enhanced. Imperial Preference in tariffs drew together current and former members of the British Empire (excluding Canada) into a mutually beneficial bloc of freer trade and payments. But other countries that traded intensively with the UK also continued to use sterling. With high proportions of trade taking place with Britain, or among the members of the Commonwealth and Empire, it made sense also to stabilize exchange rates with sterling and to hold official reserves in sterling to defend those pegs.

In September 1931 the British government was forced to devalue sterling against gold and the $US. A group of countries including the Empire and Commonwealth (except for Canada) but also Portugal, Scandinavia, and later some Latin American countries that traded heavily with the UK followed by retaining their peg to the pound. This group became known as the sterling bloc, which marks the beginnings of a more deliberately managed link among the economies of the British Empire. As trade protectionism and competitive devaluation threatened to dislocate the international economy, imperial monetary and trade cooperation were addressed at the 1932 Imperial Economic Conference in Ottawa, Canada. The outcome was the famous system of Empire Trade Preference through which members of the Commonwealth and Empire traded with each other on terms more favorable than with the rest of the world. This discriminatory system was to become a major target of American policymakers in the planning for the postwar trading framework. In terms of monetary policy, the British made clear at the conference that they intended to allow the pound

to float against gold and other currencies while seeking to stabilize exchanges within the Empire. The Empire and Commonwealth in turn made clear their willingness to follow. For the rest of the interwar period, the sterling bloc pegged their exchange rates to the pound and floated with it.

Despite its name, the sterling bloc remained an informal association of countries. There was some discussion at the time of an imperial currency or at least an imperial currency area. But this could have encouraged interference in British monetary policy by members of the Commonwealth and Empire, and it could also cause friction within the Commonwealth. The best policy for Britain was to stabilize the exchanges and ensure the attractiveness of sterling in international trade and payments through sound policies rather than introducing an element of compulsion or complicating matters through closer links. In 1933 the Chancellor of the Exchequer (Chamberlain) expressed his view that

> The more I see of this difficult and complicated question (of imperial monetary union) the more I am convinced that it would be wrong to lay down rigid rules in conformity with theoretical considerations. There are too many factors in the case which do not lend themselves to theory. (quoted in Drummond 1981)

In other words, as European states have learned, monetary union is not merely an economic problem but involves complex political and strategic obstacles. As a result, there was no formal pooling of reserves, common exchange controls, or common policy-making involved in the sterling bloc. Indeed, it was in this interwar period that many Commonwealth countries and India established their own central banks (with close personal and professional links to the Bank of England). It was also in this interwar period that many new colonial currency boards were established.

The sterling bloc clung together in the wake of successive shocks to the international economy in the interwar period. With the onset of the war in 1939, sterling convertibility was suspended, and monetary relations were determined by a series of emergency Defence Finance Regulations, which defined a set of "scheduled territories" within which sterling was used more freely. These regulations were destined to remain in force throughout the 1950s and shaped the postwar sterling monetary system.

The Sterling Area in the Era of Controls 1945–1958

The vision of wartime planners in Britain and America had been to create a deliberately, and to some extent collectively, managed international economy that would prevent the crises that had plagued the interwar years. As noted above, the institution designed at Bretton Woods in 1944 to achieve this – the International Monetary Fund – imposed pegged exchange rates and promoted convertibility of currencies to facilitate multilateral trade and payments. In fact, convertibility of the major currencies was not possible until the end of 1958, and even then only for current account transactions. Controls were maintained on capital flows (especially

short-term capital) throughout the 1950s and 1960s, by which time the US dollar was under pressure and the monetary system was beginning to crumble. Convertibility in a pegged exchange rate system as designed at Bretton Woods in 1944 can be said to have operated (if at all) for less than 10 years – starting in December 1958 with the onset of convertible currencies and ending in March 1968 when a two-tiered gold price was introduced. In the meantime other arrangements were devised.

At the end of the war, sterling made up about 80% of global foreign exchange reserves, and half the world's trade was denominated in sterling, so the functioning of the international monetary system depended on careful management of the pound. But the productive power of the British and European economies had been seriously undermined by 6 years of destructive war effort. As the peacetime reconstruction began, the global demand for American goods and investment far outstripped the dollar resources to buy them, so if European currencies were convertible to the dollar they would quickly exhaust the meager supply of dollars held in European central banks. This, indeed, was the lesson learned in the short-lived move to sterling convertibility in August of 1947, which had to be suspended after only 6 weeks before it bankrupted the Bank of England. Sterling convertibility in 1947 was forced on Britain by the US government as a condition of an American Loan, but rather than ushering in the nondiscriminatory payments system to which the US government aspired, this episode confirmed that any move to the multilateral payments envisioned at Bretton Woods in 1944 would be postponed indefinitely. While trade liberalization was embraced and trade quotas were removed during the 1950s, capital controls and exchange controls were intensified to protect national currencies from speculative pressure.

Consistent with the prevailing belief that international economic relations should be controlled and managed, the sterling area system was more clearly defined and more closely regulated than the interwar sterling bloc. The sterling area should also be seen as part of the general context of regional solutions to payments problems in the postwar period, alongside the European initiatives which resulted in the European Payments Union which facilitated multilateral payments in Europe from 1950 to 1958 and the EEC. Britain was an important (although sometimes reluctant) member of the European Payments Union, but they remained outside subsequent European systems like the Coal and Steel Community of 1952 and the Economic Community of 1957. After the debacle of 1947, countries came to favor less ambitious solutions to the payments problems which the IMF had failed to solve, rather than the global "one world" vision of the wartime planners.

The sterling area included Commonwealth countries (except for Canada) and all the formal British dependencies, plus a few other members left over from the sterling bloc such as Egypt, Iraq, Kuwait, Persian Gulf States, Burma, Jordan, and Iceland. Canada was excluded because of its close monetary and commercial links with the USA, while Ireland was included because of its close economic and monetary links with the UK. Members of the sterling area agreed to maintain fixed exchange rates with sterling, to hold the bulk of their foreign exchange reserves in sterling, and to impose exchange control in common with Britain to protect against possible flight from sterling to other currencies (in particular the $US). In return, members enjoyed

freer trade with Britain and freer access to British capital than other countries. Access to the London capital market had particular importance in the 1950s because other markets were heavily restricted by capital controls under the Bretton Woods system. This was an important advantage given the need for finance and investment among most of the constitutionally independent members of the sterling area such as Australia and New Zealand. On the other hand, if a country opted to leave the sterling area, it would step outside the boundaries of the exchange controls, and the sterling assets that it held would no longer be convertible to other currencies. In the late 1940s and early 1950s, a few members such as Iraq, Lebanon, and Egypt left the sterling area and had their assets "blocked." The system thus operated to support collective interests of its members in the stability of sterling and freer trade and investment flows, underpinned by carrots and sticks.

In the early years after 1945, the independent members of the sterling area held periodic meetings under the auspices of regular Commonwealth summits to coordinate trade policy and domestic macroeconomic policy in order to maintain their pegged exchange rates. Since all the members of the area held their foreign exchange reserves in sterling, this meant that they sold or pooled all their US$ and other currency earnings in London. In the first 6 years after the end of the war, there was a scarcity of US$ to buy essential capital and consumer goods, and so annual targets were set for imports from the dollar area (essentially the USA and Canada). However, as the dollar shortage receded so did the willingness of Commonwealth countries to coordinate, their trade programs evaporated, and the targets were abandoned by 1952 (Schenk 2013). Nevertheless, members continued to hold the bulk of their reserves in sterling to support their trade with the UK and each other and to protect their exchange rate peg to the pound. Since the UK usually ran a trade surplus with other members of the sterling area, who in turn usually ran a surplus with the rest of the world, this meant that Britain had access to the foreign currency earnings of the Empire and Commonwealth through pooled foreign reserves in London. The advantages of this specialization were felt especially in the raw materials boom and bust associated with the Korean War of 1950–1952 when the massive surpluses of the overseas sterling area offset the substantial deficit run by the UK. The balance of payments of the sterling area as a whole was relatively stable as a result.

Britain's war debts defined the early years of the sterling area. These debts were popularly known as the "sterling balances" and were owed primarily to colonies in South Asia, which were due for independence, including India, Pakistan, and Ceylon. Table 1 shows the distribution and amount of sterling balances. At the end of the war, these debts amounted to four times British foreign exchange reserves so they could not immediately be redeemed in dollars or other currency. Instead, the creditors were forced to continue to hold their assets in sterling or to spend them on imports from the UK or other sterling area countries who would accept sterling. Britain engaged in a long and complex set of negotiations with its primary creditors (especially India) in the late 1940s and early 1950s. As de Paiva Abreu (2017) has described, India and other territories negotiated staged releases of their sterling assets in the late 1940s to buy capital goods and other imports, gold purchases, and

Table 1 Distribution of gross official sterling debt December 1945

	Amount £ million	Percent of total
India/Pakistan	1,300	37
Other colonies	515	15
Egypt/Sudan	400	11
Ireland	190	5
Palestine/trans-Jordan	120	3
Australia	110	3
Argentina[a]	105	3
Total	2,740	77
Other	782	23
Total	3,522	100
Sterling area total	2950	84

Source: Schenk (2010)
[a]Argentina was not a formal member of the sterling area

pension-related expenditure. The value of their assets was also reduced by the 30% devaluation of sterling in 1949 and by subsequent inflation.

The impact of sterling area membership obviously differed between the colonies of Britain and the independent states of the Commonwealth. The colonies had no choice in their status because they had currency boards that enforced a direct link between the issue of local currency and sterling. The currency boards that operated in British colonies generally followed a common pattern. The issue of currency was often delegated to a currency board operated through the expatriate British commercial banks in the colonies. In the cases of West and East Africa, the currency board was physically located in London, but others, such as Hong Kong and Malaya/Singapore, were located in their home territory. The currency board supplied local currency on demand in return for sterling at a fixed exchange rate and vice versa. The automatic nature of the currency supply meant that the currency board had no discretionary power over the rate of growth of currency in circulation. Since the main economic activity in the colonies was foreign trade, the money supply moved directly with the balance of payments. In order to ensure the convertibility of the local currency to sterling at a fixed exchange rate, local currency was backed 100–110% by sterling assets. The boards thus provided a relatively cheap system of administering a local currency and gave that currency stability in terms of sterling. The two main criticisms of the system were that its automatic nature made the colonial economies vulnerable to fluctuations in the balance of payments and that the 100% reserve policy deprived the colonies of much needed development capital (Narsey 2016).

The independent members of the Commonwealth (such as Australia, South Africa, and New Zealand) all had central banks that operated the exchange control and ensured the fixed exchange rate with sterling was maintained. For the independent countries, the rationale was that the bulk of their trade was with the UK so it made sense to maintain stability in their exchange rates and to avoid exchange risk by keeping their reserves in sterling. As the competitiveness of British products

and the size of the British market waned compared to the booming continental European economies in the late 1950s and early 1960s, this commercial rationale for the monetary link became more tenuous. Another rationale was preferred access to the London capital market – the Commonwealth countries had ambitious plans for accelerating their industrialization, which relied on foreign investment. A fixed exchange rate with sterling was also considered to improve the "credit rating" of members in the eyes of private foreign investors since it constrained imprudent monetary policy in economies dependent on global primary product markets and shallow domestic foreign exchange markets. Finally, it is important to recognize that there were not many viable alternatives to sterling for these countries. This was a period when fixed exchange rates were the norm and fluctuations were frowned upon by the IMF and the international community. The only real alternative to pegging to sterling was to peg to the $US. In the 1950s, members did not have enough $US to build up sufficient reserves to establish such a peg, and by the mid-1960s, the $US was not such an attractive currency to use as an anchor – as the American ran persistent balance of payments deficits and the fixed link between the $US and gold was threatened.

The advantages and disadvantages for Britain of being the hub of the sterling area monetary system have been hotly debated. Contemporary scholars asserted that the sterling area was a drain on Britain's resources; that it represented an expensive effort to cling on to the glories of Britain's nineteenth century imperial role (e.g., Strange 1971; Shonfield 1958). In this scenario, the sterling area as a monetary system was seen as central to the British government's efforts to sustain the international use of sterling when it would have been wiser (with the benefit of hindsight) to have abandoned this role. There are three major burdens that the sterling area is supposed to have placed on the weakened British economy in the postwar period; sterling balances, excessive foreign investment as the condition which the overseas members of the system demanded in return for holding sterling and pegging their exchange rates, and that the sterling area system inhibited British economic policy-making. The debate on the implications of the sterling area should be viewed in the context of contemporary critiques of the relatively poor economic performance of Britain after the war, especially compared to its European neighbors.

First of all, to address the sterling balances. One of the rules of the sterling area was that members agreed to keep their foreign exchange reserves denominated in sterling, pooling their other foreign exchange earnings in London by converting them to sterling at the Bank of England. The result was sizeable reserves held in the form of short-term British government securities by members of the sterling area – the so-called sterling balances. These represented potential claims since they could be converted in London for the currency needed by the holder. Since one of the rationales of the sterling area was to conserve scarce foreign exchange, outstanding sterling balances well outweighed the value of the central reserves; the ratio of sterling liabilities to British foreign exchange reserves never fell below 2:1 during the 1950s. This "overhang" was deemed to make the British external economic position very fragile. In contemporary and subsequent accounts, the sterling balances appeared as a sword of Damocles hanging over the head of the British

economy, always ready to bankrupt the Bank of England. Critics argued that this inhibited British domestic economic policy by requiring higher interest rates to ensure that sterling was kept attractive to foreign holders and that this also ruled out devaluation, which might have helped British export competitiveness.

At the time, and since, the sterling balances were identified with extraordinary short-term debts which the UK accumulated to members of the Empire and Commonwealth during the war. In fact, after 1952 they were the "normal" foreign exchange reserves of these countries and were unlikely to be "cashed in" all at once except in times of crisis when emergency measures could be taken to protect the central reserves, such as drawing on the UKs IMF quotas. One of the major rationales for a currency area is that it conserves on the amount of foreign exchange which must lie idle – the sterling balances could (optimistically) be viewed therefore merely as an indication that the system was working as it should, especially as the nominal value of total sterling balances was remarkably stable throughout the decade. Even through the process of the end of Empire as many colonies became independent through the 1960s, they did not seek to run down all their foreign exchange reserves or to remove themselves from the sterling area because they benefited from the confidence and stability of the link with sterling. Two of the largest newly independent countries – Malaysia and Ghana – went so far as to continue to operate currency boards for that reason (Schenk 1997).

By the 1960s, Fig. 1 shows that there had been a major geographical redistribution of overseas sterling holdings, particularly in favor of the Far East (Singapore and Hong Kong) and the Middle East (primarily Kuwait), while holdings of Australia and New Zealand were more stable.

The increase in sterling held by territories in the Far East reflected the acceleration of industrialization, particularly in Hong Kong and Singapore, and the growth of the money supply and government reserves that accompanied increased prosperity in

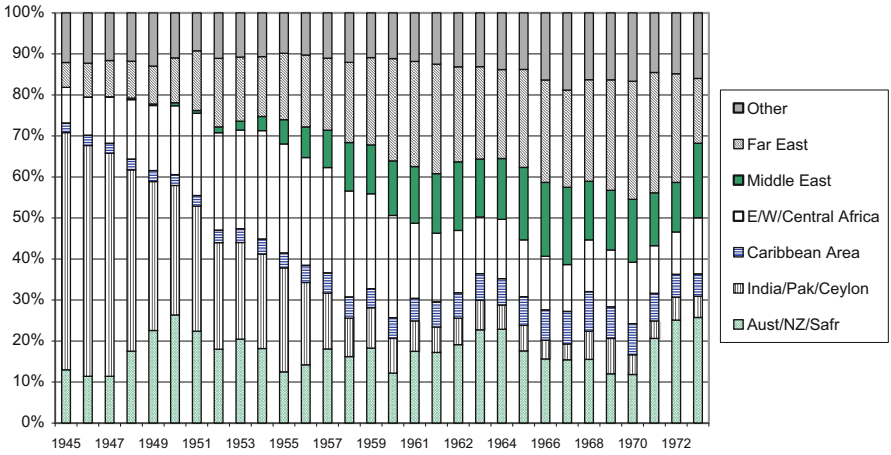


Fig. 1 Distribution of OSA sterling balances 1945–1973. (Source: Schenk 2010)

these former and continuing colonies. The former colonies kept their currencies tightly linked with sterling and pledged on independence at the end of the 1950s to maintain 100% sterling reserves (Schenk 1997). The rise in sterling balances therefore reflected the legacy of the colonial monetary system as well as the reality that the US$ was also unstable and gold was in short supply. The rise of Middle Eastern sterling balances reflected the international agreements to denominate a proportion of their oil trade in sterling. As their mineral resources were exploited and their reserves increased, so too did their sterling balances. These developments all show that the pattern of economic development on the geographical edges of Britain's former formal and informal Empire profoundly affected the distribution of sterling assets.

The second burden that the system was supposed to have placed on Britain was that of requiring excessive foreign investment to keep the members loyal. The complaint was that the sterling area prevented the imposition of capital controls as an instrument to protect domestic industry. This was supposed to have drained the British economy of scarce capital resources, leaving its industry weakened. Long-term investment in the sterling area amounted to about £200m p.a. in the 1950s which was never more than 1.5% of GDP or about 8% of domestic savings (Schenk 1994). Incremental Capital Output Ratios suggest that the problem with British industry was that the productivity of investment was low relative to other countries. Imposing capital controls was unlikely to make any significant impact on growth rates at home and indeed might have been detrimental in terms of inhibiting exports.

The final criticism of the sterling area is more fundamental to any consideration of monetary union – this is the loss of policy sovereignty. It has been argued that because Britain was the hub of the sterling area, its policy choices were constrained, and the interests of the area as a whole were put ahead of the interests of Britain alone. This argument is most commonly expressed with reference to the exchange rate. The argument goes that if Britain had been "free" to devalue sterling or move to a more flexible regime and that this would have enhanced the competitiveness of British production and increased growth in the postwar period (Burnham 2003). Thus, the adherence to the sterling area monetary system and the large value of sterling assets held outside the UK inhibited timely devaluation. Looking more closely, these arguments underestimate the influence of domestic policy commitments as a main constraint on exchange rate policy. The political consensus established during the war ensured that full employment and price stability were the primary goals of the postwar governments, and they believed that devaluation would generate inflation and also, through raising the price of imports and the cost of production, would raise unemployment levels. Furthermore there is no overwhelming evidence that devaluation would have improved Britain's balance of payments in the medium term. British exports were uncompetitive in terms of quality and after-sales service rather than price alone.

Some have blamed the sterling area for Britain's misguided hesitation over joining in European integration at the early stages. In this argument, the sterling area is considered to have been a credible alternative to European integration for

Britain. This, also, is a fallacy. Despite some imperial aspirations to the contrary, it was never likely that the sterling area would persist as a viable system in perpetuity or that it genuinely offered a viable alternative to economic relations with continental Europe. Policy-makers at the time discussed and rejected intensifying the link with the sterling area as a bulwark against the rest of the world. The truth was that by the early 1950s the other members of the sterling area were unwilling to reinforce their ties to the relatively slow-growing British economy. Retreating into a tight sterling system was not a viable option either economically or politically. Political as well as economic relations with the USA, Europe, and the rest of the Commonwealth depended on the continued commitment to achieving multilateral trade and payments at some time in the future. When Britain began to apply for membership of the EEC in the early 1960s, the sterling area was expected to fit in with Britain's plans, not be a fallback alternative. While many countries in the sterling area, such as Australia and New Zealand raised serious objections because of their close trade relations with the UK, as Ashton (2007) argued "although its essential context was always apparent, the Commonwealth was not a major impediment in Britain's negotiations over Europe."

Another group of literature has focused on the impact of the sterling area on members other than the UK. One important area of debate is the extent to which sterling area relationships were important in the process of decolonization. Thus Hinds (2001) and Krozewski (2001) argued that the financial links between the colonies and the sterling area and the risk to sterling arising from constitutional independence were important factors in British attitudes to unwinding the Empire. For Hinds, the financial burden of the colonies encouraged decolonization, while for Krozewski, the potential foreign exchange earnings of the colonies were an obstacle to decolonization. This argument was challenged by Schenk (2010) who noted that the colonies already had considerable independence in how they used their sterling assets and the ability to diversify parts of their colonial monetary backing from the early 1950s. Moreover, after independence most territories retained their sterling area membership and obligations in order to ensure monetary stability and access to international capital markets. More recently, the debate has returned, with specific attention to individual cases. Sutton (2015, 2016) argued that the dollar earning capacity of Malaya from rubber and tin exports was a vital imperative for the British in their negotiations with Malayan officials in the 1950s. The British sought "at all costs" to keep Malaya within the sterling area, although the Malayans recognized that remaining in the sterling area was in their best interests while they ran a substantial trade deficit with the UK. The result was a substantial commitment of British development and military aid that accompanied independence in August 1957 and a continuation of the economic relationship between Malaysia and Britain. White (2017) takes a benign view of these arrangements and compares the "open approach" of the decolonization settlement for Malaysia and Singapore to that of the Netherlands and Indonesia to explain the more favorable economic performance of Malaysia and Singapore in the postcolonial era. Nayamunda (2017) examines the case of Southern

Rhodesia (colonial Zimbabwe) and the Federation of Rhodesia and Nyasaland from 1953 and finds that here, also, monetary independence was a crucial factor in the drive to constitutional independence during the 1950s. Conversely, Maekawa (2014) finds little evidence for neocolonialism in Britain's relations with East African countries in the 1960s and 1970s so that "British interests hardly modified the nature of the development policies pursued by the newly independent states of East Africa." Further research on the role of individual members of the sterling area promises new insights on the development of the global system.

In December 1958, Britain formally allowed sterling to be convertible for current account transactions (i.e., mainly international trade), but this privilege was restricted to residents outside the sterling area. This step was taken in concert with the convertibility of other West European currencies and marked the (much delayed) adherence to the multilateral trade and payments envisioned at Bretton Woods in 1944. British residents and residents of sterling area countries were still subject to strict exchange control imposed by their respective monetary authorities. With the end of the extreme dollar shortage by the mid-1950s, the need for closely coordinated macroeconomic policies was not so urgent. At the same time, members of the sterling area developed more ambitious industrialization policies in an effort to diversify their economies. As a result, the complementarity of the economies which had been instrumental in cushioning the system through the primary product booms and busts of the early postwar years receded. The complementarity of the economies in the system was a strength in the early 1950s when the surpluses of primary product producers could offset the deficits of manufacturers and vice versa. By the 1960s, however, this complementarity was becoming a disadvantage since the aspirations and therefore the macroeconomic priorities of the members diverged. The UK still strove for price stability, while the developing countries strove for expansion, even at the expense of inflation in the short term. Also by this time, a persistent imbalance had arisen with the overseas sterling area being consistent net creditors to the central reserves and the UK becoming a net drawer.

In summary, by 1959, the sterling area was a very different entity than it had been a decade earlier. Within the sterling area, the war debts to South Asia no longer dominated. Through the decade, European central banks concentrated their reserves in US$ and gold so that sterling's role as a reserve currency retreated to the sterling area countries. At the same time, the trade and reserves pooling elements of the sterling area eroded as members sought to advance their own national economic interests without reference to common sterling area interests. On the other hand, access to the London capital market and the ability of their central banks to convert sterling more or less on demand continued to be important to sterling area countries. Despite the constitutional upheavals of decolonization, the sterling area provided a structure that left the monetary links between newly independent states and the sterling area largely unchanged. In the 1960s, however, the significance of the sterling area began to fade as the wider international monetary system began to crumble. Criticism of sterling's exchange rate shifted from the sterling area to Britain's own economic management.

The Sterling Area and International Monetary Cooperation 1959–1967

Once European currencies moved to convertibility on current account at the end of 1958, the pegged exchange rate system began to falter and required a series of institutional supports from the leading industrialized economies. In the context of the Cold War, the international monetary system was believed to be fundamental for the stability of the capitalist economy. This political commitment to prolonging the pegged exchange rate regime throughout the 1960s drew European and North American countries together in a range of cooperative innovations. During this period there were several strains in the system: sterling and the French franc were under repeated pressure to depreciate in exchange markets, the gold value of the US dollar was often viewed as too high, and the German Deutsche Mark peg was sometimes considered too low. The result was periodic corrections in exchange rates, often after sustained market pressure, resulting in abrupt and disruptive changes. Sterling attracted special attention in this context because of its importance in the international monetary system; about a third of global foreign exchange reserves were still held in sterling in the 1960s. Devaluation would damage the international monetary system as a whole and also undermine confidence in the US$ as the main international currency. If the value of sterling dropped, speculators were expected to turn their attention more forcefully against the gold value of the dollar. Both the dollar and sterling were particularly vulnerable because of the large outstanding assets held by other countries. Thus the central banks of the major industrialized countries worked together to support the dollar through swaps and by intervening to hold the market price of gold at the official price of $35/oz from 1961, until this effort became unsustainable in March 1968. In the end, the prediction that a sterling devaluation would force a devaluation of the dollar proved true.

The arrangement of systems of bilateral support for the international monetary system was developed during the exchange crises of 1960–1961 when pressure on the US$ and sterling led to concerted efforts by US and European central banks to support existing exchange rates (Toniolo 2005). In March and June 1961, the UK garnered bilateral support totalling $US904 million from European and US central banks. As Toniolo relates, the Bank of England and the US Fed were enthusiastic about the ability of the combined forces of central banks to forestall short-term speculative pressure and sought to extend or formalize the arrangements. The creditors in Europe were less enthusiastic about extending the system, but the initiative was nevertheless formalized through the Bank for International Settlements. The Bank for International Settlements had been established in the 1930s and gathered the central bankers of the major industrial countries to discuss issues in the international monetary and banking system (Toniolo 2005).

Table 2 shows the support that the Bank of England negotiated with its partner central banks at the Bank for International Settlements (BIS). The USA was always the largest contributor, as the richest country and with the most to lose from a collapse in the pound. The support took two main forms. The Bilateral Concerté were lines of short-term credit offered to the Bank of England in bilateral agreements by central

Table 2 Multilateral support for sterling 1963–1968

Format	Date of agreement	Total amount of credit available
Bilateral Concerté	March 1963	$250m
Bilateral Concerté	September 1964	$1,000m
Bilateral Concerté	November 1964	$2,530m
Group Arrangement	June 1966	$910m
BIS multilateral support	October 1967	$250m
Group Arrangement	September 1968	$2,000m

Source: Schenk (2010)

banks including West Germany, Switzerland, Italy, Belgium, the Netherlands, Sweden, and the USA. The credits were bilateral, but the BIS would inform each creditor of the details of the total operation. They were repayable within 3 months and were to be used to cover short-term speculative attacks on sterling. After the November 1964 Bilateral Concerté, discussion turned to longer-term support, and a new structure was created, known as Group Arrangements, which offered collective lines of credit for 12 months, renewable on agreement of the participants. The Group Arrangements had broader participation, including the Bilateral Concerté countries plus Canada, Japan, Austria, Denmark, and Norway. The First Group Arrangement took several months to negotiate and was developed to respond to anticipated pressure on the Bank of England's foreign exchange reserves that might arise from a decision by sterling area countries to diversify their reserves away from the pound. The terms, therefore, anticipated the breakup of the sterling area and were designed to help the British reserves cope with this event. The credit was explicitly not to be used for Britain's own balance of payments problems, for with the British government was supposed to take its own corrective action. The June 1966 First Group Arrangement was hastily renewed in March 1967 and then again, with more discussion, in February 1968. On this second occasion, the contributing countries insisted that new, more permanent, arrangements would need to be negotiated in the aftermath of the devaluation of sterling in November 1967. In October 1967, the BIS had acted as principle in offering an extra $250m in emergency credit in advance of the devaluation, to be collected afterward through the member central banks.

These multilateral discussions by the Bank of England in the early 1960s presupposed that the sterling area's demise was inevitable because of the deteriorating attractions of sterling as a store of value, but political understanding sometimes lagged behind the economic realities. Within the UK Treasury and the Bank of England, there was a clear desire to extricate the British economy and balance of payments from the vulnerability of sterling as an international currency. Indeed new controls were imposed on third-party use of sterling in international trade, and restrictions were placed on the free movement of capital to major sterling area members such as Australia, which undermined the very structure of the sterling area by violating its "rules." At the end of 1962, Chancellor of the Exchequer Reginald Maudling noted privately that "I regard it as a major aim of policy to free the UK economy from the inhibitions of reserve currency status," and JS Fforde

of the Bank of England remarked in 1964, "I do not think that at the highest levels in the Bank there would be dissent from the proposition that to get rid of reserve-currency status while maintaining our trading currency position would be a most desirable achievement" (Schenk 2010). However, Ashton (2007) describes how Harold Wilson, first in opposition and later as a fresh incumbent Prime Minister in 1965, hoped to reinforce Commonwealth economic ties. But he was quickly disabused of this idea by Treasury officials. They explained that trade with the Commonwealth had peaked over a decade earlier and was now in prolonged decline, that countries were concerned primarily with their own development and access to capital outside the UK, and that Britain would do better to develop trade relations with other countries. The international role of sterling relied on the continuation of the sterling area, but the British commitment to this international role was waning.

Whether the exchange rate problems of the 1960s were mainly internal to the political and economic weaknesses of Britain itself or to the structure of the international monetary system as a whole is debated in the literature, as is whether the 1967 devaluation should have happened sooner (Newton 2010; Oliver 2011). Certainly, the sterling area played a part in the pressure on sterling, since these countries were the largest holders of sterling at the time. The greatest decline in sterling reserves happened in the Middle East due to political instability in the region in the summer of 1967. Countries outside the sterling area also reduced their sterling holdings when swap agreements with the Bank of England unwound. Meanwhile, private holders also lost faith in the Labour Government's ability to resist devaluation. On 18 November 1967, the line could no longer be held, and the Chancellor of the Exchequer took the decision to devalue sterling by 14.3%. While the US administration, European ministers, and the IMF were very hastily consulted, no members of the sterling area were told in advance that the value of their reserves would overnight be reduced by such a significant amount. The bitter sense of betrayal felt by loyal sterling area members marked an important turning point for sterling area relations.

The Sterling Area and the End of Bretton Woods 1968–1972

During the 5 years following the devaluation of sterling in 1967, the sterling area was dramatically transformed, to the point where its formal dismantling in 1972 was merely a footnote to the abandonment of sterling's pegged exchange rate regime. The most important innovation was the Sterling Agreements starting in 1968 that controlled the diversification away from sterling and provided a dollar value guarantee for most of the sterling held by sterling area monetary authorities. By the end of the 1960s, the sterling area had become a set of bilaterally negotiated, public, and managed statutory agreements.

In the wake of the devaluation of November 1967, members of the sterling area began to accelerate the diversification of their reserves, but some had begun earlier. Singapore secretly stopped accumulating sterling in the months before the devaluation and revealed in November that their sterling reserves amounted to only

50% of total compared to 77% in June 1967. In the mid-1960s, Australia and New Zealand had also struggled with their allegiance to sterling. They were torn between wanting to sell off their sterling assets, which they expected to fall in value, and the need to avoid prompting a collapse of sterling if it were known that they were selling out of sterling (Singleton and Schenk 2015). The benefits of sterling area membership had eroded as other sources of investment capital and aid were available in the USA, Europe, and multilateral agencies such as the World Bank. In any case, the sterling area was responsible for only a part of the global pressure on sterling in the summer and autumn of 1967 (Newton 2010). Losing over 14% of the dollar value of their reserves in 1967 was a final straw for many members of the sterling area, but a rapid diversification risked the value of remaining sterling reserves. This time the international community came to the rescue with a plan to manage the final retreat from the sterling area.

The Group Arrangement and other multilateral credits arranged through the Bank for International Settlements and the International Monetary Fund proved insufficient to avoid a devaluation, but they were important to the aftermath. At the Bank for International Settlements, the Bank of England pressed its partners to offer yet another Group Arrangement to protect the international monetary system from a sterling collapse. In March 1968, the fall in the private market price of gold and the suspension of the gold convertibility of the dollar except for central banks added further incentive to create a lasting solution for sterling. This time other central banks required more assurances from the British authorities, including an offer to guarantee the dollar value of 90% of the sterling held in the reserves of sterling area countries. This would remove the main incentive for the sterling area to diversify away from sterling: an anticipated further fall in the pound's value. In return, the British authorities required each sterling area country to promise formally to keep a minimum proportion of their reserves in sterling. What followed was a dramatic reversal in fortune as Bank of England and Treasury officials travelled to 34 countries and territories (many former or current colonies) through the summer of 1968 to negotiate formal bilateral agreements to limit diversification away from sterling. Both sides were aware that without these agreements in place by the deadline in September, the final Group Arrangement would not be forthcoming, so the sterling area countries were in a strong bargaining position. As students rioted against war and oppression across Europe and the USA, British bureaucrats finalized the effective end of the sterling area monetary system.

Table 3 shows the minimum sterling proportions agreed with 34 members of the sterling area by September 1968, which allowed the Second Group Arrangement of $2,000m to be offered through the Bank for International Settlements. The deal was that if sterling fell below these thresholds, the country would lose the dollar value guarantee. On the other hand, if sterling fell more than 1% below the current parity to the US dollar ($2.40), then members of the sterling area were entitled to compensation. The largest single sterling holder by this time was the colony of Hong Kong, which agreed swiftly to continue to hold almost all of its reserves in sterling, primarily as backing for their currency issue. Australia, Malaysia, and Singapore all set their threshold at 40% of reserves, which was close to where they had reached

Table 3 1968 Sterling Agreements: minimum sterling proportions

East Caribbean Currency Authority	100
Gambia	100
Hong Kong	99
Barbados	97
Mauritius	95
British Honduras	90
Bahamas	80
Bermuda	80
Ceylon	80
Ghana	80
Guyana	80
Malawi	80
Trinidad	80
Malta	75
Bahrain	70
New Zealand	70
Sierra Leone	70
Zambia	65
Nigeria	60
Jamaica	57
Ireland	55
Uganda	51
Cyprus	50
Dubai	50
Iceland	45
Australia	40
Malaysia	40
Pakistan	40
Singapore	40
Jordan	25
Tanzania	25
Kuwait	25
Libya	18
India	13

Source: Schenk (2010)

by the time the negotiations were concluded. At this point the New Zealand economy was much more tightly linked to British trade and investment and agreed to keep 70% of its reserves in sterling. India, the largest sterling holder in 1945, had already spent its sterling reserves by 1968 and emerged with the lowest minimum threshold. The Sterling Agreements were sealed by exchange of intergovernmental letters, most of which were then published by the British government. In order to monitor the Agreements, each country reported the denomination of their reserves on a monthly basis to the Bank of England.

The Agreements were considered a success at the time since there was no defection from sterling among the participants. On the other hand, the main alternative reserve asset, the US dollar, was under considerable pressure and was also expected to fall in value, so the benefits of diversification were dulled. Moreover, returns on sterling were relatively high, and of course most sterling reserves were covered by the value guarantee. Rather than falling, the value of sterling held in the reserves of sterling area countries increased by about £2,000m from the time of signing the Sterling Agreements and when they were due for renewal in September 1971. The share of sterling in reserves also rose sharply in the 1st year but then fell back to about 54% by 1971, which was about where it had started in 1968. The British government agreed to renew the Sterling Agreements in September 1971, reducing the minimum sterling proportions by 10% across the board to allow an orderly diversification and to discourage further accumulation of sterling reserves.

The renewal deals were concluded in the summer of 1971 before the Nixon shock on August, which severed the link between the dollar and gold and provoked a devaluation of the dollar to be negotiated at the Smithsonian Conference in December 1971. The new exchange rate for the pound was increased from $2.40 to $2.60, but the trigger rate for the Bank of England to pay compensation was not changed. This reduced the value of the Sterling Agreements for the participants and was bitterly disputed, but the British held firm. In the event, the new exchange rate soon came under pressure, and in June 1972 sterling was allowed to float free of its peg, and compensation of £58m became due. The Sterling Agreements were renewed again in September 1973 for a further 6 months, and the trigger point for compensation was adjusted this time. This cost the Bank of England a further £100m in compensation as sterling slumped against the dollar through the economic turmoil of the OPEC oil crisis. In March 1974 the final set of Sterling Agreements were concluded to take the system through to the end of that year. No further compensation was paid, and the minimum sterling proportions were gradually reduced. The persistence of these agreements even after the formal end of the Sterling Area and the collapse of the Bretton Woods pegged exchange rate system demonstrates the continued need to manage even the final vestiges of sterling's international role.

After the Smithsonian Agreement of December 1971 increased the margins of flexibility within the pegged exchange rate system, some sterling area countries abandoned sterling and switched their peg to the US dollar, including India and Pakistan, Kenya, Nigeria, and Ghana. The evolution of trade and investment patterns meant that maintaining the peg to sterling and enduring wider fluctuations against the dollar was less attractive than pegging directly to the dollar. Six months later, the sterling area was finally abandoned quietly as a footnote to the floating of the pound in June 1972. The list of "scheduled territories" under exchange control legislation was reduced to the British Isles and Ireland. By 1971, sterling accounted for less than 10% of global reserves, but still the Sterling Agreements were extended to 1974, and a final "safety net" line of credit was offered by Britain's partners in the Bank for International Settlements in 1977.

The history of the sterling area in the 1960s demonstrates the risks and disadvantages from issuing a national currency that is used as an international reserve

asset and commercial currency. The accumulation of assets by central banks and other holders outside Britain were almost always viewed as a source of weakness for the British economy rather than an expression of confidence in Britain's economic power. Contemporary observers could clearly identify the disconnect between the waning economic competitiveness and imperial power of the British economy on the one hand and the continued international role for the British pound on the other. However, even when many in the British political and public spheres would have welcomed the quiet retirement of sterling, the fragility of the international monetary system built on a weakened dollar made it difficult to end sterling's international status. The risk that a sudden tipping point would disrupt the global payments system meant that Britain was able to attract considerable multilateral support by the end of the decade but only by offering to guarantee the dollar value of sterling reserves.

Conclusions

The sterling area international monetary system was in place for over 20 years and at its start involved close to half of the world's trade, finishing with 34 members from every edge of the globe. Arising out of wartime exchange controls and British war debts, it was instrumental in restoring multilateral trade and payments among a large and widespread group of developed and developing countries in the postwar period. The failure of the Bretton Woods institutions to manage the process of recovery left the members of the sterling area to rely on the wartime controls as a platform to allow the freeing up of payments, trade, and capital flows. But the system also adapted. In the 1950s the war debts were retired and replaced by accumulations of reserves in developing regions. At the same time, the defensive impetus for cooperation relaxed with the end of the dollar shortage, so that the prospects for any tighter relationship receded and the autonomous members of the sterling area sought to diversify their international economic contacts away from the traditional focus on the British economy. By the 1960s, the sterling area was caught up in the arrangements to prop up the Bretton Woods pegged exchange rate system. The continuation of sterling as an important reserve asset for sterling area countries meant that reforms to the Bretton Woods system required solutions for sterling as well as the dollar as international currencies. The vulnerability of the dollar to a collapse of the pound meant that the British government and Bank of England were able to attract wider support to manage the sterling area. The devaluation in 1967 marked a major turning point as sterling area members accelerated their turn away from Britain, and the monetary relations were formalized through the Sterling Agreements. The sterling area was the last stage in the erosion of Britain's domination of international payments: a process that can be traced from the heyday of the gold standard, through the cruel volatility of the interwar period, to the shattered hopes for a managed utopia in the postwar period. After 1972, London sustained its role as a global financial center, but this was based not on the importance of its own currency but rather through the proliferation of Eurodollar business.

Cross-References

- ▶ Currency Boards
- ▶ International Currencies in the Lens of History
- ▶ International Monetary Regimes: The Bretton Woods System

References

Ashton SR (2007) British government perspectives on the commonwealth, 1964–71: An asset or a liability? Journal of Imperial and Commonwealth History 35(1):73–94
Burnham P (2003) Remaking the postwar world economy: robot and British policy in the 1950s. Palgrave Macmillan, London
De Paiva Abreu M (2017) Britain as a debtor: Indian sterling balances 1940–53. Econ Hist Rev 70 (2):586–604
Drummond I (1981) The floating pound and the sterling area, 1931–1939. Cambridge University Press, Cambridge
Eichengreen B, Flandreau M (2009) The rise and fall of the dollar (or when did the dollar replace Sterling as the leading reserve currency?). Eur Rev Econ Hist 13(3):377–411
Hinds A (2001) Britain's Sterling colonial policy and decolonization, 1939–1958. Greenwood Press, Westport
Krozewski G (2001) Money and the end of empire: British international economic policy and the colonies 1947–58. Springer, Cambridge
Maekawa I (2014) Neo-Colonialism reconsidered: a case study of East Africa in the 1960s and 1970s. J Imp Commonw Hist 43(2):317–341
Narsey W (2016) British imperialism and the making of colonial currency systems. Palgrave Macmillan, Basingstoke
Nayamunda T (2017) British Sterling imperialism, settler colonialism and the political economy of money and finance in southern Rhodesia, 1945–1962. Afr Econ Hist 45(1):77–109
Newton S (2010) The Sterling devaluation of 1967, the international economy and post-war social democracy. Engl Hist Rev cxxv:912–945
Oliver MJ (2011) The management of Sterling 1964–1967. Engl Hist Rev CXXVI(520):582–613
Schenk CR (1997) Monetary institutions in newly independent countries: the experience of Malaya, Ghana and Nigeria in the 1950s. Financ Hist Rev 4:181–198
Schenk C (2010) The decline of Sterling: managing the retreat of an international currency. Cambridge University Press, Cambridge
Schenk CR (2013) The Sterling area and economic disintegration. Gesch Ges 39(2):177–196
Shonfield A (1958) British economic policy since the war. Penguin, Harmondsworth
Singleton J, Schenk CR (2015) The shift from sterling to the dollar 1965–76: evidence from Australia and New Zealand. Aust Econ Hist Rev 68(4):1154–1176
Strange S (1971) Sterling and British policy: a political study of an international currency in decline. Oxford University Press, Oxford
Sutton A (2015) The political economy of imperial relations: Britain, the sterling area, and Malaya 1945–1960. Palgrave Macmillan, Houndmills
Sutton A (2016) British imperialism and the political economy of Malayan independence. J Imp Commonw Hist 44(3):470–491
Toniolo G (2005) Central Bank cooperation at the BIS 1930–1973. Cambridge University Press, Cambridge, UK
White N (2017) The settlement of decolonization and post-colonial economic development: Indonesia, Malaysia and Singapore compared. Bijdragen tot de taal-, land- en volkenkunde/ J Humanit Soc Sci SE Asia 173(2–3):208–241

Currency Blocs: The Yen

31

Michael Schiltz

破綻しますよ、いつかは破綻しますよ、それは *(hatan shimasuyo; itsuka hatan shimasuyo, sore wa)*
"Of course [such scheme] is bound to implode; of course it will implode at some time."
Ohara Masahiroshi 小原正弘*, former Yokohama Specie Bank banker and once appointee to the Federal Reserve Bank of China (*中國聯合準備銀行*)*
in: S.a. En no sensō "The War of the [Continental] Yen"『圓の戦争』
"Japan and the yen ha[d] an overpowering dominance, and arrangements [were] made with an eye to the needs of Japan and the yen, rather than in consideration of the other areas and their currencies."
Hunsberger, "The Yen Bloc in Japan's Expansion Program," p. 251.

Contents

M. Schiltz (✉)
Department of International History, Graduate Institute, Geneva, Geneva, Switzerland
e-mail: michael.schiltz@gmail.com

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_30

Abstract

Historically, unilateral political attempts at boosting the liquidity and prestige of a currency have rarely been successful. In the context of catastrophic events as, for instance, a World War, several countries have nevertheless experimented with schemes to develop currency blocs. This chapter describes the Japanese experiment with establishing an economic zone in which the yen was either the main or the sole currency of reference. We proceed chronologically. Taiwan and Korea are explained as consolidating currency policy in the nascent Japanese empire. World War I temporarily reversed Japan's balance of payments problem. It heralded bold attempts at incorporating China within the Japanese *yen-bloc*, by means of aggressive lending schemes. Although the latter failed, the *yen-bloc* became an important political objective throughout the 1930s. The chapter explores the financial technology that governed the monetary set-up of the wartime empire, and it attempts to explain why it contained the seeds of its destruction.

Keywords

Imperialism · Gold-exchange standard · Blocism · Japan

Introduction

Is it possible to hack liquidity? Are there ways to boost the credibility of a currency? Put in the vocabulary of international rivalry, what are the chances of success in attempting to readjust the monetary pecking order to one's advantage and to challenge the position of a key currency?

If anything, the historical record would lead us to be rather skeptical. Monetary regime changes are conspicuously few and can hardly be said to have succeeded because of political initiative or imperialist intent. The United States' National Monetary Commission (1908–1912), among others set up to study the ways in which the international use of the dollar could be promoted, might appear as a successful exception but is probably best understood as a belated and corrective reaction to the steady rise of the dollar as reserve currency from the late nineteenth century onwards. The catastrophe of World War I merely consolidated a long-term shift in the monetary force field; on the surface of the broker's "course of the exchange," such evolutions remain only marginally visible, if at all. Only after World War I did it become obvious that this map had to be redrawn as well.

Other attempts at core currency status, not accidentally developed in reaction to a succinctly political aspiration, were less successful, if not to say outrightly disastrous. The Nazi's attempt at developing a *Großraumwirtschaft*, in which a group of inconvertible currencies was tied (pegged) to the German *Reichsmark*, is one such example. Japanese attempts at developing a yen-bloc are another. The Japanese project is, in a way, *sui generis* for it dated back to the early days of Japanese imperialism. Tracing its evolution is the mainstay of this chapter. We will witness that the contradiction between Japan's aspiring ambitions, on the one hand, and the

chronic lack of capital and credibility, on the other, led to a schism within the policy making community: pragmatists were in a constant struggle with hard-nosed ideologists. This is not to say that cleavages in the policy-making constituency developed along the lines of imperialist vs. anti-imperialist. As policy goals, expansion, and imperialism were not contested. Instead, the discussion concentrated on the means of securing real political and economic influence in East Asia, and on the political styles and tactics that would best secure a degree of autonomy and independence in a geopolitical environment that was, in fact, fundamentally adverse to both prospects.

An Early House-Call: Taiwan

> From now on, international conflict will not consist of invading countries and their peoples by means of military might [*wanryoku*], but of occupying [them] by means of financial power [*kinryoku*]. In other words, it is a fact that we will have to wage conflict not [according to the principles of] military tactics [*senryaku*], but [of] business strategy [*shōryaku*].
> —Gotō Shinpei

Perhaps surprisingly, yen diplomacy started long before the tumultuous 1930s and was closely bound up with nascent Japanese imperialism in the wake of the country's adoption of the gold standard (1897). Although several, if not most, observers have been led to stress the long-term beneficial effects of the monetary reform effort, short- and mid-term difficulties in the wake of this policy shift have received far less attention. Seen in hindsight, gold standard adoption was an audacious and profoundly risky bet at attaining superpower status. The outcome of this strategy was not only unforeseeable to its authors (the Meiji oligarchs) but also crucially dependent on the long-term commitment to an agenda that required very considerable sacrifices from the country's population and its institutions (Ōkurashō 1900). It also put into question Japan's commitment to the new imperial subjects of the freshly acquired overseas territories.

Nowhere was confusion so obvious as in the case of Taiwan. As Japan's first Asia colony, but especially in view of the state of its underdevelopment, it only seemed a natural starting point for the hands-on policies and the "steady sense of imperial purpose" for which Japanese expansionism became famous. There was not a single institutional remnant of earlier attempts at colonization (as there were in Korea); no rich pre-modern legacy that reformers felt to be at the core of their own identity (as in the case of the Chinese "cultural sphere"); and, especially, nothing of the types of infrastructure we commonly associate with colonial modernity (telegraphy, trams, and railway networks, and so on). Indeed, Taiwan was so far off the imperialist compass that General Charles LeGendre had suggested a Japanese expedition to the island, implying that it was simply for the taking (Thomson 1971).

Even so, a narrow yet important fissure developed in Japan's decision-making constituency from the very outset. The strategy of assimilationism for which the Taiwanese case has become known was not rigorously adhered to in the "hard" field of finance and fiscal matters. As Japan's reformers were soon to discover, the

society-wide imperialization of the colonial subjects was a long-term and, particularly, expensive undertaking. Famously, some Meiji leaders seemed to have pondered the sale of the island to France. Although this never happened, the Japanese faced several formidable tasks. Already in the immediate aftermath of Taiwan's annexation, there was a lot of soul-searching to be done about the validity of fully incorporating Taiwan's economy within the one of the mainland. In this sense, Taiwan was Japan's first encounter with the limits of expansionism and empire.

In the field of economic policy making, this manifested itself first in the context of the establishment of a Taiwanese central bank (factually one of the earliest central banks created, as Bytheway and Metzler have reminded us) (Bytheway and Metzler 2016, p. 19). First, it was difficult to find people willing to sit on a founding committee (*sōritsu iinkai*). When solicited, many of the newly formed Meiji elite politely demurred. Such reluctance was to a certain degree understandable: although in the immediate aftermath of the Sino-Japanese War investment in Taiwan increased, returns on these investments were far from certain and they were therefore considered risky by Japan's vested business interests. It is probably for this reason as well that, in the autumn of 1898, the industrialist Yasuda Zenjirō proposed to drop the concept of a bank of issue at all and, instead, to transform the Bank of Taiwan into a Sino-Japanese Bank (*nisshin ginkō*), with a mission of furthering trade. As a "financial institution [. . .] specializing in the trade with China and Korea [. . .], it should establish branch offices both in China and Korea" in order to "further Sino-Japanese trade relations."

Eventually, the central bank plan was saved by Gotō Shinpei, the visionary bureaucrat behind several of Japan's imperial schemes and the country's first money doctor (accidentally or not, he also was a medical doctor by training). The Bank of Taiwan (BOT) was central to Gotō's interest in the long-term development of the island and especially its integration with Japan's modernity while avoiding enforcement through military force. Cleverly playing to the expansionist forces in Tokyo, Gotō stressed the missionary importance of the BOT as a modern institution and a powerful means to gain influence for Japan farther south. His plea fell upon receptive ears. In June 1897, the BOT opened its doors in Taipei; its capital was five million yen, of which one-fourth was paid up. The foundations of Japan's southward economic advance had been laid.

Or had they? In the immediate aftermath of BOT establishment, and linked with the adoption of the gold standard in Japan proper, the question for Taiwan's monetary destiny once more uncovered schisms within Tokyo's policy-making constituency. Matsukata Masayoshi's stance had, from the very outset, been very sympathetic to Taiwan's assimilation, as we understand from his design of the island's institutions. Not only was Taiwan's financial and monetary administration subordinate to the Minister of State for Finance (Matsukata himself). The BOT was not granted the privilege of note-issue, as would have been the case for an independent central bank, but had only the right to issue bearer notes (Bank of Taiwan Law, Article 8).

Still, Ministry of Finance officials, understanding that the specie- and capital-poor Japanese government would have grave difficulties upholding the gold standard

in Japan proper, were forced to propose a hybrid solution to Taiwan's currency question. It was decided to leave core elements of the silver standard system intact, at least "for the time being" (Matsukata 1899, p. 383). This measure was certainly inspired by adoption of the gold standard in Japan proper. After all, an important element of the policy would be the liquidation of the country's accumulated silver reserves. At the time, fear of large amounts of exported silver yen coming back for redemption was one of the objections to the plan for bringing Japan onto gold (Soyeda 1899, p. 472). By offsetting the effects of the inflow of re-imported coins through putting silver to use in newly acquired territories, the Ministry of Finance probably thought it had found a key solution to alleviating at least a part of the pressure inherent in the monetary reform. For Taiwan, however, this meant that transaction costs *vis-à-vis* the Japanese mainland were bolstered, especially because the valuation of the Taiwanese "stamped" silver coin was based on the market rate of exchange in Hong Kong. Attempts at explaining the delayed integration of Taiwan into the Japanese economy soon made way for a plea to "cover Taiwan's expenditures by means of its receipts" (Namikata 1985, p. 70) – in effect, a recognition of the fact that Taiwan would have to be economically independent, more or less as a colony.

Eventually, the unexpected (and short-lived) upshot in the international silver price around 1902 would force yet another change of course (for a general discussion, see Kemmerer 1912). With the silver price becoming unpredictable, speculative exchange operations, mostly from Hong Kong and Amoy (Xiamen), but also from within Japan, were behind sharp shifts in Taiwan's money supply and cumulative losses at the BOT. In 1906, after years of administrative tug-of-war, the Japanese Government acknowledged that Taiwan was in practice incorporated within Japan's gold standard system. As for the BOT, it would become an institution in the forefront of Japanese imperialism in Southeast-Asia.

Korea: The First Monetary Laboratory

> Unlike other imperial states Japan has never been accused of having neglected her colonies; on the contrary, if any charge may be made, it is that Korea [. . .] suffered from excessive attention.
> —Hyman Kublin

Around the same time, i.e., more or less coinciding with the end of the Russo-Japanese War, another reformer was preparing the complete overhaul of Korean monetary affairs. This is not to say that this meddling on the peninsula was Japan's first. As a matter of fact, Tokyo had tried to gain influence for decades, more or less since the early 1870s. This had to do with the strategic position of Korea, which was quite in contrast with, say, Taiwan. Often dubbed the "dagger aimed at the heart of Japan," the Meiji reformers had decided that Korea was to function as a last line of defense of the homeland, both a geographic and political buffer against foreign interventionism. Monetarily, this position translated in (a) fierce attempts at, among

others, controlling the exchange rate of indigenous currencies, (b) (abortive) proposals for loans issued in order to gain financial leverage, and (c) an aggressive commercial presence of Japan's First Bank (*daiichi ginkō*).

Still, it would be Russo-Japanese rivalry that would have the greatest impact on the peninsula's economy. Throughout the 1890s, both parties implemented a series of monetary reforms that were cloaked in the discourse of assistance, yet clearly aimed at destroying each other's influence and legacy. Russia, from its side, used anti-Japanese opinion to quickly expand its grip on the political class and hasten the acquisition of concessions. Sergei Witte, the Russian finance minister who had become dissatisfied with similar efforts to obtain railroad concessions in China and other interests through the Russo-Chinese Bank (a heavily subsidized bank operating with French and Belgian money), found Korea a less contested target for his expansionist plans. The new Korean government was eager to help. Most forcefully, it resorted to a scheme of issuing large amounts of nickel currency (*paekt'ong*) of low denomination, partly in an attempt to drive Japanese currency out of the market.

But the Koreans, just as the Chinese, also attempted to play the imperialist powers against themselves. In 1898, they solicited Japanese authorities for a loan, with the express aim of bringing some order to Korea's perennially chaotic currency system. Shibusawa Eiichi's First Bank, too, would gladly undertake efforts at Korean monetary reform if they seemed profitable. This turned out to be the case. In 1900, Shibusawa therefore relentlessly pushed for negotiations with McLeavy Brown, an Englishman who had been in charge of the national finances since 1893 (his job being to see to it that Korea's foreign creditors were properly repaid). In 1901, it was agreed that that the First Bank issues banknotes, and that it lends them to the Korean government when it was in need of capital. Although the Ministry of Finance first objected on the grounds that note issuing was reserved to officially established banks, it later approved, with the stipulations that First Bank notes would circulate only under approval of the Korean government and only in Korea, hence not contradicting the legal provision that prohibited privately issued money on Japanese soil. First Bank notes went into circulation in 1902.

Unsurprisingly, none of the reforms were to much avail. If anything, they only added to the country's monetary complexity. Overcoinage, mostly of nickel coins (so called *paekt'ong*), and corollary inflationary pressures furthermore led to the massive destruction of homegrown wealth. What Megata Tanetarō, Japan special envoy to Korea, inherited in 1905 was thus an economy that was not only profoundly debased, yet also characterized by strong regional differentiation [illustration?]. The cure Korea was to be prescribed came right out of the money doctor's medical textbook (Drake 1993; Rosenberg 2003). Put in the dispassionate terminology of monetary advise, its core objectives were the thorough uniformization of Korea's money, financial practices, and, by extension, institutions. If we look more critically at its effects and its implementation, though, we are struck by the severity of its societal impact: the disenfranchising of certain Korean classes, especially the bourgeoisie or *yangban*; the confusion arising from neglect of existing financial practices

and relief measures; and especially the profoundly deflationary pressure on the Korean economy, with foreseeable consequences.

The most detrimental consequences of reform arose from the readjustment of Korea's monetary system, especially in the nickel coin regions. As a matter of fact, the rate Megata imposed for the latter to be redeemed was primarily set with the objective of boosting the relative share of already circulating Japanese currency in Korea. The societal impact of this decision must have been immediate. It was certainly advantageous to the *yen*-holding commercial classes. Vice versa, it was a direct blow to Korean commoners. This was not all. If reports of the time are to be a guide, it is clear that the process of redemption contained extraordinary bias in favor of members of, again, the (mostly Japanese and Chinese) merchant classes. They not only were the main holders of Grade A coins (lower grade coins were excluded from redemption), but they also benefited from the system set up to manage the process of redemption. During the first 3 months, the only applications accepted for currency transactions were for individual exchanges of between 1000 and 10,000 *wŏn* – amounts that were, for obvious reasons, out of the league of Korean commoners. This resulted in remarkable imbalances in exchange by nationality, with a large number of abstentions by Japanese nationals (131 individuals out of 178 total abstentions). Presumably, they filed their applications with the aim of making a quick profit, but then did not manage to buy the required number of coins by the exchange date.

Copper coins (*yŏpchŏn*), too, were to be redeemed. Here, the problem was not so much an artificially set exchange rate but rises in the international copper price. Large-scale exportation of what was the medium of exchange in the peninsula's (mostly remote) provinces resulted in a nagging deflation. For a lot of Koreans, reform thus came at a tremendous cost. Continuing shortages of cash depressed prices of farm products, increasing farmers' tax burdens, and once more forcing many into bankruptcy. Japanese measures at relief read as a cynical way of furthering imperial tactics. Seoul merchants, for instance, could pledge their real estate and receive loans on their merchandise, but the latter was strictly overseen, and loans were only granted in the new currency.

Although often overlooked in economic analyses, the profits accrued to Japanese institutions (more specifically, the First Bank, now turned into Korea's national bank, the *Bank of Chosen* (BOC)) were inversely proportional. Clearly, seigniorage occupied a large share of the latter. Remarkably, however, the BOT was much more than a "banker of banks." The *Economic History of Chosen*, published in commemoration of the bank's tenth anniversary, could rightfully claim that "so diverse and manifold are the services rendered by the bank that, since its establishment, there has been scarcely any reform undertaken in this country [. . .] which can absolutely disclaim the Bank's assistance given it in one form or another. It has done a great deal more than a central bank, as such is understood in most countries, ought to do" (Chosŏn Ŭnhaeng 1920). Conveniently left out of the discussion, however, was the process of *creative wealth destruction* that had made this activity possible in the first place. The same dictum applies to all other institutions of Megata's making. There were quite a few of them. Almost overnight, the Megata reform departed from the

decentralized, even fragmented structure of Korea's economy and attempted to install a modern financial system characterized by sound dealings, and in which several types of institutions were to address different sectors of the credit business. All institutions, including agricultural and industrial banks, established in accordance with Imperial Ordinance 13 (March 1906), built on the credit pyramid imposed by Tokyo and its imperialist agenda.

In this context, it is therefore instructive to reiterate Metzler's and Schiltz' indication of the profoundly *political* and *hegemonic* dynamics behind the reform effort, in fact creating and cementing a core-periphery distinction within Japan's nascent empire in Asia (Metzler 2006; Schiltz 2012a). As pointed out in the epitaph to this chapter, "Japan and the yen ha[d] an overpowering dominance, and arrangements [were] made with an eye to the needs of Japan and the yen, rather than in consideration of the other areas and their currencies." This had very tangible implications. Although the currency reports of the day indicate that Korea did adopt the gold standard, Korea's standard was, by its very design, a *gold-exchange standard* levered upon the Japanese yen. As in Japan proper (the standard of which was arguably levered on the British pound), gold coins were never in circulation. Neither did the First Bank hold much gold as a convertible reserve. First Bank notes were convertible into BOJ notes, which were convertible into gold only in Tokyo: "thus there was almost no connection between the issue of Daiichi Ginkō notes and the gold reserve." This disconnect was further reinforced by certain amendments with regard to the issue of notes. Although the bank was in principle backed by a gold money reserve or BOJ convertible notes held at the First Bank's Seoul branch, it was also allowed to issue notes backed by government bonds and other credit securities, and it could issue notes beyond the limits of convertibility on the condition of approval by the Finance Minister.

This model of the inverted credit pyramid, de facto a technocratically managed structure of debt claims built upon a very limited amount of physical gold, would nevertheless become the blueprint for all of Japan's financial domination in Asia. In the following paragraphs, we will explore what this entailed for the region's economies.

World War I: Japan's Balance-of-Payment Surplus, and Foreign Loans

> As we cannot expect the newly established Chinese government to unify the currency in Manchuria overnight, it is up to us [Japanese] to unify Manchuria's currency, further trade among Japan, Korea, and Manchuria, and gradually expand our economic zone to the south.
> —Nishihara Kamezō

World War I was, in many ways, a watershed in the history of modern Japan and its fragile gold standard. As said before, and typically erased from the Meiji leaders' attempts at remembering this period, gold standard adoption had been a mixed blessing. In the long run, expansionary military build-up programs in the wake of

both the Sino-Japanese and Russo-Japanese War, together with the country's turn to industrialization depressed the state's finances. Steep imports of raw materials and machinery meant that Japan would run constant trade deficits for the whole period between 1894 and 1914, with the exceptions of 1906 and 1909. This was in sharp contrast to the years leading up to the Sino-Japanese War. In effect, the foreign borrowing necessary to maintain the course towards a nascent empire would be massive; at the eve of World War I, it stood at 36% of GNP. De facto, it led to a situation in which "[Japan] floated overseas bonds in order to maintain the gold standard," as one Bank of Japan official put it (Fukai 1941, p. 80). What made this unsustainable debt burden nevertheless possible was not the health of Japan's economy but the Anglo-Japanese alliance.

World War I changed the picture of a nation that seemed on a path of unsustainable borrowing. As could be expected, the very outbreak of the war was disruptive. The prices of products Japan was importing from the European belligerents rose sharply. Much worse, however, were changes in the British and, to a lesser degree, American attitudes with respect to willingness to lend to Japan. Whereas Japan's creditors had already grown wary of the country's apparent insatiable hunger for new loans (between 1904 and 1913, Japan had borrowed approximately one billion yen in British and American funds), the British decision to suspend gold shipments suddenly upset the fragile balance on which the finances of the late Meiji state rested. Most importantly, it strangulated its lifeline to the London capital market, the financial foundation of Japan's import and export trade.

Soon, however, the tide turned to Japan's favor. Diminished competition from European producers led to the consequent (and unprecedented) demand for Japanese products. European-produced beet sugar, to take one example, more or less disappeared from world markets during the war, thereby greatly increasing the price of cane sugar and the profits of Taiwan's sugar producers – especially the Suzuki concern's sugar interests and its bank, the semi-governmental Bank of Taiwan (BOT). What was more, Japan also was to build up a very large gold reserve, in the form of foreign-held currencies. Through the smooth working of its state sponsored exchange bank, the Yokohama Specie Bank (YSB; 横濱正金銀行), ¥94.4 million of gold poured into the country in 1916; in 1917, the number rose to ¥247.2 million. According to leading financier Inoue Junnosuke (井上準之助), the war economy had supplied Japan with enough financial resources potentially to liquidate all its outstanding debts, both foreign and domestic (Inoue and de Bunsen 1931, p. 38).

Given the country's dependence on foreign trade for development, however, this was not what happened. It may not be well known but, in the period between 1915 and 1918, Japan lent a total of ¥640,627,000 to Great Britain, France, and Russia; a syndicate of Japanese banks functioned as underwriter. Itō Masanao has made clear that these loans were primarily economic in nature, and stemmed from problems associated with limited specie exchange (Itō 1989). This was especially the case after 1917 (the year in which the United States too placed an embargo on the export of gold). Facing the prospect of not being able to remit its growing trade surpluses, Tokyo decided to engage in lending in an effort to "fund its own trade." Imports of

raw cotton from India were a key commodity. Gold shipments were immediately transferred to Indian wholesalers, in order to enable further imports into Japan and thus sustain furious economic growth. Japan's first experiment with foreign lending was not entirely successful. Although Great-Britain and France would pay back their debts by the end of 1924, the Bolshevik revolution of 1917 would be the cause behind Russia's default. Approximately one third of the loan issues, or ¥220 million, would never be recovered.

There were also other loan schemes, and these carried all the hallmarks of a pronounced political interest. The initiation of the latter was, again, connected with events on the European theater of war. Most prominently, wartime expediency forced Great Britain to abandon its aggressive economic and political activities in China. Both Japan and the United States, temporary winners of hostilities in Europe, would declare their candidacies at filling the power vacuum. At stake were the politics of "exclusive rights" that the European powers had wrought upon China, which they regarded as immutable and inalienable.

The American reaction to the opportunity represented, rhetorically at least, a break with the concession imperialism that had been the hallmark of European meddling in Chinese affairs. As the United States had largely sailed an isolationist course roughly until 1890, its influence in Asia had effectively been undercut by a series of contracts (for mining rights, building railways, a telecommunications infrastructure, and so on) through which China's sovereignty had been further eroded. The United States' own adoption of the gold standard, in particular, put an end to isolationism and saw the birth of a strand in American politics that stressed greater American involvement in international affairs. The US government sponsored the establishment of a Commission on International Exchange whose aim was to investigate the possibility of establishing gold-exchange standards in, among other countries, Panama, Cuba, the Dominican Republic, and Mexico (United States 1903). Its ambition was the establishment of a *dollar bloc*, no more, no less. The currencies of these countries would not be based on the pound sterling (the system pioneered by the British in India) but on the US gold dollar. China remained, however, the *nec plus ultra* in the Great Power scramble for concessions and favorable trade agreements, although US interests there had been substantially weakened.

Preempting a confrontation with the USA, the Terauchi government retorted to the Pan-Asian ideology and tactics that would otherwise characterize much of the 1930s. Concretely, the administration sought out to sideline the multilateral China consortium, a syndicate of British, French, German, and American bankers committed to sharing equally all future business in China. The architect of Japan's strategy was Nishihara Kamezō 西原亀三, at the time an unknown Japanese businessman and adventurer, who had sought to make his fortune in Korea (interestingly, Nishihara had been a harsh critic of the Megata reform; Megata allegedly tried to have him expelled from Korea).

During a series of several (partially secret) missions to China in early 1916, Nishihara would negotiate several loans on behalf of the Japanese government. After 1916, Nishihara's activities accelerated. When both Ōkura and Co. and YSB proved

unwilling to lend to the Bank of Communications, Nishihara chose to side with Transportation Minister Cao Rulin (曹汝霖) in the new regime led by Duan Qirui (段祺瑞). He signed a preliminary loan for ¥5 million. In the course of 1917 and 1918, negotiations concerned even larger loan issues. Clearly, the stakes had gotten higher, yet Nishihara wanted to reinforce and cement his role as a political fixer.

Ultimately, another five loans were concluded, in the following chronological order:

1. The Kirin–Hueining Railway Primary Loan (June 18, 1918): ¥10 million for linking Manchuria and northern Korea
2. The Mine and Forestry Loan (August 2, 1918): ¥30 million for developing gold mines and forests in the Russian zone of northern Manchuria
3. The Manmō Four-Railway Loan (September 28, 1918): ¥20 million for linking Manchuria and Mongolia
4. The Shandong Two-Railway Preliminary Loan (September 28, 1918): ¥20 million for linking Shandong with central and southern China; the railway was to be a jointly-owned Chinese and Japanese line
5. The War Participation Loan (September 28, 1918): ¥20 million

These, together with the earlier loans, amounted to no less than ¥145 million. Interestingly, three of the loans had been hastily concluded on the Terauchi cabinet's final day in office.

A full discussion of the Nishihara loans falls beyond the scope of the chapter. Here we suffice with pointing out the rationale behind them. Nishihara was particularly concerned with schemes for economic cooperation and the need to develop a unified currency zone – the "yen bloc" (*en burokku* 円ブロック). Especially after the US entry into World War I (1917), this policy stance hardened. Probably blind to the irony, Tokyo proposed a "Monroe Doctrine for East Asia, or Pan-Asianism" (*Tōyō no Monrō sunawachi zen-Tōyō shugi* 東洋のモンロー即ち全東洋主義). The notion of an "East Asian self-sufficient zone" (*Tōyō jikyūken* 東洋自給圏), in particular, was a radical departure from former foreign policy ideology.

The idea certainly went beyond the earlier set-up of gold-exchange standards in Taiwan and Korea. Nishihara and his co-conspirators envisaged a more radical connection between the Chinese and Japanese monetary systems than yen-based gold exchange standards (*engawase ken* 円為替圏 or yen *exchange* bloc). What they had in mind was a full-fledged bloc, encompassing all the countries in Japan's sphere of influence, in which the Japanese yen would be the sole standard against which other currencies would fluctuate (*enkei tsūkaken* 円系通貨圏). Gold currency notes of the same type as Japanese currency were to be issued, backed by Japanese gold notes. In Nishihara's original proposals, the convertibility of local currency with Japanese paper yen was set at the rate of 1:1; Bank of Chosen notes too were to be allowed for circulation in China, at the same rate. This would wipe out transaction costs and clearly favor Japanese commerce at the expense of the Western powers. Further imperial overtones were obvious: as the host nation, China would "invite" high-profile Japanese to oversee the reforms. In an attempt to sugarcoat the

arrangement, Nishihara promised Japanese assistance in increasing Chinese custom duties with the aim of stimulating the country's domestic development and trade, in exchange for "the abolition of export tariffs on cotton, wool, iron, and copper, and possibly two or three other commodities." It was difficult not to see the geopolitical implications of such an arrangement.

Although the Nishihara loans were eventually written off (cf. infra), Nishihara must be considered as a visionary, albeit an imperialist one. If anything, his arguments foreshadow the grim realities of the 1930s and prefigure the German and Japanese militarist belief that war was inevitable. After this war, he predicted, the world would fall apart into several blocs, "be it in the form of economic wars between the powers, or [in the form of] the unification of one British empire, or realized through extreme protectionist measures [*kyokutan naru hogo seisaku* 極端なる保護政策]." His predictions would turn into a self-fulfilling prophecy, propelled by Japan's increasingly unbridled military apparatus.

And yet, as said above, hardliners had to shelf their plans for the time being. Once the foreign powers had been alerted to the magnitude of the Nishihara loans, they conspired for the loans' immediate termination. To its international embarrassment, the Japanese government was pressured into writing them off completely, with the exception of a token repayment of ¥5 million. Thus, ¥140 million was unaccounted for, which probably went to line the pockets of corrupt Chinese officials. In hindsight, the main effect of the Nishihara loans may thus very well have been merely to prolong China's civil war. Consecutive Japanese administrations implored the loss of foreign exchange and prestige this episode implied. The financier and statesman Inoue Junnosuke put it as follows in 1926: "these investments with the central and provincial governments of China—investments running to several hundred million yen—resulted in a dead loss, and today Japan can recover neither the capital which she thus locked up nor one penny of interest on it. To put the matter in a nutshell, I would say that foreign investment was not practiced by this country, and that such trifling investments as were effected might just as well have been thrown into the sea" (Inoue and de Bunsen 1931, p. 37).

The 1930s: Manchuria

> We pioneered a new realm [*atarashii tenchi*], and had the genuine intention of partaking in the building of a new country. Owen Lattimore called Manchuria the "cradle of conflict" in East Asia. Until the construction of Manchukuo, it was just that. But turning [that situation] around, and making it into the cradle of East Asia's peace was the honest intention of those who took part in the building of Manchukuo.
> —Hoshino Naoki, Mihatenu yume (1963)

Much later, i.e., after the crisis-ridden 1920s, blocist ideology would make a comeback. And once more, the object of Japan's currency imperialism was China, in particular the country's Northeastern regions. Actually, Japanese financial interests in Manchuria were long-standing; they dated from the 1890s, when

nascent Japanese imperialism started to flex its muscles on the Asian mainland. The Bank of Chosen developed its function as an overseas bank or *kaigai ginkō* because Korea's trade deficit with the Japanese mainland and its consequent balance-of-payments problem made it necessary for the bank to venture into those countries with which it had a trade surplus. Bank of Chosen bills thus almost necessarily came to circulate in the adjacent territory of South Manchuria. According to the *Economic History of Manchuria*, not accidentally published by the bank, "the demand was always present, and the Bank in most cases had simply to follow it" (Chosŏn Ŭnhaeng 1920, p. 285). As could be expected, it greatly expanded its network of branches in Manchuria. Since the outset of its establishment, it had maintained a branch in Andong (immediately over the Yalu River); to this were added branches in Mukden, Dairen, and Changchun (all founded 1913), and, later, Sipingjieh (1914), Kaiyuan (1915), Harbin (1916), Fujiadian (1916), Yingkou (1916), and Longjing (1917). Remarkably, from 1918 onward, the bank's business in Manchuria and mainland Japan equaled and even overtook its business in Korea proper, in terms of both deposits and loans. Important for the discussion, BOC was the most vocal proponent of incorporating Manchuria within a gold-yen bloc, an orientation that is related to its cadre's closeness to the Terauchi government and Nishihara (Shōda Kazue, finance minister at the time, was a former BOC governor).

BOC did, however, have to reckon with a formidable competitor: the Yokohama Specie Bank, which had been charged with money issuance in the aftermath of the Russo-Japanese war. Much has been said of the fact that it was strongly opposed to large scale monetary reform, and that it sought to keep silver as an integral part of Manchuria's byzantine monetary fabric. Several authors have explained it as pragmatic, yet this misses the point. As I argue in forthcoming research (Schiltz forthcoming), it points directly to a concern with its management strategy of hedging all foreign exchange risk, which had been mandated in the early 1890s, and which stemmed from the volatility of the world silver price. For the same reason, YSB also resisted the demands to give preferential treatment to Japanese customers; its silver using clientele was key to its business success. In any case, YSB remained a firm and stabilizing factor in Japanese policy versus Manchuria during the whole prewar period. Unlike BOC, which in the 1920s even came on the verge of bankruptcy, YSB and its currency enjoyed a sterling reputation among Chinese, Manchurian, and Japanese business circles.

The early 1930s, however, reignited the gold versus silver debate in a novel fashion. Unsurprisingly, military hostilities were the catalyst of the feud. As is well known, the Japanese military used an alleged Chinese provocation as pretext for an occupation of cities and towns along the line of the South Manchurian railway, which soon extended to all of Manchuria. Shortly thereafter, they created the semisovereign state of Manchukuo, with the aim of turning it into an industrial powerhouse for the militarists. Politically, Tokyo confirmed its international isolation by withdrawing from the League of Nations. Takahashi Korekiyo's decision to take Japan off the gold standard signaled the end of Japanese internationalism in financial matters.

The impact of Japanese control over Manchuria was immediate and total. Together with establishing Japanese military and political dominance, the occupying forces set out to control the region by economic and monetary means. They immediately confiscated all banks, put them under the supervision of employees of SMRC, YSB, or BOC, and seized both their assets and ledgers. BOC was eager to declare its candidacy at leading the new country's monetary organization, but, given Japan's own retreat from the gold standard, found it difficult to convince the political class. YSB silver advocates were more successful. Stressing the importance of stabilizing Manchuria rather than implementing yet another large-scale reform effort, and pointing out Japan's own need for bullion, they explained that a unification of Manchurian money on a silver bullion standard would be the only viable option. Their proposals were accepted in 1932, although ultimately all references to the convertibility of the Manchurian paper yuan were omitted from the charter of the Central Bank of Manchukuo! Thus, from a rather optimistic assessment of Manchuria's monetary and financial future, even the gold proponents made a complete U-turn. Still, the process of currency unification was considered a success. In the words of a contemporary observer: "Thus far the chief benefit of Japanese rule, the new banking system, has abolished fiat currency which the peasants formerly were forced to accept by avaricious warlords. As remarked earlier, they issued crisp but soon worthless paper [. . .] in exchange for the great soya-bean crop [. . .] and for other cereals, which they then sold for hard money. [. . .] The Central Bank of Manchukuo [. . .] converted most of the more reputable old notes. It succeeded in stabilizing the Manchukuo yuan [. . .] at a par with the China dollar" (Snow 1933, pp. 271–272).

In the end, however, it was sharp swing in the international silver price that would necessitate the reorientation of Japan's monetary position in Manchuria. Initially, Manchuria's silver standard had been a boon. Not in the least, it had functioned as a buffer against the Great Depression which, by the early 1930s, was in full swing and had rapidly enveloped the whole world. Yet Roosevelt's controversial silver buying policy (1934) turned the tide. As explained in *The Money Doctors from Japan* (Schiltz 2012a), and contrary to the official explanation, the program's rationale was not so much economic as geopolitical. Roosevelt was determined to spite Tokyo for its own unilateral policies. The silver purchase program was a subtle, yet powerful, way to do so. The latter explains, among others, why the policy was also so detrimental to Chinese interests, a consequence of which Roosevelt had been warned. If anything, the effect of the silver purchase program was a *maelstrom* of deflation. Put simply, as silver was not a mere commodity, but China's money, the outflow of silver sent commodity prices plummeting, translating into a fall in income and employment (Friedman 1992). Within Manchuria, it became increasingly difficult to maintain stable exchange rates between the still circulating silver and gold notes.

Eventually, and clearly frustrated by American policy, Manchurian authorities opted for a radical solution. In September 1935, they agreed to couple the Manchurian yuan to the yen, in other words, to have the economy of Japan proper subsume the Manchurian economy (*naichika*). In a fateful twist of irony, Roosevelt's controversial silver purchase program had only further buttressed Japan's choice for autarky.

From the Outbreak of the Second Sino-Japanese War (1937)

In the aftermath of this Manchurian episode, the technocratic grip on the monetary affairs of the *gaichi* was strengthened even further. In a typical example of overconfidence that characterized the bureaucratic elite, the Kyoto economist Matsuoka Kōji hailed the gold-exchange standard as the future of monetary arrangements (1936; 1938); establishing an economic zone on the basis of the latter was Japan's destiny, an "important and pressing" need.

His view is actually doubly meaningful, because he is correctly credited for being the first to describe Japan's standard as a gold-exchange standard levered upon the British pound. And indeed, if the set-up of monetary standards in China, Mongolia, the Dutch Indies, etc. seems similar to, in particular, the country's early experiences with gold monometallism, this is *because it is*. The mechanism behind Japan's gold-standard was described long ago by Kojima Hitoshi (1981). As in the context of the Indian gold-exchange standard, the key was a currency board (Jpn. *azukeai kanjō* 預け合い勘定, literally "joint custody account"). The function of the latter was to aid with pegging or fixing the value of the soft currency vis-à-vis the value of the core currency, primarily by selling/buying foreign exchange between legally stipulated band (Keynes 1913; Kemmerer 1905; United States 1903). One way in which Japan's gold standard set itself apart from the Indian set-up is that the accounts of the Japanese government and the Bank of Japan in London were administered by a semi-official Japanese institution (the Yokohama Specie Bank) rather than an foreign entity. Another is that these accounts included a "specie reserve account" on the basis of which paper money in Japan was issued.

The latter had led to heated debates in Japanese monetary policy circles. Some had compared the unorthodox practice to the "Ninben" katsuo-don (bonito and rice bowl) shop, which had famously issued its own bills on the security of its reserves of fish still swimming in the sea (Metzler 2006, p. 98). And yet, difficulties of maintaining the yen par notwithstanding, there is no denying that this system worked. In the case of colonial monetary systems, however, the mechanism and its metaphor were stretched to its very limits. Importantly for our discussion, "anchors" were levered upon each other. Concretely, the monetary systems of Japan's periphery (colonial yen, expressed as 圓, rather than the yen 円 used within the Japanese islands) were anchored to the yen in the way the yen was anchored to gold. As put by Metzler: "What we could call first-order money in this system—gold—was leveraged into a larger supply of gold-based national monies, such as the British pound—'second-order' money. This second-order money in turn constituted the monetary base for banks who leveraged it into a much greater volume of money by making loans and creating bank deposits—'third-order' money. In a like way, gold-exchange standards in peripheral countries pyramided gold-based foreign exchange money such as the British pound into third-order or fourth-order money" (Metzler 2006, p. 37).

Clearly, there were and are limits to leverage. As the epitaph to this chapter reminds us, the pitcher goes so often to the well that it is broken at last. In August 1941, when the Roosevelt administration implemented a financial siege and made

the already inconvertible yen illiquid (Miller 2007), something had to give. That "something" turned out to be any meaningful limitation, economic or not, on the arbitrary expansion of the money supply and thereby, debt claims. As the chain connecting the proverbial gold anchor to Japan's gold-exchange boats was then factually severed, the boats drifted and swayed according to the mercurial demands of Japan's war machine.

As could be expected, military dominance led to a dangerous experiment with "inflation capital": freshly created (and inconvertible) yen that were used to buy what the Japanese military wanted, and that denied the captive regions and countries the foreign exchange that might have been used for their own growth and industrialization. What this meant for countries and territories occupied by Japan we know from their rates of inflation: "From the beginning of the Pacific War to the end of the War, the amount of currency issued in the several regions [of the Greater East-Asian Financial Sphere] increased, and the price index [of these regions] shows an identical pattern. Yet, one should not lose sight of the fact that this increase took the form of a rising wave, originating in the center and growing toward the periphery [.. .]. Comparing the outset of the war with its end, prices in Tokyo rose by a factor of one-and-a-half, whereas they rose extraordinarily [elsewhere], i.e., by a factor of 350 in Singapore, and by a factor of 1,850 in Rangoon" (Yamamoto 1997, p. 19).

As explored in a recent documentary by Japan's public national broadcasting company, the folly that was the "Greater East-Asian Financial Sphere" (Shimazaki 1987; 1988a, b, c, d) leaves its traces even today. In the accounts of the Ministry of Finance, one entry entitled "extraordinary war expenditures" still shows a debt running into more than 41 billion yen. Unlikely to ever be repaid, this loan, created out of thin air by the penstrokes of Japan's military planners, nevertheless presents very real goods and services that were extracted from the occupied territories and put to the service of the imperialist center. In Japan's imperialist project, for which, according to Ishiwara Kanji, "the war was to be used to feed the war" (*sensō wo motte sensō wo yashinau* 戦争をもって戦争を養う), some players thus turned out to be more equal than the others. Simply put, not even one domestic Japanese yen had been used for the conquering of Asia (Tatai 2014). Vice versa, if viewed through the eyes of the occupied regions and territories, they had been paying for their own occupation. Clearly, as has been argued in the context of the Nazi economy: "[u]nlike communism, fascism was not for export, and the collaboration and the sensibilities of the conquered peoples were of relatively little concern to the conquerors" (Wolfe 1955, p. 400).

Cross-References

- ▶ International Monetary Regimes: the Gold Standard
- ▶ International Monetary Regimes: the Interwar Gold Exchange Standard
- ▶ Money in Wars

► The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy

References

S.a. En no sensō (the war of the continental yen) "圓の戦争." (2011) *NHKスペシャル*. Tokyo. http://www6.nhk.or.jp/special/detail/index.html?aid=20110814

Bytheway SJ (2005) Nihon keizai to gaikoku shihon - 1858–1939 (the Japanese economy and foreign capital - 1858–1939). 日本経済と外国資本—1858–1939. 東京: 刀水書房

Bytheway SJ, Metzler M (2016) Central banks and gold: how Tokyo, London, and New York shaped the modern world. Cornell studies in money. Cornell University Press, Ithaca

Chosŏn Ŭnhaeng (1920) Economic history of Chosen, Comp. in commemoration of the Decennial of the Bank of Chosen. Chosen, Seoul

Drake PW (1993) Money doctors, foreign debts, and economic reforms in Latin America from the 1890s to the Present. SR Books, Wilmington. http://www.amazon.com/dp/0842024344

Friedman M (1992) Franklin D. Roosevelt, silver, and China. J Polit Econ 100(1):62–83. http://www.jstor.org/stable/2138806

Fukai Eigo 深井英五 (1941) Kaiko nanajunen (Looking back at 70 years of my life). 『回顧七十年』. 岩波書店. http://ci.nii.ac.jp/ncid/BN02993778

Hoshino Naoki 星野直樹 (1963) Mihatenu yume: Manshūkoku gaishi (Dreaming an impossible dream: an unofficial history of Manchukuo). 『見果てぬ夢：満州国外史』. ダイヤモンド社. http://ci.nii.ac.jp/ncid/BN0663518X

Hunsberger WS (1938) The yen bloc in Japan's expansion program. Far Eastern Survey 7(22): 251–258. https://doi.org/10.2307/3021427

Inoue J, De Bunsen EH (1931) Problems of the Japanese exchange, 1914–1926. Macmillan, London

Itō Masanao 伊藤正直 (1989) Nihon no taigai kin'yū to kin'yū seisaku, 1914–1936 (Japan's international finance and financial policies, 1914–1936). 『日本の対外金融と金融政策：1914～1936』. 名古屋大学出版会. http://ci.nii.ac.jp/ncid/BN04118008

Japan. Ōkurashō (1900) Report on the post-bellum Financial Administration in Japan, 1896–1900. Printed at the Government Press, Tokyo

Kemmerer EW (1905) The establishment of the gold exchange standard in the Philippines. Q J Econ 19(4):585–609. https://doi.org/10.2307/1885290

Kemmerer EW (1912) The recent rise in the price of silver and some of its monetary consequences. Q J Econ 26(2):215–274. https://doi.org/10.2307/1884764

Keynes JM (1913) Indian currency and finance. Macmillan, London

Kojima Hitoshi 小島仁 (1981) Nihon no kinhon'isei jidai (1897–1917) (Japan's gold-standard era (1897–1917)). 『日本の金本位制時代：1897～1917：円の対外関係を中心とする考察』. 日本経済評論社

Matsukata M (ed) (1899) Report of the adoption of the gold standard in Japan. Printed at the Government Press, Tokyo

Matsuoka Kōji 松岡孝児 (1936) Kin kawase hon'isei no kenkyū (An inquiry into the gold exchange standard). 『金爲替本位制の研究』. 日本評論社

Matsuoka K (1938) L'Etalon de change or en Extrême-Orient, par Kôji Matsuoka,…. P. Geuthner/Mitsukoshi, Paris/Tôkyo

Metzler M (2006) Lever of empire: the international gold standard and the crisis of liberalism in Prewar Japan. Twentieth-Century Japan 17. University of California Press, Berkeley

Miller ES (2007) Bankrupting the enemy: The U.S. Financial Siege of Japan before Pearl Harbor. Naval Institute Press, Annapolis

Namikata Shōichi 波形昭一 (1985) Nihon shokuminchi kin'yū seisakushi no kenkyū (A study of the history of financial policies towards Japan's colonies). 『日本植民地金融政策史の研究』. 早稲田大学出版部. http://ci.nii.ac.jp/ncid/BN00205982

Rosenberg ES (2003) Financial missionaries to the world: The politics and culture of dollar diplomacy, 1900–1930. American Encounters/Global Interactions. Duke University Press, Durham

Schiltz M (2012a) The money doctors from Japan: finance, imperialism, and the building of the Yen Bloc, 1895–1937. Harvard University Asia Center, Cambridge

Schiltz M (2012b) Money on the road to empire: Japan's adoption of gold Monometallism, 1873–971. Econ Hist Rev 65(3):1147–1168. https://doi.org/10.1111/j.1468-0289.2011.00619.x

Schiltz M (forthcoming) On an Even Keel. Silver Risk

Shimazaki Kyūya 島崎久弥 (1987) "'Dai toa kin'yūken' no keisei katei" (The process of the formation of the "Greater East-Asian Financial Sphere") 1.「大東亜金融圏」の形成過程-1-. 『経済系』 153 (October):p1–40. https://ci.nii.ac.jp/naid/40000862587

Shimazaki Kyūya 島崎久弥 (1988a) "'Dai toa kin'yūken' no keisei katei" (The process of the formation of the "Greater East-Asian Financial Sphere") 2.「大東亜金融圏」の形成過程-2-. 『経済系』 154 (January):p19–49. https://ci.nii.ac.jp/naid/40000862591

Shimazaki Kyūya 島崎久弥 (1988b) "'Dai toa kin'yūken' no keisei katei" (The process of the formation of the "Greater East-Asian Financial Sphere") 3.「大東亜金融圏」の形成過程-3-. 『経済系』 155 (April):p74–108. https://ci.nii.ac.jp/naid/40000862619

Shimazaki Kyūya 島崎久弥 (1988c) "'Dai toa kin'yūken' no keisei katei" (The process of the formation of the "Greater East-Asian Financial Sphere") 4.「大東亜金融圏」の形成過程-4-. 『経済系』 156 (July):p30–75. https://ci.nii.ac.jp/naid/40000862631

Shimazaki Kyūya 島崎久弥 (1988d) "'Dai toa kin'yūken' no keisei katei" (The process of the formation of the "Greater East-Asian Financial Sphere") 5.「大東亜金融圏」の形成過程-5-. 『経済系』 157 (October):p1–50. https://ci.nii.ac.jp/naid/40000862637

Snow E (1933) Far eastern front. H. Smith & R. Haas, New York

Soyeda J (1899) Letter from Japan. Econ J 9(35):469–477. https://doi.org/10.2307/2957078

Tatai Yoshio 多田井喜生 (2014) Shōwa no meisō: 'dai-ni Manshūkoku ni tsukarete (An aberration of the Shōwa period: haunted by a 'second Manchuria'). 『昭和の迷走:「第二満州国」に憑かれて』. 筑摩書房, Tōkyō

Thomson SC (1971) Filibustering to Formosa: general Charles LeGendre and the Japanese. Pac Hist Rev 40(4):442–456. https://doi.org/10.2307/3637704

United States (1903) Stability of international exchange. Report on the introduction of the gold-exchange standard into China and other silver-using countries. Government Printing Office, Washington, DC

Wolfe M (1955) The development of Nazi monetary policy. J Econ Hist 15(4):392–402. https://doi.org/10.1017/S0022050700056552

Yamamoto Yūzō 山本有造 (1997) 'Dai Tōa kin'yūken' ron (A theory of the Greater East-Asian Financial Sphere).「大東亜金融圏」論 (1940年代日本植民地帝国の諸相). 『人文學報』 79:1–26. https://doi.org/10.14989/48494

European Monetary Integration

32

Emmanuel Mourlon-Druol

Contents

Research supporting this chapter has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 716849), project EURECON: The Making of a Lopsided Union, Economic Integration in the European Economic Community, 1957–1992. This chapter was written while the author was Jean Monnet Fellow at the European University Institute's Robert Schuman Centre.

E. Mourlon-Druol (✉)
Adam Smith Business School, University of Glasgow, Glasgow, UK
e-mail: Emmanuel.Mourlon-Druol@glasgow.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_26

Abstract

This chapter analyzes the process of monetary integration in Europe between 1945 and 1992, when the Treaty of Maastricht introduced the euro as the single currency of most member states of the European Union (EU). It introduces the main theories used in dealing with the monetary integration of Europe and provides the institutional and global context in which the discussions took place. It analyzes the different types of monetary integration that Europe used and debated, including the use of the unit of account, the debate over the introduction of a parallel currency, the exchange rate coordination, and the creation of a single currency. It finally presents the different phases of the creation of the euro.

Keywords

European Unit of Account (EUA) · Exchange Rate Mechanism (ERM) · European Monetary System (EMS) · European Currency Unit (ECU) · Parallel currency · Euro

Introduction

Europe's quest for the creation of a single currency is by all accounts an unprecedented one. Historically, currencies have been backed by a state, whether a small unitary state or a federation. The euro, created in 1999, is a currency without a state. The European single currency thus poses many challenges to the historian wishing to understand what peculiar forces drove the euro's unusual constitution.

As this is a handbook on the history of money, this chapter looks exclusively at the currency aspect of European monetary integration. The economic dimension will only occasionally be mentioned for a fuller account see (Mourlon-Druol 2018). The chapter first recounts the main theories that provide useful background to understand the debates about currency integration in Europe; then it sets out the institutional and global economic and financial contexts in which the debates took place, the four different types of monetary integration that were discussed in Europe, and finally the creation of the euro.

What does "European monetary integration" mean? What is "Europe" in the context of monetary integration? The expression traditionally covers Western European countries, thus excluding the USSR and Eastern European countries under Soviet influence. "Europe" focuses on the member states of the European Economic Community (EEC), founded in 1958 by the Treaty of Rome and transformed into the European Union (EU) in 1992 by the Treaty of Maastricht.

What is "monetary?" Money has three commonly accepted functions. Money is a means of exchange, a unit of account, and a store of value. A "unit of account" means that money is used as a common measure to indicate the value of goods and services in an economy. A "means of exchange" reflects the fact that the use of money solves the issue of the lack of coincidence of wishes. A "store of value" underlines the fact that money is a safe value. Monetary integration, in the European

context of multiple nation-states, therefore includes important developments about the use of a single unit of account – but not yet single currency – across the EEC. The unit of account was a critical tool for the functioning of the EEC, with regard to the common budget and the running of specific common policies that required a financial instrument for the cross-comparison of prices.

Finally, what is "integration?" In this chapter, for the sake of simplicity, "integration" will be used as a catchword that encompasses the meanings often attributed to "cooperation" (Cooper 2006). "Integration," in a European context, involves sensu stricto, the transfer of competences into some sort of nonnational executive body. This process is also known as supranational integration. Because we know where the story of European monetary integration ended, we tend to use the word anachronistically and superimpose a number of preconceptions of the 1990s to much different contexts and ambitions. As a consequence, the use of the word "integration" tends to leave aside the many alternative options that were explored but did not materialize to achieve European monetary integration (see section "The Creation of the Euro").

Monetary Theories and European Integration

In thinking about European monetary integration, three sets of ideas regularly come back in the debates: the relationship between money and the state, the "economists" versus "monetarists" wrangle, and the theory of optimum currency areas (OCA). This section examines them in turn.

The Relationship Between Money and the State

Traditional thinking about money holds that the creation of currency is the domain of the state. The influential work *The State Theory of Money* of German economist Georg Friedrich Knapp established the chartalist school of monetary theory. This school of thought holds that government, rather than spontaneous relations of exchange, can issue money (Knapp 1924). In that sense, the monetary integration of the EEC represents a novel way of issuing money, through an independent and supranational central bank, the ECB, which is itself part of a highly complex institutional structure and judicial order, neither fully federal nor simply intergovernmental, the EEC/EU. As the then president of the ECB Wim Duisenberg put it in 2002, the euro "is not backed by the durability of the metal or by the authority of the state" (Duisenberg 2002).

Contemporary discussions have regularly taken up this debate. Paul de Grauwe, for instance, famously quipped that "the euro is a currency without a country. To make the euro sustainable a country will have to be created" (De Grauwe 2013, 25). But in the run up to the creation of the euro, the topic was less central. Some policymakers, often federalists, did envisage the creation of a single currency as part of the broader development of a full-fledged European federation. More often

than not, however, monetary integration remained confined to technical discussions, in spite of its major political implications. This is partly due to the fact that monetary affairs were already politically very sensitive, and technically very difficult to agree upon, and policymakers preferred dealing with one issue at a time.

The Debate Between the "Economists" and the "Monetarists"

One long-standing theoretical opposition about the best way to achieve Economic and Monetary Union runs through the history of European monetary integration from the 1960s until the Treaty of Maastricht and under different guises even until the present day, namely, the opposition between the so-called economists and the so-called monetarists. Their opposition is related to the strategy to adopt to reach the goal of the creation of a single currency in the EEC/EU. According to the "economists," monetary integration in Europe could only be achieved after the EEC member states' economies would have converged. The economist approach was also dubbed the "coronation theory," in that it was only after a long period of preparation that a king or a queen was eventually crowned. According to the "monetarists," the introduction of a single currency in Europe would per se lead to the economic convergence of the member states taking part in the initiative. This brand of monetarism is not to be confused with Milton Friedman-inspired monetarist thinking. "Economists" were therefore reluctant to agree upon any advance in monetary integration and financial risk-sharing so long as EEC member states' economies had not yet converged. "Monetarists," on the opposite, argued that EEC member states' economies would only converge once European monetary integration would have been fully achieved. The differences between economists and monetarists were often blurred, as both camps held valid points. To simplify, West Germany, Denmark, and the Netherlands tended to be described as economists, while France, Italy, and the European Commission tended to be presented as monetarists.

The economist vs monetarist wrangle was an important one in the debates about monetary integration in Europe, as any change was suspended to its resolution. As detailed in section "The Snake, 1972–1978" "economists" adjourned any reform of the EEC's exchange rate system until EEC member states' economies had converged further, while "monetarists" called for more immediate monetary integration. The economist vs monetarist wrangle survived until the present day, albeit under a different form (Mourlon-Druol 2014). Many Eurozone members still oppose greater financial burden-sharing in the Eurozone until political union has been achieved, while some other members would like burden-sharing to happen immediately for the European monetary union to be sustainable.

The Theory of Optimum Currency Areas (OCA)

Finally, European monetary debates often refer to the so-called OCA theory. OCA theory originates in an eponymous article written by Canadian economist Robert

Mundell in 1961 (Mundell 1961). Mundell identified two factors which he considered necessary for a "currency area," that is, a group of countries sharing the same currency, to be "optimal," that is, not to produce any loss of well-being. These two factors were the free movement of labor and capital. Subsequently, other economists revised and refined Mundell's framework and added new criteria to the list of elements that an OCA should fulfill. To name but a few, these criteria were the openness of the economy (McKinnon 1963), the diversification of production (Kenen 1969), the financial dimension (Ingram 1969), the convergence of inflation rates (Haberler 1970; Fleming 1971), and the homogeneity of preferences within the zone (Kindleberger 1986).

Few perfect OCAs exist at all in the world, and the influence of the OCA school of thought on European monetary discussions is unclear and often exaggerated. If it is true that some European policymakers read and took into consideration OCA-related discussions, it is difficult to find direct causal links between the development of the OCA theory and the thinking about EMU. The so-called OPTICA reports produced by a group of academic economists were one of these exceptions: OPTICA stood for OPTImum Currency Area (Optica 1976, 1977). But the OCA theory was and still is frequently used as an analytical framework to analyze the development, and travails, of the European single currency.

The Institutional and Global Context of European Economic and Financial Integration

Discussions about the monetary integration of Europe occurred in specific contexts: the institutional framework of the European Economic Community (EEC) and the evolution of the EEC in the wider framework of economic and financial globalization. Both aspects are analyzed in turn in this section.

European Monetary Integration in Europe at the End of the Second World War

Before moving to the institutional and global context in which European monetary integration developed, it is first necessary to recall the situation in the immediate postwar period. European monetary relations led to many debates about what institutional framework should govern them, well before the creation of the euro. Even before the creation of the European Economic Community in 1957, currency relations occupied a central place in European policy discussions.

The very basic question of currency convertibility – rather than any more ambitious plan for monetary integration – dominated the immediate postwar period in Western Europe (Eichengreen 1993, 1995). In 1945, European countries were constrained to trade with other countries through bilateral arrangements. Any deficit had to be offset by a surplus; trade was guided by outstanding debts, and based on US dollars, which European countries lacked. In July 1950, the Organisation for European Economic Cooperation (OEEC, inherited from the Marshall Plan) agreed

to create the European Payments Union (EPU). The EPU created a multilateral system of payments replacing the bilateral payment agreements. Only a multilateral system would be able to relaunch the European economy. Each country accumulated its deficits and surpluses with all other countries into one central account with EPU, which was debited or credited by the combined net result of all intra-European transactions. This system allowed countries not to be concerned about a deficit with some countries, since this deficit could be offset by a surplus with some other countries into the overall EPU account on a monthly basis. EPU thus removed bilateral bargaining, reduced transaction costs, and restored multilateral trade in Europe. EPU functioned from July 1950 until December 1958. EPU was of course not about creating a single currency in Europe, but it was an unavoidable stepping stone toward reconstructing a functioning European financial system and gave birth to the first reflections about the monetary organization of Europe.

The Institutional Context: The European Economic Community and Monetary Integration

The Treaty of Rome did not foresee the introduction of a single currency in the EEC. The Treaty of Rome's provisions only indicated that individual member states should consider their exchange rate as a "matter of common concern" (Article 107) and that each member state should maintain "confidence in its currency" (Article 104). Appropriate monetary coordination among member states should prevent any problem arising. This was the only way in which monetary affairs were included in the Treaty of Rome. The silence of the EEC's founding treaty on monetary integration stemmed from the fact that monetary relations were not yet really a cause for concern, since there existed a functioning international monetary system. Some policymakers such as Robert Marjolin or Robert Triffin did think that creating a monetary union in the long term would be a logical move for the EEC, given the inherent weaknesses of the Bretton Woods system, and/or the need to protect the European common market (Dyson and Maes 2016). But this did not translate into the quasi-constitutional framework of the EEC treaty in 1958.

If the Treaty of Rome did not set out in detail a plan to introduce a single currency, the EEC's founding treaty included three institutions that would be central to European monetary cooperation: the Council of Ministers, the European Commission, and the Monetary Committee. The Council of Ministers of the EEC brings together the relevant ministers of each EEC member state according to the topic tackled. A specific meeting is devoted to economics and finance – hereafter, the Finance Council – which gathers roughly once a month. All issues relevant to these policy areas are discussed during a council meeting. The Council of Ministers is an intergovernmental institution in which each EEC member state has one vote (qualified majority voting has been progressively introduced in some policy areas in the EEC/EU).

Article 105 of the Treaty of Rome outlines the creation of a Monetary Committee. The Monetary Committee was the regular institution to discuss currency-related

Table 1 List of European commissioners in charge of "economic and financial affairs" (renamed "economic and monetary affairs" in 1999 and then "economic and monetary affairs and the euro" in 2011) since 1957

1958–1967	Robert Marjolin
1967–1973	Raymond Barre
1973–1977	Wilhelm Haferkamp
1977–1985	François-Xavier Ortoli
1985–1995	Henning Christophersen
1995–1999	Yves-Thibault de Silguy
1999–2004	Pedro Solbes
2004	Joaquin Almunia
2004	Siim Kallas
2004–2010	Joaquin Almunia
2010–2014	Olli Rehn
2014	Jyrki Katainen
2014–present	Pierre Moscovici

issues relevant to the functioning of the EEC. Two members of each EEC member state take part to the meetings (one representative of the central bank, usually the deputy head, and one representative of the finance ministry or equivalent), as well as two members chosen by the European Commission. The Monetary Committee thus allowed the representation of both economic and monetary policymakers.

The European Commission, set up by the Treaty of Rome, is one of the central institutions of the EEC. The commission is administratively divided into several "Directorate Generals" – their number evolved in the history of the Commission – with DG II always being devoted to "economic and financial affairs" (and housed the secretary of the Monetary Committee). One of the commissioners is specifically in charge of economic and financial affairs (Table 1).

Although not envisaged in the Treaty of Rome, two further institutions played a central role in European monetary integration: the Committee of Central Bank Governors, created in 1964, and the European Council, created in 1974.

In April 1964, a decision of the Council of Ministers created the Committee of Governors of the central banks of the EEC – hereafter Committee of Governors (James 2012). The governor of each central bank, as well as one or two alternates, took part in the meetings. The Committee of Governors included all EEC member states, regardless of their belonging to an EEC exchange rate system. The Bank for International Settlements (BIS), in Basel, provided for the Committee of Governors' secretariat, as well as its regular meeting place.

In December 1974, EEC heads of state and government decided to create the European Council, that is, the regular meeting of EEC leaders, three times a year and whenever necessary, to discuss current problems (Mourlon-Druol 2010, 2016). Confronted with international problems that required a coordination at European level – such as the oil crisis, recession, international monetary instability – EEC leaders realized that there existed no EEC institutional mechanism allowing them to meet on a regular basis. The creation of the European Council aimed to fill that

institutional vacuum. The European Council played an often critical role in the evolution of European cooperation and integration broadly speaking. Since the European Council gathered the most powerful administrative level of the EEC's member states, it was able to provide new political impulses necessary for the development of new projects, such as the creation of the EMS, and the euro. The European Council also reflected the disagreements among EEC member states and their inability to reach satisfactory and comprehensive policy agreements, as the Eurozone crisis made plain.

Further to this, the European Monetary Cooperation Fund (EMCF) was created in April 1973 as part of the implementation of the Werner plan. The ambitions of the EMCF were not really high, and it merely limited itself to help in the administration of the European exchange rate system. But the main political and technical discussions remained in the remit of heads of government, finance ministers, and central bankers. When the EMS was created, a European Monetary Fund was also envisaged (and would have replaced the EMCF) but was never established. Discussions about the creation of an EMF have continued until the present day.

The Economic and Financial Context: Europe in a Globalizing World

European monetary integration did not develop in a policy vacuum. Two aspects critically influenced the monetary unification of Europe: the evolution of the international monetary system, and the development of the EEC common, and then single market.

The Evolution of the International Monetary System

Until the late 1960s, the Bretton Woods system provided an international monetary cocoon in which European currency relations were largely protected from global turbulence. Once the first cracks and eventual breakdown of the Bretton Woods system occurred, the question of intra-European currency stability came out in the open. For many European policymakers, one of the appeals of European monetary unification was to contribute to the reform of the international monetary system, challenge the domination of the dollar, and thereby remedy the negative effects of the US currency's domination. Throughout the 1960s, 1970s, and 1980s, the fluctuations of the dollar have thus served as a catalyst to debate European monetary integration (Grygowski 2009).

Delegates from 44 Allied nations met in the Mount Washington hotel in Bretton Woods, New Hampshire, from 1 to 22 July 1944. The interwar international system was generally perceived as a failure: the Great Depression contributed to increase international tensions that escalated into war. The delegates of the Allied nations meeting in Bretton Woods thus sought to establish a new international system that drew on the lessons from the interwar period and would support postwar reconstruction, international cooperation, and international trade. The delegates agreed to set up a new fixed but adjustable exchange rate system in which all currencies would be pegged to the US dollar – with a 1% fluctuation margin, but countries could ask

authorization to the IMF for a 10% devaluation – and only the US dollar would be convertible into gold at a fixed price of 35 dollars per ounce. Two new international institutions were created. The International Monetary Fund (IMF), established in 1945, in order to monitor the functioning of the exchange rate system and more generally to encourage international monetary cooperation and guarantee international financial stability. Also established in 1945 was the International Bank for Reconstruction and Development (IBRD), now part of the World Bank Group, in charge of supporting lending for postwar reconstruction and to less developed countries. The exchange rate system effectively entered into function in 1958, when currencies became fully convertible. Before that date, many countries maintained exchange controls, which made that the free convertibility of one currency into another set out in the Bretton Woods agreement was not implemented.

The ups and downs of the dollar and more broadly the problems in the international monetary system importantly influenced European monetary discussions. In the late 1960s, when the first problems in the Bretton Woods system became obvious, European monetary integration became more urgent (see section "The Snake, 1972–1978"). In the late 1970s, when the dollar depreciated, EEC members – and the German government in particular – feared that this would lead to import inflation to Europe. This fear motivated German chancellor Helmut Schmidt to take the lead in proposing a European response to US policy, which later materialized in the creation of the EMS. In the mid-1980s, the appreciation of the dollar conversely raised concerns in Europe and highlighted again the need to coordinate a response. This time, global discussions about establishing a system of target zones were first privileged with the Louvre accord in 1987. But soon thereafter discussions were oriented toward Europe with the appointment of the Delors committee (see section "The Delors Report"). European monetary integration thus often came to represent a regional European solution to a global problem.

European Monetary Integration and the Single Market

The development of a barrierless internal market in the EEC fed another strand of reflection for the evolution of intra-European currency relations. The Treaty of Rome set out to create a common European market with no tariffs. Initial market integration focused on the free circulation of goods, and the first step accordingly focused on the removal of tariffs imposed by EEC member states. The abolition of those tariffs however made progressively plain that there existed other types of barriers to trade, called non-tariff barriers (NTBs). NTBs became the focus of policy attention from the late 1970s and most importantly with the single market program of the 1980s up until the present day. NTBs relate to any type of qualitative barrier such as an excessive administrative burden, or a regulation, and that discriminates between goods produced nationally and imported goods (with a view to advantage national production).

Currency relations fit into these policy discussions insofar as monetary fluctuations could be considered as a potential barrier to trade. For instance, a devalued currency in one country can discourage from buying goods in a country with a higher currency value as these goods will become comparatively more expensive. In that

sense, if the states belonging to the single market were sharing a single currency, an additional NTB would be removed. In the late 1950s and 1960s, European policymakers did not mention NTBs yet (this came later in the 1970s), but some of them already pointed at the fact that the common market could, and perhaps should, logically be transformed into an EMU to function properly. Another, related line of argument was that the sharing of a single currency, with the macro- and microeconomic benefits it implied, would increase trade among participating countries (Rose 1999; Commission 1990). In the 1980s, Tommaso Padoa-Schioppa famously set out his "inconsistent quartet" (Padoa-Schioppa 1994). The Italian economist argued that among the following four policy options, free trade, full capital mobility, fixed or managed exchange rates, and monetary policy autonomy, one had to give, since not all could coexist at the same time.

Dealing with Surpluses and Deficits

In the context of the EEC common/single market and intra-EEC currency relations, the current account deficits and surpluses of individual member states could easily create difficulties. In the absence of free capital flows, deficit countries were faced with either the implementation of deflationary policies and tax rises to correct the balance – which was obviously unpalatable to the electorate – or the hope that surplus countries (chiefly Germany) would start expansionary policies, which was also unpalatable to an electorate that feared inflation. Monetary union could offer a way out of this political dilemma. Difficulties linked to the evolution of current accounts were important in the reflections about the reform of the snake (see section "The Snake, 1972–1978" below). Harold James has shown that moments of tension in intra-EEC currency relations since the 1950s closely corresponded to persistent German current account surpluses and that moments of "relaxation" corresponded to German current account deficits (James 2012). Discussions on the evolution of European monetary integration thus occurred against the backdrop question of how to deal with the persistent surpluses or deficits of some EEC member states.

A Typology of Monetary Integration in Europe

In presenting European monetary integration, very different forms of monetary cooperation, such as exchange rate coordination and Economic and Monetary Union, are often referred to as if they were interchangeable. Exchange rate coordination and EMU however entail different realities. Exchange rate coordination is concerned with the fluctuation of the value of currencies internationally, while EMU is concerned with the creation of a single currency to be shared among different member states. Similarly, a traditional confusion is made between a common currency and a single currency. Both are often used interchangeably in writing about EMU. One substantial difference exists between the two, however. If both are shared between different countries, a single currency is the only one in circulation, while a common currency may exist alongside other currencies.

This section distinguishes between four different types of monetary integration that were discussed in the EEC/Western Europe since 1945 and sometimes overlapped. It starts with the most basic one, namely, the European Unit of Account (EUA). It then looks into the question of the creation of a parallel currency – sometimes via the EUA – as an alternative strategy to the mainstream discussions about EMU that was eventually not followed but regained a degree of prominence in public debates after the creation of the euro. Third, it moves to the attempts at coordinating exchange rates in Europe, that is, the snake, and the EMS. Finally, it scrutinizes the two different plans that were advanced for the creation of the European single currency, namely, the Werner report (1970) and the Delors report (1989).

The European Unit of Account

The creation and evolution of European units of accounts a technical but critical aspect of European monetary integration. Some policymakers, whether from EEC institutions or at the national level, viewed the development of European units of account as a way to promote a European monetary identity. According to that view, a monetary unit would allow to materialize European integration.

The unit of account was not actually coined but only used for bookkeeping/accounting purposes. The European Payments Union created a European Unit of Account (EUA) in July 1950. The value of this EUA corresponded to the dollar's worth of gold under the original agreement of the Bretton Woods system, that is, 35 dollars per ounce of gold. With the establishment of the common agricultural policy, the EEC decided to use the unit of account to fix guaranteed prices. The end of the Bretton Woods system rendered this meaningless as gold lost its role in the international monetary system. The EEC's unit of account was thus reformed in 1975, and a new EUA was introduced, based on a basket of EEC currencies (Table 2). The EUA was first used for EEC budgets in 1978.

Several proposals were made in the 1970s to give a greater role to the European unit of account in the functioning of the EEC's exchange rate system (Mourlon-Druol

Table 2 Composition of the EUA in April 1975. (Source: Council Decision 75/250/EEC of 21 April 1975)

Currency	Amount of currency	Weight of currency (%)
Belgian franc	3.66	9.1
British pound sterling	0.0885	13.6
Dutch guilder	0.286	10.5
Danish krone	0.217	3
French franc	1.15	19.8
German mark	0.828	33
Irish pound	0.00759	1.1
Italian lira	109	9.5
Luxembourg franc	0.14	0.4

2012). Overall, these proposals aimed at rendering the European exchange rate more "symmetric," by spreading out more evenly the burden of adjustment between surplus and deficit countries. The idea was that the reference point would no longer be bilateral rates but would instead be to the EUA itself. This was meant to be a better indicator since the EUA was a multicurrency basket. The idea of a greater use of the EUA in the EEC exchange rate system never managed to reach a consensus among EEC members. At first sight, the introduction of the ECU in the EMS looked like a materialization of this line of thinking, but in practice the role of the ECU was limited.

The Debate Over the Creation of a Parallel Currency

Currency integration through the narrowing of exchange rates was the most famous strategy discussed among European policymakers and economists, but not the only one. The opportunity of introducing a "parallel" currency was also frequently debated, although primarily within financial and academic circles. The question of the parallel currency earned real fame in the 1980s when the UK government devised its alternative proposal to EMU based along these lines. But proposals for a parallel currency as a strategy for monetary integration in Europe had been advanced several times as early as in the 1970s.

Private banks and international institutions, rather than individual governments, have been precursors in advancing proposals for the creation of multicurrency units in Europe. Such units of account came into being against the backdrop of the international currency instability of the end of the Bretton Woods system. These units were developed as a way to hedge against increasingly volatile exchange rates, particularly within Europe, and in a transatlantic context. The EIB issued a bond in European Composite Unit, Eurco, in 1973 (Table 3). Before that, Sigmund Warburg was the first to lead such a financial innovation with the issuing of the first Eurobonds (Ferguson 2009).

1975 was the year when different initiatives supporting the introduction of a parallel currency in the EEC came out into the open (Mourlon-Druol 2012, 87–90). The most famous was the so-called All Saints' Day Manifesto signed by nine

Table 3 Composition of the Eurco in 1973

Currency	Amount of currency
German mark	0.9
French franc	1.2
British pound sterling	0.075
Italian lira	80
Dutch guilder	0.35
Belgian franc	4.5
Danish krone	0.2
Irish pound	0.005
Luxembourg franc	0.05

economists and published in *The Economist* on 1 November 1975 (The Economist 1975). Shortly after, another source of support for the idea of a parallel currency came from the series of OPTICA reports published in 1976 and 1977 (The Optica Report 1975, 1976). Both reports also advocated the introduction of a parallel currency and indeed partly involved the same academic economists.

The parallel currency approach was based on an entirely different set of assumptions than other approaches to European monetary integration. Advocates of the introduction of a parallel currency took for granted the failure of the Werner-style monetary integration strategy. They considered that monetary integration in Europe could only happen through market forces, rather than the narrowing of exchange rate fluctuations. The parallel currency, conceived as a basket of currencies with constant purchasing power, would gradually become more attractive than national monies and thus force the latter out of the market.

Much later, at the time of the Treaty of Maastricht negotiations, the UK government put forward a proposal for the introduction of a parallel currency in the EEC (Dyson and Featherstone 613–615). This was a counterproposal to the dominant option to create a single currency. But the British proposal arrived late, in the autumn of 1989, at a time when the Delors committee had already ruled out alternatives to the option of a single currency, and established momentum for EMU.

The option of introducing a parallel currency has resurfaced more recently in the context of the crisis of the euro area. The idea seems to have the same features as in the past: intellectually attractive, but practically impossible to implement (Cohen-Setton 2015).

European Exchange Rate Systems: Snake and EMS

Two types of exchange rate systems, very close to each other in their functioning, were in operation among most EEC member states and some non-EEC states from the early 1970s until the creation of the euro.

The Snake: 1972–1978

In August 1971, US President Richard M. Nixon unilaterally decided to suspend the convertibility of the US dollar into gold. Nixon declared that the USA would let the dollar freely float and announced that it was also imposing a 10% import surcharge in order to protect US-based products. The so-called closing of the gold window put a de facto end to the Bretton Woods system. After Nixon's decision to "close the gold window," the Group of Ten – that is, the group of countries (Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, the UK, and the USA) that agreed to participate in the IMF's General Agreements to Borrow (GAB) – adopted the Smithsonian Agreement in December 1971. The Smithsonian Agreement created bands of fluctuation of 2.25% related to each currency's central rate against the US dollar. This system created a "tunnel" in which European currencies could fluctuate. Further discussions took place among the Group of Ten in the following years, until the Jamaica Agreements signed in January 1976

eventually approved the end of the Bretton Woods system and allowed generalized floating. The transition to the post-Bretton Woods system between 1971 and 1976 represented an important moment for EEC currency cooperation as it forced the EEC to revise the basis on which its monetary relations were organized until then. The breakdown of the Bretton Woods system indeed put an end to the international monetary cocoon in which European currency relations had been so far developing.

Paradoxically, the end of the Bretton Woods system rendered regional monetary cooperation in Europe more urgent and more difficult (Ungerer 1997). European monetary cooperation became more urgent as the disappearance of the international monetary cocoon left Europeans with no functioning currency system. But European monetary cooperation became more difficult because the disappearance of the international monetary cocoon revived the disagreements about how to organize currency relations in a European setting. It is true that the Werner plan was still being implemented, but the Werner plan's implementation had been envisaged within the framework of a functioning international monetary system. As part of the implementation of the Werner plan, and as the first step on the way to the gradual narrowing of exchange rates in the EEC, the so-called European currency snake was adopted on 21 March 1972. The Council of Ministers agreed that fluctuation margins between EEC currencies should be limited to 2.25%.

But it became quickly clear that the snake was not an exchange rate system adapted to all EEC member states. Over some periods (see Table 4), the snake barely counted five EEC participants only: Germany, the Netherlands, Luxembourg, Belgium, and Denmark. The irony was also that many non-EEC members participated in what was meant to be an EEC exchange rate system (Norway, Sweden, Switzerland considered association). The majority of the biggest EEC member states, including two founding members, were out of the snake: Italy, France, and the UK. If the UK (in June 1972) and Italy (February 1973) were quickly forced out of the snake and never managed to get back in, the French government kept trying to rejoin the snake after it was forced out (in January 1974, rejoined in July 1975) but had to leave again (in March 1976). As an EEC-wide exchange rate mechanism, the snake thus did not appear well adapted to the different economic realities of each EEC member states.

Vivid discussions about this state of affairs ensued until the creation of the EMS and often revolved around the lines of the economist versus monetarist divide (see section "The Debate Between the 'Economists' and the 'Monetarists'"). Was it the snake that was ill-adapted to the economic reality of some EEC member states? Or was it some EEC member states that did not take the necessary steps to align their national economic policymaking with the constraints of the snake? The non-snake members argued the latter, while the snake members argued the former. Part of the explanation comes from the fact that the adjustment mechanism of the snake automatically fell on weaker currency countries. This is why weaker currency countries often advocated, as mentioned in sections "The European Unit of Account" and "The Debate Over the Creation of a Parallel Currency," the use of a composite currency unit – the EUA – as the reference point of the exchange rate system, so as to spread the burden more evenly among participants.

Table 4 Entries and exits from the snake and then ERM, 1972–1999

April 1972	Snake enters into force
May 1972	Denmark, Ireland, and the UK join the snake
	Norway associated with the snake
June 1972	Ireland and the UK, and then Denmark, leave the snake
October 1972	Denmark rejoins the snake
February 1973	Italy leaves the snake
March 1973	End of the tunnel. Sweden associated with the snake
January 1974	France leaves the snake
July 1975	France rejoins the snake
March 1976	France leaves the snake. End of the narrower margins among Benelux currencies
August 1977	Sweden leaves the snake
December 1978	Norway leaves the snake
March 1979	EMS enters into force
June 1989	Spain joins the ERM
October 1990	The UK joins the ERM
April 1992	Portugal joins the ERM
September 1992	Italy and the UK leave the ERM
August 1993	Agreement to widen the ERM band of fluctuation to 15%
January 1995	Austria joins the ERM
October 1996	Finland joins the ERM
November 1996	Italy rejoins the ERM

The European Monetary System: 1979–1992

After several months of negotiation, the EEC heads of state and government agreed on the EMS at the Brussels European Council in December 1978 (Mourlon-Druol 2012). The UK government had made plain that it did not intend to participate in the Exchange Rate Mechanism (ERM) but that it would formally belong to the EMS. The Irish and Italian governments reserved their decision, as they were not sure to be able to participate without further economic support. After additional negotiations involving bilateral loans and the setting of preferential interest rates for European Investment Bank's (EIB) loans, both governments agreed to participate in the ERM. The EMS entered into force in March 1979.

The EMS negotiations essentially centered on the role to be given to the ECU in the exchange rate system. As mentioned above, weaker currency countries wished that the EUA should be given a greater role of anchor of the system, while stronger currency countries (snake members) did not. The view of the latter prevailed, but the EUA was formally given centrality in the system. Part of this centrality came from a rechristening of the unit of account into European Currency Unit, which pleased the French government in particular as the acronym was reminiscent of the name of an old French currency and provided more visibility to the unit of account.

It must however be noted that the ECU acronym has been in use since the early 1970s, both among European institutions and private companies. In a 1974 speech, President of the European Commission François-Xavier Ortoli was the first official to draw the conscious parallel between the acronym and the old French currency. Second, the ECU was used to devise a bilateral grid of parities (so the ECU was not used as direct reference point as the weaker currency countries hoped). Third, the EMS introduced a so-called divergence indicator, in order to spread the burden of adjustment more fairly between weak and strong currency countries. The aim of that indicator was to identify the currency that was actually diverging from the ECU, whether upward and downward. The divergence indicator remained however toothless. Indeed, the new ability to pinpoint at the diverging currency was not accompanied by any automatic measure of adjustment. Only consultations among central banks were anticipated. Finally, after 1979, the use of the ECU in private markets developed and supported the discussions about the possibility of transforming it into a parallel currency.

In sum, the EMS was not very different from the snake. The value of the ECU introduced in 1979 was the same as that of the EUA in 1975 (see Table 2). The marginal modifications mentioned above, as well as an increase in the support facilities available, made it look like a new mechanism. What made a real difference was the new economic policy consensus that was sustaining the EMS, especially between the French and German government that now both aimed at fighting inflation and follow stability-oriented economic policies (McNamara 1998).

In spite of the lack of important novelty in the EMS, the first years of functioning of the new EEC exchange rate system appeared successful, in the sense that no major realignment occurred. The new expansionary economic policy that the French government implemented after the election of François Mitterrand in 1981 soon threatened the functioning of the EMS, as the French franc's participation was uncertain. But when the French government decided to reverse the course of its economic policy in March 1983, the EMS was eventually reinforced (Duchaussoy 2011). Faced with a choice between continued participation in the EMS, and its expansionary economic policy, the French government chose the former, which de facto reinforced the EMS as a tool for European monetary integration.

This reinforcement of the credibility of the EMS did not remove the inherent problems of the European exchange rate system. In the first years of functioning of the EMS, European policymakers and central bankers in particular discussed the possible improvements that could be brought to the ERM. The so-called Basle-Nyborg agreement concluded on 12 September 1987 materialized progress on a long-term debate about the use of intramarginal interventions in the EMS (James 2012). Such interventions were aimed at anticipating that a currency reached its fluctuation limit but were not included in the EMS agreement. The Basle-Nyborg agreement formalized that the Monetary Committee and the Committee of Governors would monitor policy inconsistencies among EMS members using a set of indicators and projections (surveillance), as well as officialized the possible use of intramarginal interventions.

Plans for the Creation of a Single Currency in Europe

Two famous reports set out in detail a plan to introduce a single currency in Europe: the Werner plan in 1970 and the Delors report in 1989 (James 2012; Marsh 2009). Both reports are often compared with each other, since Jacques Delors himself claimed that his report was consciously inspired from the earlier Werner plan.

The Werner Report

The EEC heads of government meeting at The Hague Summit in 1969 decided to prioritize the creation of an EMU and tasked the Luxembourg prime minister Pierre Werner to chair a group that would devise a roadmap to achieve this goal (Danescu and Muñoz 2015). The so-called Werner committee (Table 5), composed of the presidents of the relevant EEC committees, met 14 times between March and October 1970 and released its final report on 8 October 1970. The work of the Werner committee epitomized the opposition between the "economist" and "monetarist" schools of thought (see section "The Debate Between the 'Economists' and the 'Monetarists'"), and the final report reflected a compromise between them (also called the "parallelist approach"). The Werner report set out three stages for the completion of EMU, but only the first stage was described in detail. The report did not set out how the political-institutional dimension should be implemented.

By adopting a resolution on "the achievement by stages of EMU" based on the Werner report, the Council of Ministers officially set in motion the process (Council 1971). As mentioned above, this processed famously materialized in the creation of the European currency snake in March 1972 and in the creation of the EMCF in April 1973. But the difficulties linked to the end of the Bretton Woods system revived tensions about how to reach EMU and eventually led to the unofficial abandonment of the implementation of the Werner plan in early 1974.

Table 5 Members of the Werner committee

Name	Capacity
Hubert Ansiaux	Chairman of the Committee of Governors and Governor of the National Bank of Belgium
Gerard Brouwers	Chairman of the Conjunctural Policy Committee and State Secretary in the Dutch Ministry of the Economy
Bernard Clappier	Chairman of the Monetary Committee and Deputy Governor of the Banque de France
Georges Morelli	Coordinator of the Group's secretariat and Commission official
Ugo Mosca	Director-General for Economic Affairs, DG II, European Commission
Johann Baptist Schollhorn	Chairman of the Medium-Term Economic Policy Committee and State Secretary in the Federal Ministry of the Economy
Gaetano Stammati	Chairman of the Budgetary Committee and Treasurer-General in the Italian Ministry of the Treasury
Pierre Werner	Chairman of the Group. Prime minister and finance minister of Luxembourg

The Delors Report

Meeting in Hanover in 27–28 June 1988, the EEC heads of government decided to revive the debate about creating an EMU. The European Council tasked the commission's president Jacques Delors with the preparation of a roadmap to achieve EMU (James 2012). Unlike the Werner committee, the so-called Delors committee (Table 6) was composed of all EEC central bankers, plus a small number of additional economic experts. Delors drew what he perceived as the lesson of the Werner plan's failure: the noninvolvement from the start of those officials – the central bankers – who were less keen to advance monetary integration. Further to this, Delors was careful to involve the central bankers in their personal capacity, not as representatives of their institution. Finally, the group's work was based on consensus, confidentiality, and mutual trust, which allowed the final report to propose the creation of an EMU, in spite of the presence of central bankers who could originally be assumed to be staunch opponents to it (in particular governor of the Bank of England Robert Leigh-Pemberton and president of the Bundesbank Karl-Otto Pöhl).

The final report, released in April 1989, bore some resemblance with the Werner report, in that it recommended the creation of an EMU in three stages. But the Delors report was different in substance, at least in two respects. First, the report did not set out a clear timetable for Stages 2 and 3. Too strict a timetable for completion was interpreted as one of the reasons for the failure of the Werner plan, and the committee wished to avoid falling into that trap. Second, the report attributed less space to the economic dimension of EMU and in particular the social aspects. The members of

Table 6 Members of the Delors committee

Name	Capacity
Frans Andriessen	Commissioner in charge of agriculture
Miguel Boyer	Independent expert, former Spanish finance minister
Demetrios J. Chalikias	Governor of the Bank of Greece
Carlo Azeglio Ciampi	Governor of the Bank of Italy
Maurice F. Doyle	Governor of the Central Bank of Ireland
Willem F. Duisenberg	President of the Central Bank of the Netherlands
Jean Godeaux	Governor of the National Bank of Belgium
Erik Hoffmeyer	Governor of the Central Bank of Denmark
Pierre Jaans	Governor of the Monetary Institute of Luxembourg
Alexandre Lamfalussy	General manager of the Bank for International Settlements
Jacques de Larosière	Governor of the Bank of France
Robert Leigh-Pemberton	Governor of the Bank of England
Karl Otto Pöhl	President of the Bundesbank
Mariano Rubio	Governor of the Central Bank of Spain
José Alberto Tavares Moreira	Governor of the Central Bank of Portugal
Niels Thygesen	Danish economics professor
Gunter Baer	Rapporteur
Tommaso Padoa-Schioppa	Rapporteur

the committee tackled some of these broader issues but considered that they were not belonging to their competence remit.

The Creation of the Euro

If the Delors report set out in detail how a European single currency would be created, a political decision still needed to be agreed upon so that the single currency could eventually be adopted. This section first looks at the making of the Treaty of Maastricht, including the Intergovernmental Conference (IGC), the convergence criteria, the stages for the creation of the euro, and the German reunification and then moves on to the crises of the EMS in 1992–1993 that threatened the EMU process.

The Intergovernmental Conference, the Treaty of Maastricht, and German Reunification

The EEC heads of government endorsed the Delors report, but the creation of an EMU required a change in the treaties and thus the convening of an intergovernmental conference (IGC). The Strasbourg European Council of December 1989 agreed to convene the IGC, which began its works in December 1990 (Dyson and Featherstone 1999; James 2012). The Treaty of Maastricht was signed in February 1992 and entered into force on 1 January 1993.

The Treaty of Maastricht set out four convergence criteria to join the euro, related to price stability, government finances, exchange rate stability, and convergence of interest rates (Protocol No.13 2010). "Price stability" meant that for a period of 1 year before the examination of the situation of the candidate country, the inflation rate should not exceed 1.5% above that of the three best performing member countries. "Government finances" is the most famous part of the criteria and entails the provisions that deficits should not exceed 3% of GDP and debt 60% of GDP. "Exchange rate stability" refers to the respect of the ERM's fluctuation margins for the past 2 years prior to the examination of the candidate country (with no devaluation). Finally, "convergence of interest rates" means that "observed over a period of 1 year before the examination, a member state has had an average nominal long-term interest rate that does not exceed by more than two percentage points that of, at most, the three best performing member states in terms of price stability."

Building on the conclusions of the Delors report, the Treaty of Maastricht envisaged three stages for the creation of EMU. *Stage 1* (from 1 July 1990 to 31 December 1993) was mostly preparatory and included the improvement of economic convergence, the increase of cooperation among central banks, the completion of the freedom of movement of capital, and the free use of the ECU. *Stage 2* (from 1 January 1994 to December 1998) involved more concrete institutional steps, including the creation of the European Monetary Institute and the granting of independence to national central banks. *Stage 3* (from 1 January 1999) represented

the final steps toward the creation of the single currency and included the creation of the European System of Central Banks and the European Central Bank to conduct a single monetary policy, the irrevocable fixing of exchange rates, the entry into force of the Stability and Growth Pact, and of course the introduction of the euro banknotes and coins.

The fall of the Berlin wall and the German reunification process encouraged the French government to seek a firmer timetable for the creation of EMU (Dyson and Featherstone 1999). It certainly did not initiate, however, the debate about the advantages and disadvantages of creating a single currency in the EEC. In the autumn of 1989, the process was already well in motion, and the issue was on the agenda for many years, if not decades. Convening the IGC and agreeing on a clear timetable for EMU however allowed to bind a reunified Germany to the EEC.

The Crisis of the ERM: 1992–1993

Just as EMU seemed on track, two severe currency crises hit the EMS in 1992 and 1993 (James 2012). Both originated in currency speculation and threatened the very existence of the European exchange rate system. They also laid bare the power of increased financial flows in a Europe where capital movements were now fully liberalized.

Currency speculation arose on fertile ground. No currency realignment occurred in the EMS between 1987 and 1992, while at the same time, a number of competitiveness indicators had begun to diverge significantly (Feiertag 2016). What served as a trigger for the speculative attacks was the difficult ratification process of the Treaty of Maastricht introducing the European single currency. In June 1992, the Danish electorate voted against the ratification of the Treaty of Maastricht. The result of the French referendum, to be held in September, was uncertain. Speculation mounted over the summer, in particular, against the British pound sterling and the Italian lira. On 16 September 1992, retrospectively called "Black Wednesday," speculators attacked the pound sterling. The level at which the pound joined the ERM in October 1990 had given birth to much debate. Given the volumes of financial transactions involved, the Bank of England could not counter the attacks, and the pound left the ERM.

Inauspicious economic figures in France in July 1993 gave birth to renewed currency speculation, this time against the French franc, and also the Belgian franc, the Spanish peseta, the Portuguese escudo, and the German mark. Faced with the inability to maintain the regular EMS parities due to the pressure of speculation, the Monetary Committee agreed to widen the margins of fluctuation of the ERM from +/−2.25% (or 6% for some members) to +/− 15%. These new bands allowed to render speculation ineffective as they were too large to be targeted but also showed that the EMS was effectively abandoned. As Karl-Otto Pöhl indeed anticipated in the discussions of the Delors committee, the EMS crises highlighted that in a European and global context of free capital flows, the EEC faced a binary choice: either opt for full floating or adopt a single currency. In the end, the EMS crises of the early 1990s thus reinforced European policymakers' determination to implement the latter.

Conclusions: European Monetary Integration After Maastricht

The monetary integration of Europe did not stop at Maastricht. If the Treaty signed in 1992 set out how the single currency would be created, the following three decades actually witnessed its inception in practice and revealed the strengths and weaknesses of the Maastricht construct.

The first 10 years of the euro proved successful on many counts. The introduction of a new single currency among eleven countries occurred smoothly, which was no small technical feat. On the whole, economic indicators tended to be positive, and even if this could not automatically be attributed to the euro, it contributed to create a sense that the single currency was a success and that the pessimistic forecasts that some economists had made before its creation were wrong.

Behind these positive economic indicators, some underlying weaknesses were being confirmed or emerging. The introduction of the euro was associated with an immediate adjustment of prices that gave birth to a lasting memory of sudden inflation caused by European monetary unification. The introduction of the euro supported a movement of capital from Northern European countries to Southern European countries. This movement could have allowed an economic adjustment between North and South. Instead, it proved to be nurturing some bubbles and did not lead Southern European countries to undertake some needed, and promised, structural economic reforms.

The outbreak of the so-called Eurozone crisis in 2008 and 2009 made plain the original weaknesses of the euro area (Mourlon-Druol 2014). The so-called Eurozone crisis started against the backdrop of the global financial crisis that severely troubled the international financial system. The crisis started in late 2009, when the new Greek government led by George Papandreou announced substantially revised figures for its deficit, from an estimation of 6–8% to 15.7%. Set aside the question of the failure to provide accurate figures, this revision casted doubt on the Greek government's ability to finance its deficit. This launched a series of bailout negotiations and spread fear about the ability of other countries – Portugal, Spain, Italy, and Ireland – to service their debt and deficit (Sandbu 2015).

The Eurozone crisis confirmed that the process of European monetary integration did not stop with the Treaty of Maastricht. In terms of monetary policymaking, significant policy changes have been put in place since the outbreak of the Eurozone crisis (Claeys 2017). The ECB developed new monetary policy operations, including the Securities Market Programme (SMP) in 2010 in order to buy sovereign bonds on secondary markets, the Long-Term Refinancing Operations in 2011 to support bank lending and liquidity, the Outright Monetary Transactions (OMT) in 2012 which was similar to SMP but with no ex ante quantitative limit, and the Public Sector Purchase Programme (PSPP), better known as "quantitative easing," in 2015. The ECB also developed a new role as supranational banking supervisor within the framework of the European banking union.

Finally, since 1999, the euro has enlarged several times (Table 7). Due to the Eurozone crisis, most attention and debate focused on the possibility of a Eurozone member departing the euro area. But in reality no single Eurozone member has left

Table 7 Enlargements of the Eurozone

1999	Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain
2001	Greece
2007	Slovenia
2008	Cyprus, Malta
2009	Slovakia
2011	Estonia
2014	Latvia
2015	Lithuania

yet, at least at the time of writing, and instead eight have joined. This certainly is one of the paradoxes of European monetary integration: it is heavily commented, challenged, and criticized; but it retains a strong power of attraction.

Cross-References

- ► International Monetary Regimes: The Bretton Woods System
- ► The Evolution of Monetary Policy (Goals and Targets) in Western Europe
- ► The Historical Evolution of Central Banking

References

'The All Saints' Day Manifesto for European Monetary Union'. The Economist, 1 November 1975

Claeys G (2017) The missing pieces of the euro architecture. Bruegel Policy Contribution No. 28, October 2017

Cohen-Setton J (2015) The economics of parallel currencies. Bruegel blog post. http://bruegel.org/2015/06/the-economics-of-parallel-currencies/. Accessed 31 July 2018

Commission of the European Communities, Directorate-General for Economic and Financial Affairs, One market, one money. An evaluation of the potential benefits and costs of forming an economic and monetary union, n° 44, October 1990

Cooper R (2006) Almost a Century of Central Bank Cooperation. BIS working papers, n° 198, February 2006

Council of Ministers, Resolution on the achievement by stages of economic and monetary union in the Community, 22 March 1971

Council Decision 75/250/EEC of 21 April 1975 on the definition and conversion of the European unit of account used for expressing the amounts of aid mentioned in Article 42 of the ACP-EEC convention of Lomé

Danescu E, Muñoz S (eds) (2015) Pierre Werner et l'Europe : pensée, action, enseignements/Pierre Werner and Europe : his approach, action, and legacy. Peter Lang, Brussels

de Grauwe P (2013) Design failures in the Eurozone: can they be fixed? European Economy, Economic papers 491, April 2013

Duchaussoy V (2011) La Banque de France et l'Etat, de Giscard a Mitterrand: enjeux de pouvoir ou resurgence du mur d'argent? 1978–1984. L'Harmattan, Paris

Duisenberg W (2002) Acceptance speech of the International Charlemagne Prize for 2002, 9 May 2002. Available at https://www.ecb.europa.eu/press/key/date/2002/html/sp020509.en.html
Dyson K, Featherstone K (1999) The road to Maastricht: negotiating economic and Monetary Union. Oxford University Press, Oxford
Dyson K, Maes I (eds) (2016) Architects of the Euro: intellectuals in the making of European Monetary Union. Oxford University Press, Oxford
Eichengreen B (1993) Reconstructing Europe's trade and payments: the European Payments Union. Manchester University Press, Manchester
Eichengreen B (ed) (1995) Europe's postwar recovery. Cambridge University Press, Cambridge
Feiertag O (2016) Les banques centrales et les États dans la crise de change de 1992–1993: le baptême de la monnaie unique. In: Feiertag O, Margairaz M (eds) Les banques centrales et l'État-nation/The central banks and the nation state. Les Presses de Sciences Po, Paris, pp 617–644
Ferguson N (2009) Siegmund Warburg, the City of London and the financial roots of European integration. Bus Hist 51(3):364–382
Fleming M (1971) On exchange rate unification. Financ Anal J 20(1):1102–1103
Grygowski D (2009) Les États-Unis et l'unification monétaire de l'Europe. Peter Lang, Brussels
Haberler G (1970) The international monetary system: some developments and discussions. In: Halm GN (ed) Approaches to greater flexibility of exchange rates. Princeton University Press, Princeton, pp 115–123
Ingram J (1969) Comment: the optimum currency problem. In: Mundell R, Swoboda A (eds) Monetary problems of the international economy. Chicago University Press, Chicago, pp 95–100
James H (2012) Making the European Monetary Union: the role of the Committee of Central Bank Governors and the origins of the European Central Bank. Harvard University Press, Cambridge, MA
Kenen P (1969) The theory of optimum currency areas: an eclectic view. In: Mundell R, Swoboda A (eds) Monetary problems of the international economy. Chicago University Press, Chicago, pp 41–60
Kindleberger CP (1986) International public goods without international government. Am Econ Rev 76(1):1–13
Knapp GF (1924) The state theory of money. Macmillan, London
Marsh D (2009) The Euro: the politics of the new global currency. Yale University Press, New Haven
McKinnon R (1963) Optimal currency areas. Am Econ Rev 53:717–724
McNamara KR (1998) The currency of ideas: monetary politics in the European Union. Cornell University Press, Ithaca
Mourlon-Druol E (2010) Filling the EEC leadership vacuum? The creation of the European Council in 1974. Cold War Hist 10(3):315–339
Mourlon-Druol E (2012) A Europe made of money: the emergence of the European monetary system. Cornell University Press, Ithaca
Mourlon-Druol E (2014) Don't blame the Euro: historical reflections on the roots of the Eurocrisis. West Eur Polit 37:1282–1296
Mourlon-Druol E (2016) Steering Europe: explaining the rise of the European Council, 1975–1986. Contemp Eur Hist 25(3):409–437
Mourlon-Druol E (2018) History of economic and monetary union. In: Amtenbrink F, Hermann C (eds) The EU law of economic and monetary union. Oxford University Press, Oxford
Mundell R (1961) A theory of optimum currency areas. Am Econ Rev 51(4):657–665
Padoa-Schioppa T (1994) The road to monetary union in Europe: the emperor, the kings, and the genies. Oxford University Press, Oxford
Protocol No. 13 on the convergence criteria, Official Journal of the European Union, C83, 30 March 2010
Rose A (1999) One Money, One Market: Estimating the Effect of Common Currencies on Trade, NBER Working Paper No. 7432
Sandbu M (2015) Europe's orphan: the future of the Euro and the politics of debt. Princeton University Press, Princeton

The Optica Report 1975 (1976) Towards economic equilibrium and monetary unification in Europe. Commission of the European Communities, Brussels

The Optica report 1976 (1977) Inflation and exchange rates: evidence and policy guidelines for the European Community. Commission of the European Communities, Brussels

Ungerer H (1997) A concise history of European monetary integration, from EPU to EMU. Quorum Books, Westport

Werner P (1970) Report to the Council and the Commission on the realisation by stages of economic and monetary union in the Community. Available at http://ec.europa.eu/economy_finance/emu_history/documentation/chapter5/19701008en72realisationbystage.pdf

Part IX

Central Banking and Monetary Policy

The Historical Evolution of Central Banking 33

Stefano Ugolini

Contents

Abstract

"Central banking" is what a central bank does, but the definition of "central bank" is less straightforward than it may appear at first sight. Following Ugolini (The evolution of central banking: theory and history. Palgrave Macmillan, London, 2017), this chapter defines central banking as the provision of public policies aimed at fostering monetary and financial stability and surveys the historical

S. Ugolini (✉)
Sciences Po Toulouse and LEREPS, University of Toulouse, Toulouse, France
e-mail: stefano.ugolini@ut-capitole.fr

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_31

evolution of such policies in the West from the Middle Ages to today. It shows that institutional equilibria mattered a lot in shaping the way stabilization policies were implemented: central banking evolved in markedly distinct ways in city states (like Venice, Amsterdam, Hamburg, Barcelona, or Genoa), centralized territorial polities (like Naples, Sweden, England, Austria, or France), or decentralized territorial polities (like the United States or the European Union). As a result, the historical evolution of central banking does not appear to have been driven by the "survival of the fittest," but rather by the constant adaptation of policymaking to changing political economy equilibria.

Keywords
Central banking · Monetary institutions · Public policy · Political economy

Introduction

What is *central banking*? Obviously, central banking is what a central bank does. But what is a *central bank*, and what does it do? Here, the answer is less straightforward. Today, the organizations falling under the label of "central banks" are not all alike and make a bunch of quite different things. Things get even more complicated if we move backward in times: many of nowadays' central banks did not look like central banks when they were created, while other organizations that appeared to behave somewhat like central banks never evolved into proper central banks. The wording "central bank" started to be used to indicate the Bank of England only in the second half of the nineteenth century; it had already been used before, but to indicate a different thing – i.e., the headquarters of a multi-branch commercial bank, not a privileged bank of issue.

In order to find a way out of this ontological conundrum, this chapter follows the strategy proposed by Ugolini (2017) and defines central banking as *a family of public policies aimed at fostering monetary and financial stability*, whose provision is nowadays generally (albeit not necessarily) performed by those organizations that we call central banks. This strategy allows tracking the evolution of central banking to a period well preceding the appearance of modern central banks. As a result, this chapter covers a much longer and broader spectrum than most available general surveys like, e.g., Goodhart (1988), Capie et al. (1994), Giannini (2011), or Bordo and Siklos (2018). It focuses on 12 case studies that have been identified by previous scholarly research as relevant episodes in the evolution of central banking, and on which more abundant historical evidence is actually available. These include (in chronological order) Venice (thirteenth to eighteenth centuries), Barcelona (fifteenth to nineteenth), Genoa (fifteenth to eighteenth), Naples (sixteenth to eighteenth), Amsterdam (seventeenth to eighteenth), Hamburg (seventeenth to nineteenth), Sweden (seventeenth to nineteenth), England (seventeenth to twentieth), Austria (eighteenth to nineteenth), France (eighteenth to twentieth), the United States (eighteenth to twentieth), and the European Union (twentieth to twenty-first).

If central banking can legitimately be seen as public policy, then central banking cannot be analyzed separately from the political context in which it is implemented. The novelty of this survey consists of organizing historical evidence on the evolution of central banking in systematic relation to the institutional equilibria in which this evolution took place. The idea is to show that institutional frameworks were crucial in shaping the way the quest for monetary and financial stability was organized over time. The rest is structured as follows. After clarifying the adopted definition of central banking (section "Conceptual Framework"), the chapter reviews the development of central banking in city states (section "Central Banking in City States"), in centralized territorial polities (section "Central Banking in Centralized Territorial Polities"), as well as in decentralized territorial polities (section "Central Banking in Decentralized Territorial Polities"). Some concluding remarks are drawn in the end (section "Conclusions").

Conceptual Framework

Ugolini (2017) proposes to see central banking as the joint provision of a set of public policies aimed at addressing market failures in the financial sector. Actually, there exist at least four rationales for public intervention in this domain: the first two can be associated to contemporary central banks' *financial stability* mandate, while the last two can be associated to their *monetary stability* mandate.

First, public intervention can be justified by the need of securing the efficiency of the "financial plumbing" of the economy, i.e., of regulating the *management of the payment system* (the system allowing for the transfer of funds from one point to any other point of an economy). As any other network infrastructure, the payment system is a natural monopoly: the supply of payment services is subject to scale economies, while their demand is subject to network externalities. In order to prevent monopolists from capturing rents, natural monopolies need to be regulated by public authorities. Regulation is also needed in view of the "essential" nature of the payment infrastructure, whose disruption can entail huge costs to real economic activity.

Second, the banking sector is inherently unstable in view of the extreme maturity and quality transformation business performed by banks. Banks are intermediaries funding their activity through demandable liabilities (deposits) and reinvesting into highly idiosyncratic long-term assets (loans): by construction, they are therefore prone to liquidity crises. An organization that is centrally situated within the payment system can (by virtue of its privileged position) create liquidity by lending to counterparties and thus act as a *lender of last resort* in the event of liquidity crises. In view of its monopoly of crisis-time lending, such an organization enjoys a nonnegligible market power and is therefore able to impose a number of conditions (*regulatory standards*) aiming at reducing risk-taking by counterparties.

Third, there exists a complementarity between the public sector's and the private sector's cash flows: the former is characteristically cash-strapped when the latter is cash-rich (after the administration's purchases of goods and services are paid for and

before taxes are collected) and vice versa. This complementarity naturally calls for the reciprocal smoothing of cash flows: the public sector can purchase on credit goods and services from the private sector, only to resorb such debt as soon as the latter will pay taxes. Because of its widespread acceptability, short-term debt eligible for tax payments naturally has a tendency to be used as a means of payment in transaction among third parties – or differently said, to be monetized. *Debt monetization* is a function that can well be provided by private intermediaries, who can advance funds to both public and private sector on the one hand while issuing monetary instruments to the private sector on the other hand. However, a disruption in the provision of this function by private intermediaries can entail substantial costs, as it can potentially jeopardize the public and private sectors' ability to conduct their regular activities. Hence, there exist good reasons for political authorities to secure the continuity of public and/or private debt monetization under any financial condition.

Fourth, public *and private* debt monetization by private intermediaries may not necessarily be optimal, as it can lead to excessive price volatility – which is obnoxious to real economic activity. In order to minimize such volatility, public authorities can intervene to stabilize the value of monetary instruments both by setting limits to private money creation and by compensating for shortfalls in private money creation. This stabilization function is commonly known under the heading of *monetary policy.*

A priori, each one of these four functions might be conceivably provided separately from the others. In reality, however, this was seldom the case, as spillovers exist between them. As a matter of fact, (1) a banking organization issuing demandable liabilities (deposits or banknotes) that is centrally situated within an economy (for structural or juridical reasons) will naturally assume a leading role in the payment system; (2) the need to protect itself from counterparty risk will encourage the organization to take some regulatory role, which will be enhanced by its market power; (3) its lending activities will de facto consist of (public and/or private) debt monetization; and (4) the management of such lending activities will impact the value of the demandable liabilities issued by the organization and thus de facto consist of monetary policymaking. In what follows, the historical development of mechanisms linking the issuance of credit money to lending operations will therefore be reviewed, with a particular focus on their implications in terms of public policy.

Central Banking in City States

Central Banking in City States: Issues

City states have been a pervasive presence in European geopolitics from the Middle Ages to the Napoleonic Wars. European city states typically originated from self-governing urban institutions, which gained increasing autonomy from feudal rulers in the central centuries of the Middle Ages. Among the many hundreds of such polities, the most famous ones were the so-called merchant republics – i.e., city

states ruled by oligarchies of businessmen which had succeeded in becoming leading international trade centers (Lindemann 2015). In view of their relatively high degree of financial development, it is not surprising that the earliest attempts at elaborating central banking policies actually appeared in these polities.

Merchant republics featured an institutional equilibrium that allowed for practicing central banking directly through a branch of government. The monetary authorities of the merchant republics interacted with the same kinds of counterparties on both sides of their balance sheet: on the liabilities side, they predominantly borrowed from domestic businessmen (who were the ones that held and used the monetary instruments they issued), while on the assets side, they predominantly lent to a government that was tightly controlled by the very same businessmen. This perfect alignment of interests (Stasavage 2012) guaranteed the time consistency of monetary policy: public debt monetization could have inflationary outcomes only as long as these were justified by a superior goal (in general, militarily defending the commercial interests of the republic), and the ensuing inflation was expected to be compensated by deflation as soon as conditions came back to normality. This explains why, in these polities, central banking policies were implemented directly through a division of government or a state-owned public bank.

If European merchant republics presented many common features, each one had its own peculiar institutional equilibrium. The example that was closest to the "ideal type" of the merchant republic was Venice, a city state that had secured its full independence from territorial rulers since the early Middle Ages and that was run for centuries by a cohesive elite of businessmen. Amsterdam and Hamburg also were largely autonomous cities controlled by stable merchant oligarchies, but they were not fully sovereign states (they were formally part of, respectively, the United Provinces and the Holy Roman Empire). Barcelona long fought to retain its autonomy with respect to the Crown of Aragon, whose interference in local affairs gradually increased over the centuries. By contrast, Genoa was a de facto independent republic which was nonetheless characterized by harsh intestine infight within the ruling oligarchy (mostly consisting of feudal rather than bourgeois families). As the rest of this section will show, these institutional dissimilarities go a long way in explaining the different solutions found to the problems experienced by domestic financial infrastructures.

Venice

Under many respects, the history of central banking in the West starts in late-medieval Venice. This should not be understood as meaning that Venetians "invented" central banking. Rather, it should be understood as meaning that Venetians were the first to experience on a large scale the problems to which central banking is supposed to be a solution, and were thus forced to find some tentative solutions earlier than elsewhere. The reason why this was the case was the early sophistication of Venice's payment infrastructure. In view of its unique geographical situation, the city's economy precociously experienced a high

degree of monetization: as a result, by the late Middle Ages the deposit banking sector had already developed to an extent that was long unknown elsewhere. Already in the early fourteenth century, bank transfers were widely used by the population in payment of up to relatively small sums, and were considered by jurisdiction as legal means to discharge debt (Mueller 1997). Such a pervasive role of bank transfers was very exceptional: as a matter of comparison, in another major late-medieval financial center like Florence (home to the most important international investment banks of the time), "dematerialized" payments played a significant role in intragroup transactions, but pure market transactions had to be cleared in cash (Padgett and McLean 2011). This means that Venice was probably the first place in the West to establish a truly modern payment system. This primacy did not come without strings attached, though: the price to pay for economizing on cash on a large scale was the inherent instability of fractional-reserve banking (Diamond and Dybvig 1983). And indeed, the history of Venetian banking was punctuated by a succession of violent liquidity crises jeopardizing the continuous provision of payment services.

Venetian political authorities were particularly concerned with potential disruptions in the domestic financial infrastructure. Besides entailing large negative effects on economic activity, such disruptions compromised the functioning of the "machine of the state" by preventing the government from smoothing its irregular cash flows. This was a particularly sensitive issue in Venice, where the sustenance of the metropolitan population strictly depended on the Republic's victualling activities: the government's inability to secure the provision of flour to the city at stable prices would have triggered social unrest. At the same time, Venetian authorities were also concerned with keeping – as much as possible – an arm-length distance from the operation of the financial infrastructure. As a result, government intervention occurred only when private initiative was lacking. In the six centuries that separated the rise of its Mediterranean colonial empire from the fall of the Republic, such a situation occurred twice: first, in the early phases of consolidation of Venetian institutions following the colonial expansion (thirteenth to fourteenth centuries) and then at the time of their reorganization after the decline of its geopolitical power (seventeenth to eighteenth centuries).

Starting from 1282, the Venetian state centralized most of its financial activities to the most important of its public victualling agencies, the Grain Office (Camera del Frumento). Besides its traditional mission to buy grain from importers and resell flour to the metropolitan population, the agency was now charged with the issuance of the floating debt on the security of streams of fiscal revenues. In addition, the Office was also allowed to collect deposits from the general public, both voluntary and forced (some types of funds, like dowries, had to be compulsorily deposited with it). The centralization of all these financial activities transformed the Grain Office into a sort of "state bank." The agency implemented the monetization of the public debt by borrowing from the general public on the security of future tax revenues. The deposit liabilities of the Office established themselves as the standard means of payment used both by merchants for big commercial transactions and by savers for petty day-by-day transactions (Mueller 1997).

Venice's early "state bank" worked rather successfully until the mid-fourteenth century, when it started to get into troubles because of the bad performance of a number of subsidized loans extended to the private sector according to political criteria. Depositors' trust decreased as the Grain Office was forced to postpone the payment of interests on savings accounts, and the reform of 1365 actually terminated the "state bank" model. In the meantime, legislators had provided the status of legal money to deposits with chartered banks, and the private deposit banking sector had actually started to thrive. As a result, the government found it expedient to step back from direct intervention in the financial system and externalized the tasks previously accomplished by the Grain Office (provision of payment services and public debt monetization) to the chartered private banks operating on the Rialto. In so doing, the government discharged onto the private sector all responsibilities for the occurrence of payment accidents as the ones that had occurred to the Grain Office (which had raised widespread criticism), but these benefits came at the cost of increased financial instability. As a matter of fact, private deposit banks fell victims to a long series of liquidity crises between the fourteenth and the sixteenth century, thus jeopardizing the working of the payment system and the government's ability to borrow. In view of the high costs of such disruptions for the economy as a whole, Venetian authorities were obliged to intervene with the aim of reducing instability. Over the decades, they developed a wide range of regulatory tools that were not unlike those in force nowadays. These included ex ante interventions like the establishment of legal restrictions to operations, of specific supervisory bodies, and of disclosure requirements, but also ex post interventions like lending of last resort (Ugolini 2017, pp. 120–131). As a result, despite its willingness to stay at arm's length from the financial sector, the government was actually obliged to provide some central banking functions during this period. Petitions for the creation of a public bank were repeatedly voiced in the aftermath of crises, but they were systematically rejected as long as an alternative was viable (Mueller 1997).

When the last chartered bank fell victim to the umpteenth depositors' run, however, the government had no other choice than taking back responsibility for the operation of the domestic financial infrastructure. After lengthy discussions, in 1587, the creation of a public bank (called Banco della Piazza di Rialto) was finally approved. This was not a state bank, but a public concession: a monopoly of deposit banking was created, but its management was delegated to a private concessioner under his full personal liability. Deposits were made attractive to the public by unpegging their value from the one of circulating specie, thus sheltering them from the general degradation the latter was experiencing at the time (Roberds and Velde 2016). Moreover, a forced demand for deposits was created by making bills of exchange compulsorily payable in bank. This was meant not only as a subsidy to the bank but also as a supervisory device aimed at preserving the quality of circulating credit instruments (Ugolini 2017, pp. 55–61). These measures were apparently effective in securing the success of the Banco della Piazza (Luzzatto 1934; Tucci 1991).

Geopolitical factors, however, soon prompted new unintended changes to the design of Venetian monetary institutions. In 1619, the government faced difficulties in promptly repaying in specie purveyors that had delivered silver bullion to the Mint, and upon their request, it made credits with the government transferable to

third parties (a practice commonly adopted three centuries earlier by the Grain Office). Increasing military spending during the 1620s made the practice widespread and thus transformed a provisional device into a stable one. Known as the Banco del Giro, the new mechanism (actually a division of the government) provided for the monetization of the public debt through the issuance of inconvertible money, which could be used (as an alternative to the one issued by the Banco della Piazza) for the payment of bills of exchange. Thus, the Banco del Giro was probably the first example of a purely fiduciary state-issued legal-tender money. Despite their inconvertibility, deposits with the new bank soon established themselves as the standard means of payment (also in view of the state's big role in the domestic economy) and thus outcompeted the payment business of the Banco della Piazza, which was eventually closed down in 1638. The Banco del Giro thus became the new "central bank" of Venice and operated rather successfully until the fall of the Republic, when it was liquidated by the Napoleonic administration (Luzzatto 1934; Tucci 1991).

Amsterdam and Hamburg

Amsterdam and Hamburg were city states with fairly similar institutional equilibria (Lindemann 2015), which faced the same kind of monetary problems at more or less the same time and which tried to address them with roughly similar solutions. In the last decades of the sixteenth and early decades of the seventeenth century, a massive wave of debasements (motivated by continuous warfare) took place in Europe. Debased coins started invading the circulation of international trade centers, thus compromising the fluidity of monetary transactions. In order to redress the quality of the circulating medium, in 1609, the City of Amsterdam opened a public exchange bank (the Wisselbank) whose mission consisted of withdrawing debased coins and releasing good-quality ones. As Venice had done some years earlier, Amsterdam created a forced demand for bank deposits by ruling that bills of exchange had to be compulsorily payable in bank. The measure was only partially successful, as the high fees imposed on the redemption of deposits into good specie discouraged potential depositors. In order to increase the attractiveness of bank money, in 1683, redemption fees were lowered, but the bank stopped paying in higher-quality coins than the ones it withdrew. The reform made bank money de facto inconvertible into the circulating medium, thus unpegging the value of the former from the latter. This strongly increased the popularity of bank money and allowed for the establishment of the Wisselbank as the pivot of the international monetary system (Quinn and Roberds 2009). Such a success allowed the government and government-sponsored entities (esp. the Dutch East India Company) to monetize nonnegligible amounts of debt through the bank. Moreover, the bank started to indirectly behave as a lender of last resort, as it provided a liquidity backstop to the fund of mutual assistance that had been created in order to extend loans to cash-strapped merchant banks (Uittenbogaard 2009). In the last decades of the eighteenth century, the United Provinces ran into serious military difficulties, and public debt monetization increased to such a scale that the value of bank money started to depreciate. As the country was transformed into a

territorial monarchy by the Napoleonic troops, the municipal Wisselbank was closed down and replaced by the privately owned Nederlandsche Bank (Quinn and Roberds 2016).

The Hamburger Bank (founded in 1619) was closely modeled along the Wisselbank. Its aim was the same: "cleaning" the domestic circulation from debased coins. However, the bank had a harder time than its Amsterdam model in stabilizing the value of its money. After many vicissitudes, in 1770, a great innovation was introduced: the bank pegged the value of its money to silver bullion rather than specie. In so doing, Hamburg introduced the first ingot standard in modern history and became the leading market for silver in Europe (Roberds and Velde 2016). Moreover, the Hamburger Bank also developed a business of collateralized lending (on the security of commodities) to private counterparties, and during the 1763 crisis, it acted as a lender of last resort to domestic merchant banks (Bindseil 2018). As the Free City of Hamburg survived the Napoleonic Wars, its bank also survived, until it was finally merged into the Reichsbank in 1875.

Barcelona

Barcelona is officially credited with having created the first public bank in Europe (Venice's Grain Office not having been designed as a banking organization proper). Founded in 1401, the Taula de Canvi was created primarily as the bank of the municipality in the context of the latter's conflicts with the Crown of Aragon (Riu 1979). In the early decades of its existence, the organization played an active role in the domestic payment system, as it opened current accounts to local bankers and extended credit by allowing them to overdraw. After facing a number of payment incidents as well as increasing difficulties in maintaining the convertibility of its money (jeopardized both by the rarefication of silver specie throughout Europe and by domestic depositors' frequent runs), in the second half of the fifteenth century, the Taula banned bankers from its customers. Although the actual effectiveness of the ban is unclear, it seems that the bank gradually refocused its business on its original mission as treasurer to the City, leaving aside all other competences in the domestic payment system. In this capacity, the bank financed the City's efforts to gain independence from the monarch in the event of the revolts of 1462, 1640, and 1713 (Usher 1943). The Taula provided the model for the many municipal banks that were created in the territories of the Crown of Aragon in the course of the fifteenth and sixteenth centuries, most of which survived until they were absorbed into the national central bank's branch network in the nineteenth century (Roberds and Velde 2016).

Genoa

Well known for their instability, Genoa's politics took a particularly dramatic turn after the loss of its naval power in the late fourteenth century. Unable to restore the credibility of its huge public debt, in 1407, the Republic found no better solution

than externalizing the management of both the debt and a number of major fiscal revenues to a "corporation of domestic bondholders." Founded as a fully private company held by the state's creditors, the Casa di San Giorgio proved an efficient device for granting the sustainability of the public debt (Fratianni 2006). Given the eminently fiscal nature of its mandate, it may not be surprising that San Giorgio's record in the monetary domain was less impressive. In the early decades of its life, the company actually opened current accounts to depositors, thus assuming a central role in the domestic payment system. At about the same time as the Taula de Canvi, however, San Giorgio was hit by the general rarefication of silver specie and started to face serious difficulties in maintaining the convertibility of its money. Just like the Taula, in 1444 San Giorgio hence reacted by withdrawing from banking activities altogether (Heers 1961). It was only in 1675 that the company reopened banking facilities and restarted issuing uniform bank money convertible into specie, thus reacquiring a central role in the domestic payment system (Felloni 2006; Gianelli 2006). After the reform, San Giorgio apparently lived a quiet existence until the French invaders deprived it from its fiscal competences in 1797 and was definitively closed down when the Republic was annexed by the Kingdom of Sardinia in 1815.

Central Banking in Centralized Territorial Polities

Central Banking in Centralized Territorial Polities: Issues

In the course of the early modern age, territorial monarchies generally underwent a process of centralization (encouraged by the spiraling costs of warfare). Besides securing a domestic monopoly on violence, rulers endeavored to centralize an increasing number of competences and to impose themselves as the sole source of political legitimacy in their territory. Unlike in the case of "merchant republics," in territorial monarchies, the alignment of interests between the borrower and her creditors was far from warranted (Stasavage 2012). This explains why distrust of state-issued money was widespread among potential money holders. In order to reassure them about the time consistency of monetary policy, territorial monarchs found it expedient to externalize money creation to a formally independent organization. The organization would refinance itself from private depositors and banknote holders on the one hand while relending to the state on the other hand. Its administrators' autonomy with respect to the monarch was supposed to work as a guarantee that debt monetization would remain sustainable in the long term.

Territorial rulers' pursuit of the "ideal type" of the absolute monarchy was obviously confronted with the opposition of those that were set to lose from it. In the Habsburg monarchies of Spain and Austria, attempts at centralization met the resistance of the many peripheral institutional bodies (regional parliaments and municipalities) that had been inherited from preexisting polities; the same was also the case in Scandinavia. In England, the absolutist ambitions of the Stuarts were challenged by two revolutions, the second of which managed to stably introduce serious limitations to the Crown's competences. In France, the apparent success of

the centralization process fostered by the Bourbons led to a spectacular accumulation of discontent, which eventually proved explosive in 1789.

The French Revolution and the many ensuing uprisings of the first half of the nineteenth century marked the end of the model of the absolute monarchy and ushered in the model of the constitutional national government. Under this new model, the centralization of competences to the national government continued to be pursued but was now counterbalanced by the rise in the power of antagonist national bodies like a parliament or an independent judiciary. The progressive extension of voting rights until the early decades of the twentieth century increased the legitimacy of the central government, making the rationale for a fully independent monetary authority less compelling than before. By the mid- twentieth century, money-issuing organizations had de jure or de facto become part of the public sector in most Western countries.

Naples, Austria, and Sweden

During the early modern period, a number of territorial monarchs that faced difficulties in establishing their legitimacy in the eyes of their subjects conferred issuing rights to some types of nonprofit organizations run by a formally independent body. The organization would invest a sizeable amount of its resources in government debt, but its displayed autonomy with respect to the sovereign would facilitate the indirect refinancing of the public debt by depositors and money holders. This strategy was inflected differently in a number of countries according to the different institutional equilibria in force.

The strategy was first pioneered in Naples in the second half of the sixteenth century, when the Kingdom made part of the Crown of Spain. Spanish rulers were quite unpopular, and discontent often generated civil unrest (the most spectacular being the revolution of 1647). Because raising short-term funding from the public was difficult in this context, the viceroy eyed the possibility of doing so through a number of popular charities that were already conducting a retail banking business at the time (the oldest one being the local Mount of Piety). On the one hand, the money issued by the seven chartered nongovernmental organizations was declared eligible for tax payments, thus making it attractive to the public; while on the other hand, the organizations used part of the collected funds to extend loans to the government. Chartered banks of issue were formally independent, but they were strictly supervised by the viceroyalty, which also appointed a representative to their boards. The expedient proved successful and allowed the Neapolitan monetary system to withstand a number of shocks until the Napoleonic wars prompted its reorganization in 1808, with the creation of a monopolistic bank of issue called Banco delle Due Sicilie (De Rosa 1991; Balletta 2008; Velde 2018).

Spanish Habsburgs had a hard time securing short-term funding without resorting to some intermediary. For many decades, they were obliged to resort to private intermediaries, esp. Genoese private bankers (Pezzolo and Tattara 2008; Drelichman and Voth 2011). The Austrian branch of the family long followed a similar strategy.

In the context of a difficult debt restructuring, however, in 1705, a new strategy was endeavored, as the monarch approved the creation of a public bank formally controlled by the City of Vienna (the Wiener Stadtbanco). This time, it was the municipality's credit that was supposed to elicit depositors' trust, hence allowing for the refinancing of the emperor's (and not of the City's) short-term debt. Although the administrators' actual independence at times proved to be more formal than substantial, the device worked reasonably well for almost a century, until it was finally disrupted by the financing of the Napoleonic wars. In 1816, it was replaced by a new privately owned bank of issue, the Oesterreichische Nationalbank (Jobst and Kernbauer 2016).

The road to absolutism was much bumpier in Sweden, where it bred an uninterrupted confrontation between the monarch and the aristocracy (which held control of the parliament). An early attempt to create a royal bank (the Stockholms Banco, founded in 1657) failed after a few years also because of the attacks of the nobility. In the aftermath, the parliament managed to create its own bank (the Riksbank, founded in 1668). The Riksbank was involved in the financing of the public debt, but only as long as this was directed to goals that were compatible with the parliament's (read, the aristocrats') own interests. The result was a dual system of debt monetization (partly accomplished by the bank and partly accomplished by the Treasury) that created much confusion in the domestic payment system. The duality came to an end after the Napoleonic wars, when the parliament definitively established its primacy on the monarch and the Riksbank eventually emerged as the country's sole monetary institution (Heckscher 1934; Ögren 2016).

England

The seventeenth century was a period of great instability in England. After the second ousting of the Stuart dynasty in 1688, a quick normalization of the domestic political life could hardly be taken for granted. When in 1694 the new sovereigns found themselves in the urgent need to raise funds, they had to adopt the solution Genoa had engineered in 1407: resorting to a private company incorporated by creditors. Unlike San Giorgio, however, the newly created Bank of England was originally designed as a temporary organization (set to expire after only 11 years) and was only destined to be in charge of one single portion of the public debt (Broz and Grossman 2004; Quinn 2008). Moreover, while to San Giorgio money creation was a secondary business (which was almost entirely shut down between 1444 and 1675), it was a primary one to the Bank of England: in its very first years of life, the Bank essentially had non-securitized long-term government loans on the assets side of its balance sheet and redeemable banknotes on the liabilities side. Such an extreme maturity transformation business appears to have been very risky, and its success was far from warranted (Clapham 1944). Fortunately for the Bank, Londoners had already got accustomed by goldsmith bankers to the use of banknotes (Quinn 1997), and the demand for such means of payment was further increased by the famine of specie engendered by the "Great Recoinage" of 1696 (Desan 2014).

Combined with the monopoly of joint-stock banking in London conferred by its charter, these circumstances helped the Bank establish itself at the center of the domestic payment system.

After the vagaries of the early eighteenth century (the most spectacular of which being the South Sea scheme: Neal 1990), since 1751, the public debt was consistently reorganized, and the Bank assumed a de facto monopolistic position in the provision of short-term funding to the government (von Philippovich 1911). At about the same time, the Bank started to develop substantially its discount activities, thus introducing the first lending facility ever offered to the private sector by a privileged money-issuing organization. In the meantime, since the de facto adoption of the gold standard in 1717, the Bank had assumed the role of gold market maker, which allowed it to stabilize the domestic monetary circulation (Ugolini 2017, pp. 230–234). Taken together, these three innovations redesigned the company into something much more akin to a modern central bank. By the time it entered the Napoleonic wars, the Bank had unquestionably become the ultimate manager of (1) the national payment system through its banknote and deposit creation, (2) the domestic credit system through its lending facility, (3) the government's short-term funding through its role as state treasurer, and (4) the foreign exchange market through its role as gold market maker.

In the course of the long nineteenth century, the Bank of England perfected its joint provision of these four central banking functions. (1) When the Bank Act of 1844 (which made the Bank's charter perpetual) posed strict limits on the issuance of its banknotes, the Bank joined the London Clearing House and thus facilitated the use of its deposits as ultimate means of exchange in the domestic payment system (Martin-Holland 1910). (2) After usury ceilings were lifted, in the mid-nineteenth century, the Bank started to act as a lender of last resort – i.e., to provide unlimited access to its lending facility upon deposit of eligible collateral (Bignon et al. 2012). As a counterpart, in the absence of formal bank regulation, the Bank developed a sophisticated informal system of supervision of its London counterparties (Flandreau and Ugolini 2013; Ugolini 2017, pp. 131–143). (3) The Bank started to systematically offset the impact of variations in "autonomous factors" like Treasury balances, thus stabilizing money market interest rates (Wood 1939). (4) The Bank gradually adopted an official final target (foreign exchange rate stability) and an official monetary policy instrument (the market interest rate), thus developing a consistent monetary policy strategy (Bindseil 2004; Ugolini 2016).

Taken together, all these moves point to an increasing concern by the Bank about its "public responsibilities." Such a concern was already a matter of fact at least in the 1830s, but the formally private nature of the Bank (and the ensuing danger of conflicts of interests) was often used as an argument for curtailing its competences (Clapham 1944). Such an argument became obsolete during World War I, when the Treasury made it clear that the Bank was now expected to be a mere executor of the government's decisions. This paved the way to the Bank's eventual nationalization in 1946 (Capie et al. 1994).

By the interwar period, the Bank of England had come to be seen as the "ideal type" of the modern central bank, and this idea was reinforced in the interwar as

British money doctors started to advise countries throughout the world on why and how to create a central bank (Capie et al. 1994). As proved by the uneventful nature of its eventual internalization by the public sector, however, the twentieth-century Bank had almost nothing in common with the design of the private company that had been originally established in 1694.

France

Like almost any other early-modern territorial monarchy, France also met considerable difficulties in stabilizing its short-term borrowing. Under the "lame duck" regency of Philippe d'Orléans (1715–1723), John Law infamously tried to defy the (by then, already well established) principle that a state-controlled bank of issue could not thrive under an absolutist government. The eventual wreckage of Law's super-company (issuing inconvertible legal-tender banknotes on the security of volatile streams of fiscal and monopoly revenues) may be interpreted as a confirmation of the principle. However, it is not inconceivable that the whole experiment had been pushed since its inception more as a temporary redistributive mechanism than as a permanent solution to the government's financial problems: as Law himself had acknowledged, the time consistency of the sovereign's action was a necessary precondition to the long-term success of a state bank (Neal 1990; Velde 2007).

After the shock of Law's bankruptcy, it took France half a century before a new banking charter was granted to a joint-stock company. This time, the principle was respected: the Caisse d'Escompte (founded in 1776) was a fully private bank, initially enjoying no privilege and set to be focused on lending to the private sector. Soon afterward, however, its money was declared eligible for tax payments, and as a counterpart the bank was increasingly involved into government finance. After the outbreak of the Revolution, the Caisse failed to be turned into a national bank, and its large exposure to the state's ruined finances triggered its closure in 1793 (Roberds and Velde 2016).

The very same group of financiers that stood behind the Caisse d'Escompte formed, however, the backbone of the shareholders of the new Banque de France authorized by first consul Bonaparte in 1800. Of course, Napoleon expected the new company to provide useful services to the state, although under slightly different respects than the monarchs of the ancien régime. On the one hand, the bank was expected to create an efficient payment infrastructure covering the whole territories ruled by France. This would have allowed establishing a reliable payment system and, consequently, improving the management of the Treasury's cash flows (Prunaux 2016). On the other hand, the bank was expected to support government finances only indirectly, by maintaining stable conditions in the money market under whatever circumstance. To the contrary, the bank was not asked to invest directly large amounts of its resources in the public debt (Ramon 1929). In the century that followed the fall of Napoleon, the Banque de France did actually operate along the very lines of conduct the emperor had originally drawn for it. It developed the largest branch network of Europe, providing payment, lending of last resort, and

supervisory facilities to the most remote corners of the country (Jobst 2010; Bignon and Avaro 2017); it defended the stability of domestic interest rates by every possible means (Bignon et al. 2012; Bazot et al. 2016); and most notably, it maintained its direct holdings of government debt at remarkably low levels (Jobst and Ugolini 2016), to the point of pushing the restored monarchs to create alternative mechanisms for securing the absorption of government loans (the Caisse des Dépôts et Consignations, founded in 1816 and put under the control of the parliament: Boudet 2006). As a result, the Banque de France followed in the nineteenth century a trajectory that was parallel but not convergent with respect to the English model (Bignon and Flandreau 2018). Even the circumstances of its eventual nationalization were different than those of the Bank of England, as this occurred after the government had been obliged to bail out the bank in order to cover the losses it had suffered on its foreign exchange reserves (Accominotti 2009).

Central Banking in Decentralized Territorial Polities

Central Banking in Decentralized Territorial Polities: Issues

As pointed out by theories of fiscal federalism (see, e.g., Alesina and Spolaore 2003), the question of the optimal provision of public policy assumes different contours in contexts where the level of government to which policymaking is delegated becomes an endogenous variable. In such contexts, reaching an efficient provision is remarkably more complicated, as the basic issue of the "equilibrium of power" between different interest groups is doubled with the additional issue of the "equilibrium of power" between different levels of jurisdiction.

Although none of the stories told in the previous sections was totally exempt from such issues, these manifested themselves on a truly macroscopic scale only in the case of federations or supranational unions, in which interest groups were really transversal and were not necessarily embodied by the one or the other level of jurisdiction. Because of the complexity (and unpredictability) of the outcomes of collective bargaining, the evolution of central banking has been particularly tortuous in these institutional contexts. Currently, the two largest monetary areas in the world (the United States and the European Monetary Union) are both the result of this type of process.

The United States

The history of the United States has always been (and still is) characterized by the dynamic coexistence of centripetal and centrifugal forces. As a matter of fact, the monetary realm was one in which clashes were most violent. Conflicts predated the writing of the federal pact, as they were a direct consequence of Britain's colonial policy. Before independence, British rulers had kept the right of coining full-bodied specie for themselves. In the meantime, they had allowed each single colonial

authority to issue its own inconvertible money in order to facilitate the management of cash flows and had imposed rules on how monetization should be conducted; however, they had also made clear that they would bear no responsibility for any accident that might have occurred. As the performance of money creation varied widely across colonies, each would-be state was encouraged to "keep the house in order" on its own. When the War of Independence erupted, the first federal body that was created (the Continental Congress) found serious difficulties in financing the war effort, as no fiscal revenue had been transferred to it. Following the collapse in the value of the inconvertible money it had issued, in 1782, the Congress backed the foundation of the Bank of North America (a joint-stock company modeled along the original Bank of England of 1694); however, the states refused to make the new bank's money eligible for tax payments, thus killing the project. In 1787, the Constitutional Convention meant to draw a definite line between the monetary competences of the federal and state level: states would be prevented from issuing money directly but maintained the right to charter local banks of issue, while Congress would have the exclusive right to "coin money and regulate the value thereof" – viz., to strike metallic specie. While the meaning of such wording was apparently straightforward to its writers, its correct interpretation was the subject of many decades of harsh disputes (only definitively settled by the Supreme Court in the 1880s) on whether the Constitution provided federal authorities with the right to undertake or regulate the issuance of paper money (Timberlake 1993; Grubb 2003).

Given the unclear legality of federal interventions in the banking realm, the monetary solutions sponsored by Congress proved fragile. After the failed attempt to create a Bank of North America, two Banks of the United States were provided a 20-year federal charter (in 1791 and 1816, respectively). Both were Philadelphia-based joint-stock companies modeled along the Bank of England; both operated actively to create a national payment system; both provided lending facilities to the public; both acted as treasurers to the federal government; and both acted successfully to clean the domestic circulation from wartime inconvertible paper money. For all their merits, however, both faced harsh criticism from the political opponents of centralization and failed to secure the renewal of their charter when this came to expiration (Catterall 1903; Wettereau 1942; Knodell 2016; Javat 2017).

In the meantime, states had not refrained from using their prerogatives in the domain of banking. Banks of issue were chartered and regulated at the state level, and regulation varied considerably across the country. During the first half of the nineteenth century, basically all of the modern tools of banking regulation were invented or developed by different US states. Among these, the principle that the issuance of banknotes should be collateralized by a given amount of public debt became established almost everywhere (Rockoff 1991). This indirect strategy for pursuing government debt monetization through regulation would later become the cornerstone of the National Banking System created in 1863–1865 (Ugolini 2017, pp. 143–148 and 196–198).

The shifting of political equilibria during the Civil War allowed for the centralization of a number of banking competences to the federal level. Concerning money

creation, the Treasury resorted massively to the issuance of legal-tender notes (the "greenbacks") in order to finance the war and henceforth retained its issuing rights for more than a century. Concerning banking regulation, a dual system was created: on the one hand, national banks were regulated at the federal level and were henceforth authorized to issue banknotes on the security of federal debt; on the other hand, state banks remained regulated at the state level and were allowed to issue banknotes only at prohibitive costs (note that, despite their official labelling, national banks did *not* operate throughout the federation, as interstate branching remained prohibited). Despite creating a pyramidal banking system through the introduction of reserve requirements (White 1983), the reforms did not provide for a countercyclical mechanism to stabilize the payment and credit system in times of crises. The Treasury partially compensated for this lack by moving its liquid assets across national banks, but only on a limited scale (Taus 1943). Private clearinghouses developed interbank lending facilities in order to compensate for the lack of a lender of last resort, but these were only accessible to clearinghouse members, i.e., only to a minority of banks (Jaremski 2018).

The increasing cost of banking crises in the ensuing decades became a matter of popular concern, and calls for reform started to be voiced. In particular, the dramatic crisis of 1907 opened a window of opportunity that a coalition of different interest groups managed to seize (Broz 1997). In 1913, the Federal Reserve System was created. Rather than the transposition of the English model, it was an incremental development of the National Banking System. The new organization was conceived as a user-owned facility, held by the very commercial banks which joined the System. Because interstate branching was a political taboo, the banking system continued to be fragmented; in order to bring the new facility as close as possible to its users, 12 fully independent Federal Reserve Banks were created, whose areas of competence closely followed the structure of the pyramidal reserve system designed by the Acts of 1863–1865 (Jaremski and Wheelock 2017). Moreover, as a substantial share of state banks did not find it worthwhile to join, the dual regulatory system put in place during the Civil War remained unscathed: as it had been the case for clearinghouses, access to the Fed's payment and lending facilities remained restricted to members (Calomiris et al. 2016).

The creation of the Fed was certainly a further step toward centralization in the banking realm, but not the definitive one. The 12 independent Reserve Banks could fix their monetary stance regardless of the others, and this created serious coordination problems during crises (Meltzer 2003). Moreover, the dual nature of the banking system impeded Reserve Banks from extending lending of last resort to nonmember banks (Bordo and Wheelock 2013). Taken together, these limitations produced a reaction to the banking crises of the 1930s that was clearly suboptimal. Dissatisfaction motivated the Banking Acts of 1933–1935, which transformed the Fed into a de facto federal agency. The New Deal reforms particularly addressed the coordination problem, by centralizing the conduct of monetary policy to the Federal Reserve Board in Washington. They did not, however, address the question of the duality of the regulatory system, which still persists to date.

The European Union

Formally, the European Union is at present the only example of a supranational union. Such a uniqueness must not be exaggerated, as the Union might legitimately be seen as a sort of "federation in the making" like a number of nowadays' federations happened to be in the past. That said, the experience of European monetary unification was unique in that it consisted of federating a number of preexisting national central banks rather than imposing a new organization on top of them. Differently put, the process did not consist of managing the conflict between competing jurisdictions as in the case of the United States, but rather of coordinating the action of a number of peer jurisdictions maintaining both their full formal sovereignty and the theoretical option to leave the system.

The process of European monetary unification started in the 1970s, when the demise of Bretton Woods forced Europeans to continue the system on a regional basis (Mourlon-Druol 2012). The recurrent foreign exchange crises proved costly to European economies and jeopardized the process of commercial integration initiated in the 1950s. As the viability of a system of fixed but adjustable foreign exchange parities was proving incompatible with free capital mobility, European countries started to think about the possibility of "merging" their currencies into a single one. In order to minimize the risk that discussions about the distributional effects of unification might derail the process, negotiations were always kept focused on technical issues (James 2012). Technical difficulties were actually far from marginal, as monetary structures and practices differed substantially across countries. A long process of convergence was set in motion; a number of differences still persisted in the late 1990s, yet not to the point of compromising the viability of a monetary policy coordinated at the centralized level but implemented at the decentralized one (Forssbæck and Oxelheim 2003; Galvenius and Mercier 2011).

National central banks had already been granted independence by their governments before the creation of Eurosystem, but independence was made a precondition for membership. This was consistent with the technocratic nature of the project, which was believed to be viable only in isolation with respect to political interferences. The European Central Bank (established in 1998) was given charge for the coordination in the management of the payment infrastructure (through the creation of the TARGET system: Kokkola 2010) and in monetary policymaking. By contrast, banking supervision and (of course) the role of state treasurer stayed with national authorities. The fact that supervision remained confined to the decentralized level was not perceived as an issue in the beginning: financial regulation had been on the path to homogenization since the 1980s, and there was no consensus on the fact that the central bank was the best candidate for conducting supervision (Masciandaro and Quintyn 2016). The crisis which started in 2007 (entailing a number of big bailouts of banks by national fiscal authorities) showed the limits of this approach and brought back to light all the political dimensions of monetary unification that had been meticulously hidden until that moment (James 2012). In 2012, member countries agreed on a number of reforms of the Eurozone's fiscal and monetary governance: these included the creation of the so-called Banking Union, i.e., a single

supervisory mechanism and a single resolution mechanism for failing banks. As a result, the European Central Bank was eventually entrusted with the supervision of bigger Eurozone banks, while smaller ones were left under the control of national central banks. Curiously, this produced (for the first time in Europe) a dual supervisory system that somewhat echoes the American one.

Conclusions

This chapter has briefly surveyed available historical evidence on a number of relevant episodes in the evolution of the provision of monetary and financial stability. The results appear to suggest that in central banking one size does *not* fit all and that the evolution of central banking did not take place according to the principle of the "survival of the fittest" (Roberds and Velde 2016). On the one hand, the "extinction" of some organizational solutions (e.g., the early public banks created by European city states) appears to have been tied to the shake-up of the institutional equilibria in which they operated rather than to their inherent suboptimality. On the other hand, although some specific organizational solutions (e.g., the Bank of England) were actually seen as "models" to some extent, the development of central banking appears to have been mostly determined by domestic political economy factors rather than by foreign inspiration. Different paths toward monetary and financial stability have been pursued across time and space, and experimentation seems to go on today. After all, central banking is policymaking, and policymaking does not take place in a vacuum: in the future, it will presumably continue to evolve following changes in institutional equilibria as much as it did in the past.

Cross-References

- ▶ Experiments with Paper Money
- ▶ Money, Law, and Institutions
- ▶ Money, Trade, and Payments in Preindustrial Europe

References

Accominotti O (2009) The sterling trap: foreign reserve management at the Bank of France, 1928–1936. Eur Rev Econ Hist XIII:349–376

Alesina A, Spolaore E (2003) The size of nations. MIT Press, Cambridge MA

Balletta F (2008) La circolazione della moneta fiduciaria a Napoli nel Seicento e nel Settecento (1587–1805). Edizioni Scientifiche Italiane, Naples

Bazot G, Bordo MD, Monnet E (2016) International shocks and the balance sheet of the Bank of France under the classical gold standard. Explor Econ Hist LXII:87–107

Bignon V, Avaro M (2017) Socially embedded money creation: monetary policy and Bank of France counterparty risk management in late nineteenth-century France. Working paper

Bignon V, Flandreau M (2018) The other way: a narrative history of the Bank of France. In: Edvinsson R, Jacobson T, Waldenström D (eds) Sveriges Riksbank and the history of central banking. Cambridge University Press, Cambridge, pp 206–241
Bignon V, Flandreau M, Ugolini S (2012) Bagehot for beginners: the making of lending-of-last-resort operations in the mid-19th century. Econ Hist Rev LXV:580–608
Bindseil U (2004) Monetary policy implementation: theory, past, present. Oxford University Press, Oxford
Bindseil U (2018) Pre-1800 central bank operations and the origins of central banking. Working paper
Bordo MD, Siklos PL (2018) Central banks: evolution and innovation in historical perspective. In: Edvinsson R, Jacobson T, Waldenström D (eds) Sveriges Riksbank and the history of central banking. Cambridge University Press, Cambridge, pp 26–89
Bordo MD, Wheelock DC (2013) The promise and performance of the Federal Reserve as lender of last resort, 1914–1933. In: Bordo MD, Roberds W (eds) The origins, history, and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, New York, pp 59–98
Boudet J-F (2006) La Caisse des dépôts et consignations: histoire, statut, fonction. L'Harmattan, Paris
Broz JL (1997) The international origins of the Federal Reserve System. Cornell University Press, Ithaca/London
Broz JL, Grossman RS (2004) Paying for privilege: the political economy of Bank of England charters, 1694–1844. Explor Econ Hist XLI:48–72
Calomiris CW, Jaremski MS, Park H, Richardson G (2016) Liquidity risk, bank networks, and the value of joining the Federal Reserve System. Working paper
Capie F, Goodhart CAE, Schnadt N (1994) The development of central banking. In: Capie F, Goodhart CAE, Fischer S, Schnadt N (eds) The future of central banking: the tercentenary symposium of the Bank of England. Cambridge University Press, Cambridge, pp 1–231
Catterall RCH (1903) The Second Bank of the United States. University of Chicago Press, Chicago
Clapham J (1944) The Bank of England: a history. Cambridge University Press, Cambridge
De Rosa L (1991) Banchi pubblici, banchi privati e monti di pietà a Napoli nei secoli XVI-XVIII. In: Banchi pubblici, banchi privati e monti di pietà nell'Europa preindustriale. Società Ligure di Storia Patria, Genoa, pp 497–512
Desan C (2014) Making money: coin, currency, and the coming of capitalism. Oxford University Press, Oxford
Diamond DW, Dybvig PH (1983) Bank runs, deposit insurance, and liquidity. J Polit Econ XCI:401–419
Drelichman M, Voth H-J (2011) Lending to the borrower from hell: debt and default in the age of Philip II. Econ J CXXI:1205–1227
Felloni G (2006) Il credito all'erario e ai privati: forme ed evoluzioni. In: Felloni G (ed) La Casa di San Giorgio: il potere del credito. Società Ligure di Storia Patria, Genoa, pp 155–164
Flandreau M, Ugolini S (2013) Where it all began: lending of last resort and Bank of England monitoring during the Overend-Gurney panic of 1866. In: Bordo MD, Roberds W (eds) The origins, history, and future of the Federal Reserve: a return to Jekyll Island. Cambridge University Press, New York, pp 113–161
Forssbæck J, Oxelheim L (2003) Money markets and politics: a study of European financial integration and monetary policy options. Edward Elgar, Cheltenham
Fratianni M (2006) Government debt, reputation and creditors' protections: the tale of San Giorgio. Rev Financ X:487–506
Galvenius M, Mercier P (2011) The story of the Eurosystem framework. In: Mercier P, Papadia F (eds) The concrete euro: implementing monetary policy in the euro area. Oxford University Press, Oxford, pp 115–214
Gianelli G (2006) La riforma monetaria genovese del 1671-75 e l'apertura del banco di moneta corrente. In: Felloni G (ed) La Casa di San Giorgio: il potere del credito. Società Ligure di Storia Patria, Genoa, pp 121–142

Giannini C (2011) The age of central banks. Edward Elgar, Cheltenham
Goodhart CAE (1988) The evolution of central banks. MIT Press, Cambridge MA/London
Grubb F (2003) Creating the U.S. dollar currency union, 1748-1811: a quest for monetary stability or a usurpation of state sovereignty for personal gain? Am Econ Rev XCIII:1778–1798
Heckscher EF (1934) The Bank of Sweden in its connection with the Bank of Amsterdam. In: Van Dillen JG (ed) History of the principal public banks. Nijhoff, The Hague, pp 161–200
Heers J (1961) Gênes au XVe siècle: activité économique et problèmes sociaux. SEVPEN, Paris
James H (2012) Making the European Monetary Union. Belknap Press, Cambridge MA/London
Jaremski M (2018) The (dis)advantages of clearinghouses before the Fed. J Financ Econ CXXVII:435–458
Jaremski M, Wheelock DC (2017) Banker preferences, interbank connections, and the enduring structure of the Federal Reserve System. Explor Econ Hist LXVI:21–43
Javat A (2017) An analysis of the balance sheet of the First Bank of the United States. Working paper
Jobst C (2010) Gouverner une banque centrale décentralisée: l'exemple austro-hongrois (1847-1914). In: Feiertag O, Margairaz M (eds) Gouverner une banque centrale: du XVIIe siècle à nos jours. Albin Michel, Paris, pp 113–140
Jobst C, Kernbauer H (2016) The quest for stable money: central banking in Austria, 1816–2016. Campus, Frankfurt am Main
Jobst C, Ugolini S (2016) The coevolution of money markets and monetary policy, 1815–2008. In: Bordo MD, Eitrheim Ø, Flandreau M, Qvigstad JF (eds) Central banks at a crossroads: what can we learn from history? Cambridge University Press, New York, pp 145–194
Knodell J (2016) The Second Bank of the United States: "central" banker in an era of nation building, 1816–1836. Routledge, London/New York
Kokkola T (2010) The payment system: payments, securities and derivatives, and the role of the Eurosystem. European Central Bank, Frankfurt am Main
Lindemann M (2015) The merchant republics: Amsterdam, Antwerp, and Hamburg, 1648–1790. Cambridge University Press, New York
Luzzatto G (1934) Les banques publiques de Venise (siècles XVI–XVIII). In: Van Dillen JG (ed) History of the principal public banks. Nijhoff, The Hague, pp 39–78
Martin-Holland R (1910) The London Bankers Clearing House. In: The English banking system. National Monetary Commission, Washington, DC, pp 267–294
Masciandaro D, Quintyn M (2016) The governance of financial supervision: recent developments. J Econ Surv XXX:982–1006
Meltzer AH (2003) A history of the Federal Reserve, vol I: 1913–1951. University of Chicago Press, Chicago
Mourlon-Druol E (2012) A Europe made of money: the emergence of the European Monetary System. Cornell University Press, Ithaca
Mueller RC (1997) The Venetian money market: banks, panics, and the public debt, 1200–1500. Johns Hopkins University Press, Baltimore
Neal L (1990) The rise of financial capitalism: international capital markets in the age of reason. Cambridge University Press, Cambridge
Ögren A (2016) From parliamentary institution to a modern central bank: the Bank of Sweden. In: Feiertag O, Margairaz M (eds) Les banques centrales et l'État-nation. Presses de Sciences Po, Paris, pp 49–72
Padgett JF, McLean PD (2011) Economic credit in Renaissance Florence. J Mod Hist LXXXIII:1–47
Pezzolo L, Tattara G (2008) "Una fiera senza luogo": was Bisenzone an international capital market in sixteenth-century Italy? J Econ Hist LXVIII:1098–1122
Prunaux E (2016) Le rôle de le Banque de France dans l'édification de l'État napoléonien. In: Feiertag O, Margairaz M (eds) Les banques centrales et l'État-nation. Presses de Sciences Po, Paris, pp 99–135
Quinn S (1997) Goldsmith-banking: mutual acceptance and interbanker clearing in Restoration London. Explor Econ Hist XXXIV:411–432

Quinn S (2008) Securitization of sovereign debt: corporations as a sovereign debt restructuring mechanism in Britain, 1694 to 1750. Working paper
Quinn S, Roberds W (2009) An economic explanation of the early Bank of Amsterdam, debasement, bills of exchange and the emergence of the first central bank. In: Atack J, Neal L (eds) The origins and development of financial markets and institutions: from the seventeenth century to the present. Cambridge University Press, Cambridge/New York, pp 32–70
Quinn S, Roberds W (2016) Death of a reserve currency. Int J Centr Banking XII:63–103
G. Ramon (1929) Histoire de la Banque de France d'après les sources originales (Paris: Grasset)
Riu M (1979) Banking and society in late medieval and early modern Aragon. In: The dawn of modern banking. Yale University Press, New Haven/London, pp 131–168
Roberds W, Velde FR (2016) Early public banks. In Fox D, Ernst W (eds) Money in the Western legal tradition: Middle Ages to Bretton Woods. Oxford University Press, Oxford, 321–355 and 465–488
Rockoff H (1991) Lessons from the American experience with free banking. In: Capie F, Wood GE (eds) Unregulated banking: chaos or order? Macmillan, Basingstoke, pp 73–109
Stasavage D (2012) States of credit: size, power, and the development of European polities. Princeton University Press, Princeton and Oxford
Taus ER (1943) Central banking functions of the United States Treasury, 1789–1941. Columbia University Press, New York
Timberlake RH (1993) Monetary policy in the United States: an intellectual and institutional history. Chicago University Press, Chicago/London
Tucci U (1991) Il banco pubblico a Venezia. In: Banchi pubblici, banchi privati e monti di pietà nell'Europa preindustriale. Società Ligure di Storia Patria, Genoa, pp 309–325
Ugolini S (2016) Liquidity management and central bank strength: Bank of England operations reloaded, 1889–1910. Norges Bank Working Paper, 10/2016
Ugolini S (2017) The evolution of central banking: theory and history. Palgrave Macmillan, London
Uittenbogaard R (2009) Lending by the Bank of Amsterdam (1609–1802). In: Van Nieuwkerk M (ed) The Bank of Amsterdam: on the origins of central banking. Sonsbeek, Arnhem, pp 120–131
Usher AP (1943) The early history of deposit banking in Mediterranean Europe, vol I. Harvard University Press, Cambridge MA
Velde FR (2007) John Law's system. Am Econ Rev XCVII:276–279
Velde FR (2018) The Neapolitan banks in the context of early modern public banks. In: Costabile L, Neal L (eds) Financial innovation and resiliency: a comparative perspective on the public banks of Naples, 1462–1808. Palgrave Macmillan, London
von Philippovich E (1911) History of the Bank of England and its financial services to the state. National Monetary Commission, Washington, DC
Wettereau JO (1942) The branches of the First Bank of the United States. J Econ Hist II (Supplement):66–100
White EN (1983) The regulation and reform of the American banking system, 1900–1929. Princeton University Press, Princeton
Wood E (1939) English theories of central banking control, 1819–1858. Harvard University Press, Cambridge MA

The Evolution of Monetary Policy (Goals and Targets) in Western Europe

34

Duncan J. Needham

Contents

Abstract

This chapter charts the evolution of monetary policy in the United Kingdom, France, and Germany since the late nineteenth century. It shows how the monetary authorities in the three largest European economies transitioned from the classical gold standard through the gold exchange standard and the Bretton Woods regime to European Monetary Union, while dealing with war, reconstruction, and inflation along the way. It outlines the changing goals of monetary policy and the targets, instruments, and devices deployed to achieve those goals. In doing so, it highlights the constraints within which policy makers operated under different monetary regimes.

Keywords

Gold standard · Gold exchange standard · Bretton Woods · European Monetary Union · Euro

D. J. Needham (✉)
University of Cambridge, Cambridge, UK
e-mail: djn33@cam.ac.uk

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_47

Introduction

The monetary policy literature can be abstruse and occasionally strays into the forbidden (for the historian) territories of anachronism and teleology. After all, anyone writing about "monetary policy" in the nineteenth century risked being labeled a crank. There are, nonetheless, common themes in the evolution of Western European monetary policy, even if the reality often bore little resemblance to current theory. As well as the standard goals of economic growth, price stability, high employment, and external balance, monetary policy has been directed towards maintaining the convertibility of national currencies into fixed amounts of specie (and other currencies), financing war and reconstruction, and stabilizing financial systems. Monetary authorities have set targets for a number of variables believed to have enjoyed a relationship with one or more of these goals. These have included exchange rates, bank reserves, and the growth of the money supply. The principal instruments have been discount rates, liquidity and reserve requirements, and open market operations, often supported by the array of less-visible "devices" described below.

This chapter charts the evolution of Western European monetary policy since the classical gold standard in the three largest economies: the United Kingdom, France, and Germany. Limited space precludes wider geographical coverage but, as Eichengreen points out, "the smaller European countries were satellites revolving around these suns" (Eichengreen 1992). We begin with the United Kingdom, if not the conductor of the late nineteenth-century international monetary orchestra, then certainly one of its leading players.

The United Kingdom

In 1914, the UK had been on a gold standard for almost two centuries. This was more than a commitment by the Bank of England (the "Bank") to exchange sterling claims for fixed amounts of specie. Alongside free trade and balanced government budgets, it was one of the "three pillars of the anti-collectivist temple," imposing morality on business (a "golden brake on the credit machine") and governments (by limiting their recourse to the printing press) (Checkland 1983; Schumpeter 1952). By 1914 the gold standard was embedded in British political culture, viewed as a guarantor of peace, stability, and cheap imports – important for a country that could no longer feed itself (Daunton 2007; Trentmann 2009). It also had the virtue of simplicity; the nation was either on gold or it was not. And if it left, there was the expectation that it would rejoin at the same level (Bordo and Kydland 1995).

The Bank's primary goal was to mitigate financial crises by ensuring it had enough gold to meet an internal drain (Bank note holders wishing to convert into gold) and/or an external drain (gold exports). With most Issue Department liabilities (Bank notes) backed by gold under the terms of the 1844 Bank Charter Act, the Bank focused on the "Proportion" – gold held in reserve against Banking Department liabilities (public and private deposits). This was analogous to a commercial bank's

reserves and, while there was no fixed rule, the Bank appears to have targeted a Proportion of 40–45%. Exchange rates were fixed by the relative gold content of each currency, with a pound convertible into 4.86 times as much pure gold as a US dollar, for instance. If the US price level fell below the UK by more than the cost of financing, transporting, insuring, and assaying, then it became profitable to buy gold from the Bank, transport it to the USA, and have it re-minted into dollars. The Bank therefore monitored the "gold points," the prices at which gold would begin to move in and out of London. As sterling approached its gold export point, the Bank might deploy instruments to make London a more attractive home for international capital.

The primary instrument was Bank Rate. This was the rate at which eligible financial institutions (discount houses) accessed the Bank's facilities. After WW1, it was also the reference rate for (London) bank deposits and lending to different sectors of the economy. If money were plentiful and the discount houses had no need of the Bank's facilities, then market rates could be significantly lower than Bank Rate. To make Bank Rate effective, the Bank would take money "off the market" through what became known as open market operations: selling assets from its portfolio or borrowing from customers (such as the Council of India and the Bank of Japan) and the commercial banks (Sayers 1976). "Hitching-up" market rates to Bank Rate brought the discount houses "into the Bank" to rediscount two-name commercial bills or borrow against collateral at penalty rates. This was the "classical system" developed in the 1890s (Hotson 2017).

Since the Bank's "thin film of gold" often comprised less than 3% of the domestic money supply, it also employed several devices, sometimes alongside more orthodox instruments (Sayers 1976). It could move the gold points – raising the cost of speculation by paying out lesser-weight coins, raising the price of (or refusing to sell) foreign coin (not subject to the Bank Charter Act or the coinage laws), raising the price of (or refusing to sell) gold bars, raising the price of gold for exporters, becoming "awkward" when selling gold, or extending interest-free loans to importers. *In extremis*, the 1844 Bank Charter Act could be suspended and the fiduciary note issue increased by an Order in Council.

The initial response to the financial crisis that preceded the outbreak of war in 1914 was a hike in Bank Rate to 10% on 1 August. As the scale of the crisis became apparent, the rate was lowered to 5% as the Bank discounted heavily and extended loans to acceptors of bills embarrassed by the closure of the London Stock Exchange and the freezing of overseas remittances (Morgan 1952; Roberts 2014). As Sayers explains:

> a gigantic financial business had been habitually run without any significant cash stock because there was confidence in a continuous flow of money; confidence in the flow was suddenly destroyed; a stock of money was injected to restore confidence (Sayers 1976).

The Treasury issued £1 and 10 shilling legal tender currency notes, backed partly by gold and Bank notes but mostly by government securities (therefore "fiduciary"). These surpassed Bank notes by the end of 1914 as the principal circulating medium (Morgan 1952). Combined with heavy discounting and advances to the government,

this was a significant injection of liquidity. It was reflected in market rates which, in February 1915, were 4 percentage points below Bank Rate, threatening inflation and net capital outflows. The authorities responded by drawing money off the market with War Loans, Treasury bills, and Exchequer Bonds, replacing the Treasury tender in April 1915 with an unlimited tap at a rate of their choosing, drawing special deposits from the banks to the government as Ways and Means advances, and pressuring bankers to raise lending rates and prioritize particular sectors.

Before the war, Treasury bills comprised just 1% of the bill market; in 1919 with more than £1 billion outstanding, they dominated (Fletcher 1976; Moggridge 1972). This had a lasting impact on the conduct of monetary policy. Commercial bills usually liquidated underlying transactions, so issuance correlated with economic activity. Higher interest rates could be expected to depress activity while making commercial bills (and sterling) more attractive to overseas investors. But government expenditure, especially during wartime, is less sensitive to interest rates. Further, to minimize the cost of servicing the burgeoning national debt (up from 26% of GDP in 1913 to 135% in 1919), Treasury bill rate was held below market rates with unlimited volumes available at the tap rate. This replaced the discount rate on commercial bills as the rate that regulated domestic finance and capital flows in and out of London.

The Bank maintained *de jure* convertibility of the pound throughout the war. This was deemed essential for obtaining finance and supplies from the USA and for the long-term future of the City. Patriotic appeals against private hoarding were bolstered by "making it extremely difficult and inconvenient for the ordinary man to get gold" and commandeering gold imports (Eichengreen 1992; Officer 1996; Sayers 1976). Insurance for gold shipments was made prohibitively expensive, before exports were banned in May 1917. After dipping to $4.56, the exchange rate was stabilized at $4.76½ with American loans, the requisitioning of British overseas investments, and in July 1916, a rise in Bank Rate to 6%. An exchange rate below the prewar parity made it profitable to export gold, a less risky proposition in the absence of German submarines after the Armistice. So the March 1919 unpegging of sterling was swiftly followed by a statutory embargo on gold exports. Nonetheless, without significant foreign exchange reserves, the pound fell to $3.20 by February 1920 (Sayers 1976).

The authorities had announced their intention of restoring the prewar gold standard in 1918:

> it is imperative that after the war the conditions necessary to the maintenance of an effective gold standard should be restored without delay. Unless the machinery which long experience has shown to be the only effective remedy for an adverse balance of trade and an undue growth of credit is once more brought into play, there will be a grave danger of a progressive credit expansion which will result in a foreign drain of gold menacing the convertibility of our note issue and so jeopardising the international trade position of the country (Cunliffe 1918).

While there was debate about when Britain should restore the gold standard, there was less discussion of the parity, the assumption being that wartime inflation could

be reversed, especially in relation to the USA. Prices and wages were flexible during the immediate postwar inflation and subsequent deflation but became increasingly sticky, with Keynes estimating the pound was overvalued by 10% when Chancellor Winston Churchill announced the resumption of gold exports in April 1925 (Keynes 1925). As Matthews, Feinstein, and Odling-Smee point out, the authorities were "acting as if income from abroad was still high and increasing" (Matthews et al. 1982). In 1913, overseas investment income had contributed to a surplus on invisible trade of more than 10%, essential for a country that ran persistent deficits on visible trade (Middleton 2002). But Nearly a quarter of British overseas assets were sold during the war (Eichengreen 1992). With less overseas investment income, lower earnings from services such as shipping, and the booming US stock market a magnet for British capital, a positive interest rate differential with New York (sometimes also Paris and Berlin) was critical to protecting the reserves. This required hitching Treasury bill rate to Bank Rate *and* keeping Bank Rate above the New York rate. Combined with the impact of the over-valued pound on net exports, this depressed domestic activity. While the USA enjoyed the "roaring twenties," unemployment in the UK averaged more than 8% from 1921 to 1929. The Bank came under pressure, especially from Churchill, not to deflate further, and left Bank Rate unchanged at 5% from December 1925 until April 1927 (Sayers 1976). Officials devised a new instrument to defend sterling, a reserve of US dollars concealed in the Bank's "Other Securities" account. In 1927 they also revived the gold devices to discourage speculation (Sayers 1976). Finally, open market operations became a more powerful instrument in 1928 when the amalgamation of the Bank and Treasury note issues provided the Bank with additional Treasury bills to deploy.

The delicate state of government finances and the current account, against a backdrop of industrial distress, meant the Bank had few defenses against the monetary storms that blew in during the summer of 1931. As Janeway points out "Britain had already suffered nearly a decade of unemployment in excess of 1 million insured workers. This was the central economic fact that constrained monetary policy throughout the period" (Janeway 1995–1996). Denuded of reserves, and facing the collapse of several merchant banks (following the Central European banking panic), the authorities legally suspended gold sales in September 1931, floated the pound, and ushered in a period of "cheap money" that lasted two decades (Accominotti 2012). Bank Rate was lowered to 2% by June 1932, fuelling GDP growth of more than 4% per annum between 1933 and 1937. The authorities also established the Exchange Equalisation Account to manage sterling and encouraged the cartel arrangements that would provide the framework for postwar monetary management (Fletcher 1976; Goodhart and Needham 2017; Howson 1980).

With convertibility suspended, the primary goals of monetary policy became reducing the cost of debt servicing (the 1932 War Loan conversion) and encouraging economic recovery while maintaining price stability (Howson 1975). Officially, monetary control rested on the fulcrum of the banks' 10% cash-to-deposits ratio (Macmillan 1931). Unofficially, "all thought of control was obscured by the problem of keeping the [discount] market alive" (Sayers 1976). The slump had reduced the

supply of commercial bills to a trickle and the policy (with the 1919–1920 inflation in mind) of funding the national debt with gilts kept market rates below 1%.

Exchange controls were imposed immediately upon the outbreak of the Second World War. Bank Rate, after a "storm signal" hike to 4% in August 1939, quickly returned to 2% (Sayers 1956). Treasury bills were held at 1% with longer-dated gilts at 3%. New issues were controlled by the Capital Issues Committee and savings were channeled to the government via an array of products tailored for different investors (Broadberry and Howlett 1998). Clearing bank liquidity was controlled with sales of Treasury bills and Treasury Deposit Receipts which, together, financed a third of the wartime deficit (Howson 1993; Sayers 1956). These measures combined with budgetary policy and rationing to keep price rises to about 50% during the war (versus *c*. 250% during and after the First World War).

Cheap money became "cheaper money" under the 1945 Labour government. Treasury bill rate was reduced to ½% and the government issued "irredeemable" debt at 2½% (Howson 1993). This was partly to avert the type of slump that had followed the First World War, but also because, with the national debt peaking at 238% of GDP in 1946, higher interest rates would have required even higher taxes. Chancellor Hugh Dalton believed inflation could be contained with continued financial repression, while cheaper money would ease the transition to a peacetime economy and reduce the cost of political objectives such as the nationalization program and wealth redistribution (by hastening the "euthanasia of the rentier" and limiting bank profits) (Allen 2014; Howson 1993). Meanwhile the $3.75 billion Anglo-American loan would finance the current account deficit. "Cheaper money" was abandoned in 1947, the year of the convertibility crisis. Nonetheless, market rates remained below Bank Rate, forcing the Bank to rely on increased funding of the floating debt (26% of national debt in 1945) into longer maturities, and requests for credit restraint (Howson 1993).

The new Conservative government reactivated monetary policy by raising Bank Rate to 2½% in November 1951 (4% in March 1952) and unfixing Treasury bill rate. While nearly a third of Treasury bills were funded into gilts, the remaining short-term debt could still be monetized, rendering the cash ratio ineffective. Attention therefore turned to the third principal monetary policy instrument – bank reserves. The banks had maintained an informal ratio of cash, call loans, and discountable assets (commercial bills, Treasury bills, and local authority bills) against liabilities since the 1930s. In 1946, largely to ensure continued demand for Treasury bills, the Bank indicated that a 28–32% liquidity ratio should be "regarded as normal" (Bank of England 1962). In 1957 this was narrowed to 30%.

Increasing emphasis on the liquidity ratio symbolized the transition from "Macmillan era" emphasis on monetary control via the cash ratio, to the "Radcliffe era." The Radcliffe Committee was appointed in 1957 to investigate the failure of monetary policy to contain the 1955 pre-election boom. Reflecting the Keynesian tenor of the times, its 1959 report concluded that: "monetary measures cannot alone be relied upon to keep in nice balance an economy subject to major strains from both without and within. Monetary measures can help, but that is all" (Radcliffe 1959).

The goals were to support fiscal policy in maintaining full employment and manage the national debt, while keeping an eye on the fixed exchange rate.

In addition to Bank Rate, liquidity ratios, and open market operations, the authorities deployed an array of instruments including special deposits, credit controls, and hire purchase controls (Needham 2014a). Special deposits required the banks after 1958 (later also the finance houses) to post a percentage of their advances at the Bank during times of credit restraint. This turned liquid assets into illiquid assets until the period requiring restraint had passed. Intended as a short-term instrument, special deposits were in place from 1965 until 1971, varying between 1% and 3½% for the banks (Artis 1978). Re-called in 1972, they remained outstanding for nearly a decade (Capie 2010). Credit ceilings, also intended to be short-term, were in place from 1955 to 1958, 1961 to 1962, and then again from 1965 to 1971.

Credit ceilings diverted lending to less-regulated finance houses and secondary banks. By 1966 bank deposits were just 1.6 times the size of wholesale deposits having been nine times larger in 1951 (Revell 1972). This loss of control coincided with the increasing influence of the International Monetary Fund (IMF) as Britain negotiated successive loan agreements (Goodhart and Needham 2017; Needham 2014c). The IMF was particularly critical of the Bank's operations in the gilts market. Because of the large volumes of debt maturing each year, the Bank believed it had to reduce price volatility to maintain the marketability of gilts. This was on the longstanding assumption that the market operated with extrapolative expectations; higher rates created expectations of even higher rates, unless buyers could be persuaded that the next move would be downwards. Therefore, the government broker "leant into the wind," smoothing downward price movements by buying gilts and selling on a rising market.

Leaning into the wind reduced the effectiveness of the liquidity ratio. If the banks were just above the minimum, they could rely on the government broker to provide cash against gilt ("secondary reserve") sales, knowing losses would be contained. Leaning into the wind was also at odds with the IMF's focus on Domestic Credit Expansion (DCE), since it is difficult to control a price (the interest rate) and a quantity (DCE) at the same time (Polak 1997). In 1968, the prospect of a DCE ceiling as a condition for further IMF lending prompted the Bank to undertake the most thorough monetary policy review since Radcliffe. The result was a paradigm shift that underpinned "the biggest change in monetary policy since the Second World War" – Competition and Credit Control in 1971 (Capie 2010).

The Bank believed it had identified a stable demand-for-money function and a causal relationship from broad money (M3) growth to nominal GDP (Goodhart and Crockett 1970). As Goodhart points out:

> The demand-for-money functions appeared to promise that credit and money could be controlled by price (interest rates), so that ceilings could be abandoned. Although some older and more experienced officials doubted all the econometrics (quite rightly as it happened), they wished to embrace this latter message (Goodhart 1984).

In September 1971, having withdrawn liquidity from all but short-dated gilts, the authorities removed credit ceilings and withdrew restrictions on lending to less-favored sectors. The liquidity ratio was replaced by a new Reserve Asset Ratio as the emphasis shifted to more active use of Bank Rate to control M3 growth.

Competition and Credit Control was not a success. Despite internal targets, M3 grew by 72% between March 1972 and December 1973 (Needham 2014b). This was largely because Prime Minister Edward Heath refused to let tighter monetary policy impede his "dash for growth." But the authorities had also been misled by demand-for-money equations (that broke down irretrievably in 1972) into setting their initial (unpublished) M3 target too high at 20% (Needham 2014a). The Bank fashioned a new instrument, Supplementary Special Deposits (the "corset") that imposed penalties on the growth of the banks' interest-bearing deposit liabilities. While the corset exerted cosmetic control over M3 growth (by increasing disintermediation), the Bank was again persuaded that it was impossible to exercise tight control over the broad money supply. Meanwhile the subsequent inflation, which peaked at 26.9% in August 1975, convinced senior Conservatives and their monetarist advisers that there *was* a robust causal link between M3 growth and inflation. The result was the Medium-Term Financial Strategy (MTFS), launched in 1980, which laid out a four-year series of declining £M3 targets and deficit ceilings.

The MTFS shared with Competition and Credit Control a belief in the primacy of the interest rate weapon. Exchange controls were dismantled in 1979, the corset was removed in 1980, and the Reserve Asset Ratio was abolished in 1981. Unlike Heath's Conservatives, the Thatcher government was less squeamish about high nominal interest rates, initially at least. Misled into believing there to be robust link between the growth of the broad money supply and inflation and that £M3 growth could be contained by higher interest rates, the government raised interest rates to the highest level in British history, 17% in 1979. Combined with a shift from direct to indirect taxation, this induced the deepest recession since the 1930s. Recognizing within weeks that monetary policy was causing unnecessary damage, the government began lowering interest rates despite the money supply overshooting its targets (Hoskyns 2000). Inflation did fall from 21.9% in May 1980 to 3.7% in May 1984. But this was as much a consequence of falling global commodity prices as restrictive UK monetary policy (Neild 2014). The government continued to publish monetary targets until the 1990s, but with the velocity of circulation unstable, this was little more than an 'intricate dance between the monetary authorities and market practitioners' with plenty of scope for discretion (Hotson 2010).

The failure of monetarism saw the authorities gravitate back towards an exchange rate target, with Chancellor Nigel Lawson "shadowing" the D-Mark from 1987 before sterling joined the Exchange Rate Mechanism (ERM) in 1990. This policy was abandoned on 16 September 1992 after a failed defense that cost the public purse over £3 billion (Tempest 2005). In search of a new anchor, the authorities alighted upon inflation targeting. This framework remains in place with the Bank (independent since 1997) charged with containing inflation at or around 2%. The primary instruments are Bank Rate and open market operations although there has been recourse to quantitative easing since the financial crisis of 2007–2008.

France

Napoleon established the Banque de France (the "Banque") in 1800 to manage the state's finances and supply credit, chiefly by discounting three-name commercial bills. His desire for a stable discount rate (initially 5%, then 4% from 1820 to 1847) guided French monetary policy along a path haunted by the inflationary spectres of the *assignats* and John Law's Mississippi Scheme. French citizens preferred to hold coin rather than bank deposits, and the Banque's notes circulated widely only when heavily backed by gold in the later nineteenth century.

The Banque's goal was to maintain the convertibility of its notes into silver at 5 g per franc (the *franc germinale*), subsequently also at 3445 francs per kilogram of gold after France was driven onto a *de facto* gold standard by the 1840s gold rushes. With no *legal* requirement to convert its notes into gold, the Banque was on a "limping gold standard" (Redish 2000). Nonetheless, gold became embedded in French political culture as a "key element in a larger system of classical liberal belief in the efficacy of free markets, free trade, and liberal government, in which balanced budgets and gold convertibility allowed the free play of market forces with minimal government interference" (Mouré 2002).

The Banque deployed two principal instruments: discounting (eligible bills and collateralized "lombard" loans) and the gold premium – raising its gold selling price which, unlike the Bank of England's, was not fixed by statute (Bazot et al. 2016; Mouré 1991). The Banque eschewed open market operations before the First World War, believing they interfered with the self-equilibrating mechanism of the gold standard. It did, however, deploy devices, sometimes offering interest-free advances to gold importers and purchasing specie near the frontier to reduce transit costs (Bloomfield 1959; US National Monetary Commission 1910; White 1933). French banks were not required to observe reserve and liquidity ratios until after the Second World War (Casella and Eichengreen 1993; Patat and Lutfalla 1986). As Governor Pallain explained in 1907:

> In France we consider that the strength of a bank consists more in the composition of its portfolio, i.e., in the value of its commercial bills, rather than in the importance of its cash reserve. . .the proportion of cash to liabilities is less significant on account of the facilities offered by the organization of the Bank of France for the rapid conversion – in a crisis – of a good portfolio into ready money (US National Monetary Commission 1910).

In keeping with the Napoleonic desire for stable interest rates, the Banque changed its discount rate infrequently, preferring to raise the gold premium (before 1900) in response to interest rate rises elsewhere (Bazot et al. 2016). Between 1880 and 1913, the discount rate changed just 30 times (versus 194 changes in London and 116 in Berlin) and only three times to defend against gold losses (Mouré 1991; White 1933). After 1900, the Banque's reserves took the strain:

> the extent of our reserves allows us to contemplate without emotion important variations of our metallic stock, and we only exceptionally have recourse to a measure which is always painful for commerce and industry (US National Monetary Commission 1910).

Current account surpluses allowed the Banque to increase its gold hoard from 2 billion francs in 1880 to 4 billion francs in 1914 (25.22 francs equaled £1) (Kindleberger 1984; White 1933). Since it was less able than the Bank of England to draw gold from the periphery, this provided insurance against international monetary disturbances. It also provided a war chest in the event of a revanchist war with Germany (Mouré 2002).

France entered the First World War with public debt at 65% of GDP, higher than the UK and in line with Germany (Hautcouer 2005). Income tax had just been introduced, and there was little political appetite to raise it. Indeed, less than 18% of the war was paid for by taxation (Mouré 2002). Monetary policy therefore had a large role to play. The Banque's policy from the outset was to conserve rather than spend its gold. This meant suspending convertibility in August 1914 and "harvesting" nearly half the domestic circulation of coin with patriotic calls for citizens to hold notes on the premise that convertibility would eventually be restored at the prewar rate. The harvest was so successful that, despite sending 3 billion francs of gold abroad, mainly to support the exchange rate, the Banque *increased* its reserves during the war, from 4.1 billion to 5.5 billion francs (Mouré 2002).

By the Armistice, the Banque had increased its note issue from 6.8 billion to 30 billion francs, advanced 21 billion francs (*c.*12% of the cost of the war) to the government, and discounted 3.5 billion *bons de la défense nationale*. With just 12% of banknotes backed by gold, and prices *c.* 250% higher, internal convertibility at the prewar level was a distant goal (Mouré 2002). External convertibility at the wartime rate of 27 francs to the pound ended in March 1919 with the withdrawal of British and American support. With more than 70% of its balance sheet comprised of public-sector liabilities, the Banque insisted that the state begin repaying its debts and agreed with Finance Minister Frédéric François-Marsal in 1920 that 2 billion francs would be repaid annually over 13 years. This "convention" immediately ran into political difficulties, when it became clear that German reparations would not pay for French reconstruction. The government had continued to borrow from the Banque in anticipation of the Germans paying as promptly as the French had after Waterloo and defeat by Prussia in 1871 (Occhino et al. 2006). In the absence of reparations, successive governments "staggered from deficit to deficit," unable to agree which taxes should be raised (Bouvier 1988). Deficits were monetized as short-term debt holders responded to the threat of a capital levy by running down their holdings and exporting capital (Eichengreen 1992). In 1926, with annualized inflation at 346%, Prime Minister Raymond Poincaré raised taxes, funded the floating debt, and secured the foreign credits that allowed the franc to be fixed at an undervalued 20% of its prewar gold value in 1928 (Mouré 2002; Makinen and Woodward 1989).

The French (and the Americans) have been criticized for aggravating the Great Depression by hoarding sterilized gold (Accominotti 2009; Irwin 2012; Johnson 1997; Mouré 2002). Certainly, the Banque welcomed the increase in its gold reserves between 1927 and 1932 (from 7 to 27% of global reserves), blaming others for not deflating to prevent outflows (Irwin 2012). It did lower its discount rate from 3.5 to 2% in 1930–1931, in step with other central banks, in order to stem inflows. But after refusing to follow New York and London lower in early 1931, the Banque raised its

rate to 2.5% after sterling floated, believing this would "hasten a healthy liquidation and return to prosperity" (quoted in Mouré 1991). With the Banque unwilling to engage in extensive open market operations, the discount rate remained above market rates.

Charles Rist commented that France, "still a cheap country in 1930, had become a dear country by 1931" (quoted in Caron 1979). Political instability and budget deficits combined with the overvalued currency to drain the Banque's reserves. After Pierre Laval's 1935 deflation failed to balance the budget, Léon Blum's Socialist government brought the Banque under closer state control and raised the ceiling on treasury bill discounting. The government also raised domestic costs with the *Accords de Matignon* and decided that the Banque's gold would be better spent on rearming than supporting the franc. On 1 October 1936, following the Tripartite Agreement with Britain and the United States, gold convertibility ended and the franc was fixed against sterling. The proceeds of this revaluation financed the new *Fonds de Stabilisation des Changes*, charged with maintaining the franc at 105.15 to the pound. Within a year, however, continued political instability and budget deficits (financed by the Banque) drove the franc above 150. Devaluation to 179 in May 1938 finally allowed the authorities to pursue a cheap money policy and the discount rate was lowered to 2½%. Only then did the Banque overcome its longstanding distaste of open market operations which boosted earnings depressed by low discount rates (and lending to the Treasury at 0.2%). By decreasing discounting and "hitching up" market rates with open market operations, the Banque had belatedly assumed the character of a modern central bank.

After the fall of France in June 1940, the Germans introduced their new currency (the *Reichskreditkassenschein*) exchangeable at the Banque at an overvalued rate of 20 francs) and extracted a total of 862.5 billion francs (*c.* 75% of GDP) from the Vichy regime (Milward 1970; Occhino et al. 2006). The regime had no access to international capital markets and the Banque's gold had been shipped abroad, so nearly a third of the cost was borne by creating money in the accounts of the occupying forces at the Banque. As the Germans transferred these balances to pay French suppliers, the Banque paid out its own notes which displaced the *Reichskreditkassenschein*. Coupled with the undervalued franc, this generated inflationary pressure, so the regime raised taxes, imposed financial repression, and constructed a *politique de circuit* to sterilize money creation with bond sales. The private banks were encouraged to take up public-sector debt such that by the Libération, *c.* 90% of their balance sheets comprised short-term public debt (Occhino et al. 2006). Open market operations were deployed to keep interest rates low, with long-term rates maintained at around 3%. The policy was moderately successful, with inflation contained within a range of 17–24% during the occupation.

In 1945, French public debt stood at *c.* 235% of GDP, with short-term debt *c.* 61% (Casella and Eichengreen 1993). This monetary overhang was reduced with inflation (averaging 52% per annum between 1945 and 1948) and continued financial repression. The Banque and the four largest commercial banks were nationalized in 1945 and in 1946 the *Conseil National du Crédit* was instituted to administer credit controls, alongside the wartime *Commission de Contrôle des Banques* that regulated

the banks. The principal goals of postwar French monetary policy were to secure economic growth in line with the priorities of the *Commissariat Général du Plan* while maintaining the internal value of the franc and the external Bretton Woods parity (initially 119.1 francs to the dollar, devalued to 350 by 1949) (Patat and Lutfalla 1986). This first goal led the French authorities to hold the discount rate and the banks' cartelized rates below market-clearing rates, bringing the market "into the Banque." This required exchange controls and a dirigiste array of instruments to achieve price and currency stability. These built upon the wartime *politique de circuit* and included discounting, liquidity and reserve ratios, and open market operations.

Despite heavy reliance on rediscounting in the quarter-century after the Second World War, the Banque kept rates stable relative to other central banks (Dieterlin and Durand 1973). Until 1957 there were no limits on discounting Treasury bills with maturities of 3 months or less, albeit the government itself was prohibited from discounting directly (Hodgman 1974; Monnet 2018). The Banque did, however, impose ceilings (*plafonds de réescompte*) on banks eligible to rediscount commercial bills and three-name medium-term credits. While rediscounting ceilings were applied individually to each bank, the Banque sometimes moved all ceilings simultaneously (Monnet 2014). From 1951, the Banque also operated safety valves, permitting the banks to enter repurchase agreements at the *taux d'enfer* ("evil rate," which varied from 1 to 3% above the discount rate), with further discounts incurring the *taux de super-enfer* ("super-evil rate" which varied from 3 to 7% above the discount rate). In 1967 these were replaced by a fixed rate *pension* of 2.5% on discounts above ceilings. This *pension* was abolished alongside the *plafonde de réescompte* in 1972 as the Banque shifted the emphasis to open market operations.

In 1947 the banks were required to hold at least 60% of demand deposits in liquid assets – cash, Treasury bills, and commercial bills (Casella and Eichengreen 1993; Dieterlin and Durand 1973). From 1948 until 1956, they were also required to maintain short-term government holdings at 95% of the 1948 level – the "Treasury floor" (*plancher*), while investing at least 20% of increased deposits in short-term government debt. From 1956 to 1967 they had to maintain a percentage of their deposits in short-term government paper, declining from 25% in 1956 to 5% before the *plancher* was abolished in 1967. This had little to do with maintaining prudential liquidity; it was to prevent the monetization of the national debt while providing the government with cheap finance, initially for postwar reconstruction (Patat and Lutfalla 1986). It also restrained private-sector lending.

From 1961 until 1967, the banks were also required to maintain a liquidity ratio against total deposits. The *coefficient de trésorerie* comprised cash and short-term claims (against the Banque, Treasury, and Giro system) and medium-term paper exempt from the discount ceiling (Dieterlin and Durand 1973). This *coefficient* fluctuated between 30 and 36%. By 1967, balanced budgets had reduced the Treasury bill issue and steps were taken to create a more liquid money market (Hodgman 1974; Monnet 2014). The *plancher* and *coefficient de trésorerie* were replaced by a minimum reserve ratio requiring the banks (from 1967 until 1998) to

hold proportions of their sight and time deposits in non-interest bearing accounts at the Banque, and the *coefficient de retenue* which (until 1985) required a proportion of total deposits to be held in eligible medium-term liabilities (Quintyn 1993). To provide another brake on credit, between 1971 and 1987 French banks were also required to hold a percentage of their loans in non-interest bearing deposits at the Banque (Hodgman 1974).

The banks were also subject to ceilings on private-sector and nationalized industry lending – the *encadrement du crédit*. Initially intended as a short-term instrument in 1958, ceilings were applied intermittently until 1970, and continuously (with monthly or quarterly norms) from 1973 to 1985 (Argy 1983; Argy et al. 1990; Hodgman 1974; Monnet 2014; Patat and Lutfalla 1986). Until 1970, breaches were punished by reducing the bank's rediscount ceiling; from 1970, and until ceilings were abolished in 1987, banks posted a proportion of excess lending as Supplementary Reserves at the Banque. From December 1974, unused lending capacity below the ceiling could be carried forward 6 months (Quintyn 1993).

Finally, the Banque conducted open market operations in Treasury notes and eligible bills. These were used more actively from 1971 and displaced discounting as the principal instrument by the mid-1980s, reflecting official willingness for more flexible rates (Chouraqui 1981; Hodgman 1974).

In line with their foreign counterparts, and the recommendation of the European Community's Monetary Committee, the French authorities responded to rising inflation with internal monetary targets from 1973 to 1975 (Argy 1983; Cobham and Serre 1986). In October 1975, they published a 13% M2 growth target for 1976. M2, defined as currency in circulation plus total bank deposits, was chosen because it emerged from the forecasts as a residual after estimating the growth of the following counterparts:

(i) Net gold and foreign exchange
(ii) Net claims on the Treasury
(iii) Bank lending to the economy
(iv) Banks' stable resources
(v) Adjustment items

After 1958 the budget was usually in surplus so the largest counterpart was bank lending (Raymond 1983). The principal instrument, therefore, was the *encadrement*, albeit the largest overshoot was driven by the 1975 budget deficit when the Chirac government "disregarded" the M2 target in its efforts to tackle unemployment (Cobham and Serre 1986; Galbraith 1982). The focus returned to inflation with the Barre Plan in 1976, with M2 targets declining from 12.5% in 1977 to 10% in 1981.

Consistent with his attempts to boost the economy after winning the presidency in 1981, François Mitterrand raised the target to a range of 12.5–13.5% before two franc devaluations, a large current account deficit, and the prospect of another approach to the IMF forced the government down a path of economic *rigueur*. Marsh identifies this as "a turning point in the chronicle of European money" (Marsh 2009). The M2 growth target for 1983 was set at 9%, and after a third

devaluation (coordinated with the Germans), the franc remained within the ERM, rekindling hopes of monetary union.

Monetary targets are theoretically incompatible with exchange rate targets. In practice, currency reserves, exchange controls (abolished in 1990), two-tier exchange rates (1971–1974), wider exchange rate bands (1993–1999), and the array of instruments detailed above, allowed the Banque to retain some monetary autonomy even as the franc was fixed within the Bretton Woods system, the "snake," and the ERM (Hodgman 1974; Chouraqui 1981). That monetary autonomy was ceded to the European Central Bank (ECB) in January 1999. Dyson and Featherstone explain the French enthusiasm for monetary union as a "three-level game" (Dyson and Featherstone 1999). Integrated capital markets would drive the modernization of French industry. Second, monetary union would bind the German leviathan by "Europeanizing" monetary policy. Finally, as Germany could insulate herself from American unpredictability on defense by sharing French military capability, France could insulate herself from American economic unreliability by sharing German monetary credibility. The results have been mixed, with lower growth and higher unemployment outweighing the gains from market integration and an independent ECB. Inflation has remained low, however, with the ECB deploying a familiar battery of open market operations and minimum reserve requirements in pursuit of its inflation target of "below, but close to, 2% over the medium term."

Germany

The Reichsbank was chartered in 1875 "to regulate the monetary circulation of Germany, to facilitate clearings, and to see to it that available capital is productively employed" (quoted in Capie et al. 1995). As well as providing credit to the commercial sector, the Reichsbank was expected to provide state finance (Schneider 2018). This was reflected in its management since, despite being privately owned, the Council (*Bankkuratorium*) was chaired by the Reichskanzler and it was governed by a Directorat (*Reichsbankdirektorium*) appointed by the Kaiser (Holtfrerich 1988; US National Monetary Commission 1910).

The Reichsbank's sometimes-conflicting monetary goals were (i) convertibility, (ii) low interest rates, (iii) stable interest rates, and (iv) profitable discounting (Bopp 1964). The primary goal, as elsewhere, was to "maintain the value of the monetary unit as stable as possible" (Lexis quoted in Goodhart 1988). From 1873 the unit of account was the mark, containing 5 g of pure silver or 358 milligrams of pure gold. The Reichsbank's statutory note issue, initially 250 million marks (20.43 marks equaled £1), had to be covered by (at least) one-third legal tender money, the remainder by eligible commercial bills (Holtfrerich 1988; Singleton 2011). Notes above the statutory limit had to be fully backed by specie and incurred a 5% tax. In practice, the Reichbank held excess specie so, like the Banque de France, it could use its balance sheet to meet internal and external drains. The statutory note issue was raised periodically to accommodate the growing economy (Giovannini 1986).

The principal instrument was discounting, at three different rates. First, the rate available to persons and firms (66,681 in 1910) eligible to discount commercial bills and checks of maximum 3 months signed by at least two ("as a rule," three) solvent names (Bopp 1964). Second, the Lombard rate (usually ½-1% higher) applied to loans of maximum 3 months collateralized by "moveable value objects" and intended to accommodate seasonal demand (Northrop 1938). Finally, there was a "preferential" discount rate that was mainly used to compete for discount business (Bopp 1964; Northrop 1938).

Berlin had a more developed money market than Paris and the Reichsbank changed its discount rate more frequently than the Banque, confident this was "the only effective instrument for regulating the domestic demand for money" and influencing international gold flows (Bopp 1964; Eichengreen 1992). German banks held little cash. As Deutsche Bank Director Paul Mankiewitz explained, "we can at all times convert our holdings of commercial paper into cash at the Reichsbank" (US National Monetary Commission 1910).

Bloomfield identifies four episodes when the Reichsbank used open market operations to withdraw liquidity before the First World War (Bloomfield 1959). He finds no evidence of open market operations intended to ease monetary conditions or to offset international gold flows. The Reichsbank also deployed devices, imposing a premium on gold sales for export, paying out gold in Berlin only (rather than the port cities of Hamburg and Bremen), granting interest-free advances to gold importers, and delivering lightweight coins to exporters (Bopp 1964; McGouldrick 1984). It could also use moral suasion, discouraging the commercial banks from discounting below published rates if it feared gold exports, and dissuading them from demanding gold during the 1907 crisis.

Convertibility was suspended with war in August 1914. The requirement that one-third of the note issue (now *Papiermark*) be backed by specie remained, but the Reichsbank was authorized to include Treasury bills in the remaining two-thirds (Holtfrerich 1988). The tax on note issuance above the statutory limit was abolished. To facilitate war finance, the Reichsbank's loan bureaux began accepting public-sector securities in exchange for notes (*Darlehnskassenscheine*, not backed by specie) that could be used to settle tax liabilities and as backing for further Reichsbank note issuance (as if they were specie). By the end of the war, *Darlehnskassenscheine* comprised nearly double the "gold" backing of the Reichsbank issue (Northrop 1938; Gross 2014). The key problem, however, was the Berlin money market's capacity to absorb short-term government debt (Balderston 1989). The wartime rise in German debt to GDP was lower than in the UK. But Germany could not access US lending and had a less liquid bond market, so more of its borrowing was short term and more of it was monetized at the central bank (Burret et al. 2013).

The debate over German hyperinflation in the 1920s is framed by those who emphasize exchange rate depreciation and those who blame budget deficits. The former, including the wartime Minister of Finance and Reichsbank President, believed the threat of higher taxation to pay reparations (and occupation costs in the Ruhr, under Franco-Belgian occupation from 1923 to 1925 owing to non-payment) stimulated capital flight (James 1999). This generated inflation via a

depreciated exchange rate. The latter view held that "the fantastic demands" of the victors generated fiscal deficits that were monetized at the Reichsbank (Northrop 1938). Eichengreen weaves these arguments together to show how the inflationary impulse generated by the lower exchange rate impacted, and was accelerated by, the budget deficit:

> Since nominal revenues were less responsive than nominal expenditures to changes in the rate of inflation, an inflationary shock magnified the size of the budget deficit, requiring additional Reichsbank monetization to finance the shortfall. Monetization fuelled the inflation, aggravated the revenue shortfall, widened the deficit, and reinitiated the process. Only measures to restore stability in the foreign exchange market, such as a drastic reduction of reparations, could halt the explosive spiral (Eichengreen 1992).

Reparations payments underpin both explanations, as does the political instability that precluded agreement on how to finance the deficits. Socialists advocated property taxes and a capital levy; conservatives wanted lower expenditure and higher sales taxes. Despite a tax compromise and a 25% reduction in scheduled reparations payments in 1922, the Weimar government ran deficits of 35% in 1922 and 60% in 1923 (Webb quoted in Eichengreen 1992). Investors doubted the government's ability to service long-term debt, forcing the state to issue short-term bills which the Reichsbank, fearing social upheaval, continued to discount (Holtfrerich 1988). Moreover, to fulfill "state necessities," the Reichsbank kept discount rates unchanged at 5% until July 1922, despite annualized inflation above 500% (Webb 1984).

In October 1923 the government cut public-sector salaries and employment and ended subsidies to the Ruhr in an effort to balance its budget. The Reichsbank ceased discounting treasury bills and a new currency, the *Rentenmark*, backed by a 3.2 billion *Goldmark* mortgage on private property (later by Reichsbank gold), was issued by the new, independent *Rentenbank (*Kindleberger 1984; Peterson 1954). *Rentenmarks* were not legal tender but circulated widely before being replaced by Reichsmarks after October 1924. Reichsmarks were convertible into gold at the prewar rate and exchanged for *Papiermarks* at a rate of 1 trillion to one. In 1924 annual reparations were reduced to *c.* 1% of GDP under the terms of the Dawes Plan, taking further pressure off the government deficit. This stabilization, along with Franco-Belgian withdrawal from the Ruhr, encouraged the successful floatation of an 800 million Reichsmark ("Dawes") loan that ushered in further capital inflows (Kindleberger 1984; Schuker 1988).

The Dawes Committee also helped design the more independent Reichsbank that recommenced operations in 1924 (Singleton 2011). Although there was no legal requirement for convertibility until 1930, at least 40% of the note issue had to be covered by gold (three-quarters) and foreign exchange (one-quarter), with the remainder backed by eligible three-month (or less) domestic and foreign commercial bills (Northrop 1938). A subsidiary goal, unsurprising in the wake of hyperinflation, was price stability. A further goal, inherited from the prewar Reichsbank, was "to direct domestic and foreign capital to 'productive' ends" at low and stable interest rates. Discounting remained the principal instrument, with no limits on eligible

commercial bills and checks even if the Reichsbank was prohibited from owning government debt (except in its own pension fund) until 1926. Capital shortage, and the need to cover the note issue and reparations payments, meant the Reichsbank generally maintained its discount rate above other central banks in the late 1920s. This kept the discount rate above money market rates, and therefore "ineffective," forcing the Reichsbank to rely on other instruments.

The Reichsbank was prohibited from open market operations by the 1924 Bank Act. The legal position of its *Rentenbank* and *Golddiskontobank* subsidiaries was less clear and the authorities took advantage, issuing *Golddiskontobank* bills in December 1926 to tighten the money markets (Northrop 1938). The Reischbank also rationed discounting from 1924 to 1926, during the early summer of 1929, and again in June 1931 in response to the banking crisis sparked by the failure of the Kreditanstalt.

The interwar German gold standard effectively ended with the imposition of capital controls in July 1931. With Reichsmarks no longer freely convertible, the Reichsbank could operate a cheap money policy and reduced its discount rate from 15% to 4% after September 1932. This period marked the "transition from a free economic order to a controlled and political state" (Northrop 1938). This was formalized by the 1934 Reich Credit Law that placed the financial sector under the Supervisory Office for Credit chaired by Reichsbank President Schacht. The Office dictated the size of loans to individual borrowers, as well as the banks' cash ("not more than 10%") and liquidity reserves (James 1999). From 1933 the Reichsbank could engage in open market operations, with "work creation bills" an increasing source of public and rearmament finance. The maturity limit on eligible bills was still 3 months, so the *Golddiskontobank* was enlisted to discount six-month bills drawn by arms manufacturers and accepted by the *Metallurgische Forschungsgesellschaft m.b.H* (*MeFo*, backed by four major arms manufacturers) (Abelshauser 1998; James 1999). The credit banks, forbidden from buying government bonds before 1933, became "sponges" for public-sector debt (Wolf 1955). Inflation was tackled with the blunt instrument of Gauleiter Joseph Wagner's Office for Price Formation and statements such as "the first cause of the stability of our currency is the concentration camp: the currency stays stable when anyone who asks higher prices is arrested" (quoted in James 1999). In 1939, after refusing the government a loan, the Reichsbank was brought under direct control with loan and discounting volumes dictated by the Nazi regime. As James points out, this turned it, once again, into "a perfect machine for inflation" (James 1999).

Taxes brought in just 11% of government spending in 1944–1945, insufficient to pay for civil expenditure let alone military operations (James 1999). Private savings, channeled to the state via the banks, were also insufficient so the government issued short-term debt that was again monetized at the Reichsbank, increasing the note issue from 11 billion in 1939 to 73 billion Reichsmarks by 1945. Official prices, held down by rationing and fiat, rose just 11% between 1939 and 1944. By then barter and the black market had increasingly displaced normal monetary exchange. The release of pent-up inflation and the 1948 currency reform saw German savings largely wiped out for the second time in a generation.

In 1948 the *Bank deutscher Länder* (BdL) replaced the Reichsbank in the Western occupied zones. The BdL's capital came from the *Landeszentralbanken*, which were coordinated on the US Federal Reserve model. Under Allied supervision until 1951, the BdL was intended to be free of domestic political influence (Holtfrerich 1999). Its goals were "to stabilize the currency and the monetary and credit system." Maintaining the exchange rate of the D-mark, introduced in 1948, was the responsibility of the Federal Government. This was achieved by accumulating foreign exchange reserves through current account surpluses, while keeping interest rates low to facilitate reconstruction. The banks were required to hold minimum reserves set in 1948 at 10% for sight deposits and 5% for time and savings deposits. Since no interest was paid on reserves, the banks tended to use any excesses for credit creation. The BdL therefore monitored excess reserves, reducing reserve requirements in September 1949, for instance, when they fell to 7.1% (Holtfrerich 1999). The BdL also monitored the "volume of money" (*Geldvolumen*) defined as currency in circulation outside the banking system and the nonbanks sight (less than 30-day) deposits. The principal instruments were variable minimum reserve requirements (capped at 20%), discounting (bills of exchange and Lombard credits), variable discount quotas, and open market operations. The BdL also issued quantitative and qualitative lending guidance. This was in pursuit of the monetary mercantilism summarized by Holtfrerich:

> Strategy aimed, by means of a restrictive monetary policy and a greater degree of price stability in Germany than abroad, to force the German economy into exports and thus simultaneously to strengthen the D-mark and to promote economic growth and employment (Holtfrerich 1999).

The Bundesbank replaced the BdL in 1957 and was charged with regulating "the amount of money in circulation and of credit supplied to the economy with the aim of safeguarding the currency" (quoted in Hetzel 2002). With the D-mark fixed under the Bretton Woods arrangements, it sought to stabilize the price level by targeting banks' free liquid reserves – the combination of excess reserves held at the Bundesbank above the minimum requirement, unused discounting and Lombard quotas, eligible domestic money market paper, and short-term foreign assets. Targeting the proportion of free liquid reserves to bank deposits (the liquidity ratio) shifted the focus from *current* liquidity (minimum and excess reserves on existing deposits) to *potential* liquidity (future credit creation). This rested on the belief that banks sought to maintain stable liquidity ratios. The instruments remained largely the same with credit ceilings rejected in 1967 and again in 1973 (Holtfrerich 1999; von Hagen 1999).

With inflation lower in Germany than most of its major trading partners, conflict between the goals of internal price stability and maintaining the Bretton Woods parity produced persistent balance of payments surpluses. Inflows, initially through the current account but increasingly through the capital account after *de facto* convertibility in 1955, accounted for *c.*90% of the increase in central bank money between 1950 and 1970 (Holtfrerich 1999). Emminger referred to this as the

"substitute central bank" (*Erstatz-Notenbank*). This was monetary mercantilism's "open flank" and it threatened imported inflation. The 1957 Bundesbank Act therefore raised the limit on the reserve ratio on foreign deposits (which paid no interest) to 30% (raised to 100% in 1968). There were also currency swaps for the domestic banks, central bank swaps from 1961, and an increasing variety of devices (such as "voluntary" capital controls on US dollars) to encourage net capital exports. Nonetheless, capital inflows and the threat of imported inflation produced D-mark revaluations in 1961 (5%) and 1969 (9.3%) before the currency floated against the dollar, from May to December 1971, then again after March 1973.

Holtfrerich writes: "as early as the mid-1960s it was clear that the policy of controlling the banks' free liquid reserves had failed" (Holtfrerich 1999). Indeed, lending grew rapidly in 1971–1972 despite low liquidity ratios. By 1973, the Bundesbank had "virtually lost control of the money supply" (von Hagen 1999). This was despite a new instrument to deter inflows, a cash deposit on foreign loans (the *Bardepot*), set initially at 40% in 1972. In line with Mundell's theory, floating provided greater scope for monetary autonomy by closing the "open flank" of the fixed exchange rate (Mundell 1962). Also, from 1973, foreign currency assets no longer counted as central bank money (defined as currency in circulation plus the required minimum reserves on domestic deposits). This reduced excess reserves nearly to zero, allowing the Bundesbank to operate a policy of "money stock control with zero free liquid reserves" from mid-1973. The price of slowing credit growth by restricting reserves was volatile overnight interest rates, touching 40% and threatening the banking system. The conflict between tight control over central bank money and ensuring the banks could meet their reserve requirements was resolved with a medium-term target for central bank money, based on existing reserve requirements. The Bundesbank hoped to guide the banks along a medium-term path for credit growth by varying interest rates to control the medium-term growth of reserves.

The first central bank money growth target, 8% in 1975, represented a compromise between those concerned that lower interest rates should not be interpreted, especially by wage bargainers, as loosening the commitment to price stability, those who desired "cover" against political interference, and those who wanted a rules-based approach. Recognizing that monetary policy is more art than science, the targets (target ranges from December 1978) were set by committee with reference to potential growth and an "unavoidable" (later "normative") rate of inflation, sometimes with adjustments for changes in the velocity of circulation and previous developments (Baltensperger 1999).

Despite overshooting its first three targets, the Bundesbank restated its commitment to monetary targets ahead of the ERM launch in 1979. With the external value of the D-Mark in political hands, the Bundebank wished to emphasize that ERM membership would be consistent with price stability (Keegan et al. 2017). This pragmatic monetarism recognized the importance of discretion rather than precisely hitting monetary targets. The Bundesbank wanted "to make the aims of monetary policy clearer to labor and management, whose cooperation is essential if inflation is to be brought under control without detrimental effects on employment" (quoted in

Beyer et al. 2009). Pragmatic monetarism was in part, therefore, a "disguised incomes policy" (Hetzel 2002). Despite the Bundesbank's mandate to "regulate the amount of money in circulation," excess monetary growth was tolerated if inflation was in check – "monetary stabilization with discretionary elements" (Baltensperger 1999).

The ERM provided a less open flank than Bretton Woods partly because, until 1987, members were more willing to realign parities. Nonetheless, hitching the D-mark to weaker currencies such as the franc and the lire was consistent with monetary mercantilism. As one Finance Minister admitted: "the key principle of German economic policy was to persuade the French and Italians to pay to lower the value of the D-Mark so as to make Germany more competitive" (Healey 1989; Keegan et al. 2017). Chancellor Helmut Schmidt agreed, explaining that "one effect" of the ERM "would certainly be to weaken the German mark" (quoted in Mourlon-Druol 2012).

The instruments remained largely the same, with more flexible (and less "noisy") open market operations superseding discount and Lombard (and "special Lombard") rates, minimum reserve ratios, and rediscounting quotas. Increasing capital flows, particularly through the Euromarkets (free from reserve requirements), allowed German banks to circumvent minimum reserve ratios which were progressively reduced from 1985 to 2% for savings and time deposits and 5% for sight deposits by 1994 (Baltensperger 1999). Repurchase agreements, which accounted for just 6% of central bank money creation in 1980, were the principal instrument by the end of the decade.

Lower interest rates in 1986–1988 induced Germans to hold more cash, and this was held partly responsible for the monetary target overshoots. In 1988 the Bundesbank shifted to targeting M3. At this stage, the ERM was working well for the Germans. As Heisenberg points out:

> German businesses were in the internationally unique position of having their cake and eating it too – with the Bundesbank setting interest rates according to the exigencies of the German economy and the EMS limiting exchange rate fluctuations with their trading partners. The cost of occasional revaluations of the DM under the EMS regime was far less than the potential costs of monetary union with France (Heisenberg 2005).

Nonetheless, the Germans did agree to monetary union, despite the Bundesbank's reservations. Unlike the French, who argued that monetary union could drive economic convergence, the Bundesbank insisted that the single currency should crown economic union. But "no central bank, not even the Bundesbank, can detach itself from its environment" (Baltensperger 1999). Former Chief Economist Issing identifies the currency crises of 1992–1993 as the turning point when "I and others came to the conclusion that the Common Market would not survive another crisis of this dimension" (quoted in Marsh 2009). The survival of the European project, and with it the goal of German reunification, rested upon monetary union (Needham 2017). This required an ECB committed to price stability and "independent of

instructions from national governments and Community authorities" – something the epistemic community of central bankers and independent experts that had gathered in Basle, away from the politicians, to produce the Delors Report could agree (Delors 1989). Without an independent ECB, there would have been little German support for monetary union; without German support, there would be no monetary union. In many ways, the Germans are still "having their cake and eating it" with an ECB modeled on the Bundesbank singularly committed to the longstanding German monetary goal of price stability, and the Euro holding German export prices lower than they would be under a reconstituted D-mark.

Conclusions

The holy grail of monetary policy is "a conditional rule that combines flexible reactions to short-term disturbances with a credible commitment to medium-term price stability" (Baltensperger 1999). For more than a century, monetary knights have banded together in search of this prize, sometimes convinced it was within reach, more often lost in the thickets of incomplete data, reverse causation, and long and variable time lags. In the late nineteenth-century, they could mostly agree that:

> [A central bank's] main objective was to maintain convertibility of the currency into gold, the gold standard; its main control instrument was to vary interest rates to that end; it made its interest rate effective by discounting bills and, increasingly by open market operations (Capie et al. 1995).

The First World War changed everything. Patterns of trade that had underpinned the international monetary system shifted, as did the domestic supports of the "three pillars of the anti-collectivist temple" – the gold standard, free trade, and minimal balanced budgets. Interwar policy makers might be forgiven their desire to rebuild the world of 1913, even if their views of the pre-war monetary system were clouded with nostalgia. Each country's post-WW2 monetary policy was shaped in part by its interwar experience, in part by the legacy of war. All three fixed their currencies under the terms of the Bretton Woods regime. But the experience of hyperinflation caused Germans to prioritize price stability higher than the British, for whom the "searing experience" of the interwar years was unemployment (Budd 2014). The Germans, having once again defaulted, also had less public sector debt to contend with. Germany revalued while Britain and France devalued. As the Bretton Woods system crumbled in the late 1960s, attention turned to the money supply. All three nations operated internal monetary targets in the early 1970s, publishing these within months of each other in 1974–1976. In the 1980s, as the false god of monetarism was exposed, they retreated to the familiarity of fixed exchange rates, albeit with varying enthusiasm and success before alighting upon inflation targeting in the 1990s. As the Euro crisis has demonstrated, the single currency is not the holy grail. But holy grails are never easy to find.

Cross-References

- ▶ European Monetary Integration
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ The Historical Evolution of Central Banking

References

Abelshauser W (1998) Germany: guns, butter, and economic miracles. In: Harrison M (ed) The economics of World War II. Cambridge University Press, Cambridge, pp 122–176

Accominotti O (2009) The sterling trap: foreign reserves management at the Bank of France, 1928–1936. Eur Rev Econ Hist 13(3):349–376

Accominotti O (2012) London merchant banks, the central European panic, and the Sterling crisis of 1931. J Econ Hist 72(1):1–43

Allen WA (2014) Monetary policy and financial repression in Britain 1951–59. Palgrave Macmillan, Basingstoke

Argy VE (1983) France's experience with monetary and exchange rate management: march 1973 to end-1981. Banca Nazionale del Lavoro Quartr Rev 36(147):387–411

Argy VE, Brennan A, Stevens GR (1990) Monetary targeting: the international experience. Econ Rec 66(192):37–62

Artis MJ (1978) Monetary policy: part II. In: Blackaby FT (ed) British economic policy, 1960–74. Cambridge University Press, Cambridge

Balderston T (1989) War finance and inflation in Britain and Germany, 1914-1918. Econ History Rev 42(2):222–244

Baltensperger EB (1999) Monetary policy under conditions of increasing integration (1979–96). In: Bundesbank D (ed) Fifty years of the Deutsche mark: central bank and the currency in Germany since 1948. Oxford University Press, Oxford, pp 439–523

Bank of England (1962) Bank liquidity in the United Kingdom. Bank of Engl Quart Bull 2(4):248–255

Bazot G, Bordo MD, Monnet E (2016) International shocks and the balance sheet of the Bank of France under the classical gold standard. Explor Econ Hist 62:87–107

Beyer A, Gaspar V, Gerberding C, Issing O (2009) Opting out of the great inflation: German monetary policy after the breakdown of Bretton Woods, European Central Bank working paper series 1020. Dt. Bundesbank, Frankfurt

Bloomfield AI (1959) Monetary policy under the international gold standard: 1880–1915. Federal Reserve Bank of New York, New York

Bopp KR (1964) Reichsbank operations. Federal Reserve Bank of Philadelphia, Philadelphia

Bordo MD, Kydland FE (1995) The gold standard as a rule: an essay in exploration. Explor Econ Hist 32(4):423–464

Bouvier J (1988) The Banque de France and the state form 1850 to the present day. In: Toniolo G (ed) Central banks' independence in historical perspective. Walter de Gruyter, New York, pp 73–104

Broadberry SN, Howlett P (1998) The United Kingdom: victory at all costs. In: Harrison M (ed) The economics of World War II: six great powers in international comparison. Cambridge University Press, Cambridge

Budd AP (2014) The London Business School and the 1981 budget. In: Needham DJ, Hotson AC (eds) Expansionary fiscal contraction: the Thatcher government's 1981 budget in perspective. Cambridge University Press, Cambridge, pp 37–55

Burret HT, Feld LP, Köhler EA (2013) Sustainability of public debt in Germany – historical considerations and time series evidence. J Econ Stat 233(3):291–335

Capie FH (2010) The bank of England: 1950s to 1979. Cambridge University Press, New York
Capie FH, Fischer S, Goodhart CAE, Schnadt N (1995) The future of central banking: the tercentenary symposium of the Bank of England. Cambridge University Press, Cambridge
Caron F (1979) An economic history of modern France. Methuen, London
Casella A, Eichengreen BJ (1993) Halting inflation in Italy and France after the second world war. In: Bordo MD, Capie FH (eds) Monetary regimes in transition. Cambridge University Press, Cambridge
Checkland SG (1983) British public policy, 1776–1939: an economic, social and political perspective. Cambridge University Press, Cambridge
Chouraqui J-C (1981) Monetary policy and economic activity in France. In: Courakis AS (ed) Inflation, depression and economic policy in the west. Mansell & Alexandrine, London, pp 203–216
Cobham DP, Serre J-M (1986) Monetary targeting: a comparison of French and UK experience. Roy Bank Scot Rev 149:24–42
Committee on currency and foreign exchanges after the war ('Cunliffe Committee') (1918) First interim report, cmnd 9182. HMSO, London
Committee on Finance and Industry ('Macmillan Committee') (1931) Report, cmnd 3897. HMSO, London
Committee on the working of the monetary system ('Radcliffe Committee') (1959) Report, cmnd 827. HMSO, London
Committee for the Study of Economic and Monetary Union ('Delors Committee') (1989) Report on economic and monetary union in the European Community. Office for Official Publications of the European Communities, Luxembourg
Daunton MJ (2007) Wealth and welfare: an economic and social history of Britain, 1851–1951. Oxford University Press, Oxford
Dieterlin P, Durand H (1973) Monetary policy in France. In: Holbik K (ed) Monetary policy in twelve industrial countries. Federal Reserve Bank of Boston, Boston, pp 117–160
Dyson KHF, Featherstone K (1999) The road to Maastricht: negotiating economic and monetary union. Oxford University Press, Oxford
Eichengreen BJ (1992) Golden fetters: the gold standard and the great depression, 1919–1939. Oxford University Press, Oxford
Fletcher GA (1976) The discount houses in London: principles, operation and change. Macmillan, London
Galbraith JK (1982) Monetary policy in France. J Post Keynesian Econ 4(3):388–403
Giovannini A (1986) Rules of the game' during the international gold standard: England and Germany. J Int Money Financ 5(4):467–483
Goodhart CAE (1984) Monetary theory and practice: the UK experience. Macmillan, London
Goodhart CAE (1988) The evolution of central banks. MIT Press, Cambridge MA
Goodhart CAE, Crockett AD (1970) The importance of money. Bank Engl Quartr Bull 10:159–198
Goodhart CAE, Needham DJ (2017) Historical reasons for the focus on broad monetary aggregates in post-WW2 britain and the "seven years war" with the IMF. Fin Hist Rev 24(3):331–356
Gross SG (2014) War finance (Germany). In: International encyclopedia of the First World War. https://doi.org/10.15463/ie1418.10020
Hautcouer P-C (2005) Was the great war a watershed? The economics of World War One in France. In: Broadberry SN, Howlett P (eds) The economics of World War One. Cambridge University Press, Cambridge
Healey DW (1989) The time of my life. Methuen Publishing Ltd, London
Heisenberg D (2005) Taking a second look at Germany's motivation to establish economic and monetary union: a critique of 'economic interests' claims. German Polit 14(1):95–109
Hetzel RL (2002) German monetary history in the second half of the twentieth century: from the deutsche mark to the euro. Fed Res Bank Richmond Econ Quartr 88(2):29–64
Hodgman DR (1974) National monetary policies and international monetary cooperation. Little, Brown and Company, Boston
Holtfrerich C-L (1988) Relations between monetary authorities and governmental institutions: the case of Germany from the 19th century to the present. In: Central banks' independence in historical perspective edited by Gianni Toniolo. Walter de Gruyter, New York, pp 105–159

Holtfrerich C-L (1999) Monetary policy under fixed exchange rates (1948-70). In: Bundesbank D (ed) Fifty years of the Deutsche mark: central bank and the currency in Germany since 1948. Oxford University Press, Oxford, pp 307–401
Hoskyns JAHL (2000) Just in time: inside the Thatcher revolution. Aurum Press, London
Hotson AC (2010) British monetary targets, 1976 to 1987: a view from the fourth floor of the Bank of England, London school of economics financial markets group special paper 190. London School of Economics and Political Science, Financial Markets Group, London
Hotson AC (2017) Respectable banking: the search for stability in London's money and credit markets since 1695. Cambridge University Press, Cambridge
Howson SK (1975) Domestic monetary management in Britain 1919–38. Cambridge University Press, Cambridge
Howson SK (1980) Sterling's managed float: the operations of the exchange equalisation account, 1932-39. Princeton Stud Int Fin 46:1–66
Howson SK (1993) British monetary policy, 1945–51. Oxford University Press, Oxford
Irwin DA (2012) The French gold sink and the great deflation of 1929–32. Cato Papers Pub Pol 2:1–56
James H (1999) The Reichsbank 1876-1945. In: Fifty years of the Deutsche mark: central bank and the currency in Germany since 1948 edited by Deutsche Bundesbank. Oxford University Press, Oxford, pp 3–53
Janeway WH (1995-1996) The 1931 sterling crisis and the independence of the Bank of England. J Post Keynes Econ 18(2):251–268
Johnson HC (1997) Gold, France and the great depression, 1919–1932. Yale University Press, London
Keegan WJG, Marsh DW, Roberts RW (2017) Six days in september: black wednesday, Brexit and the making of Europe. OMFIF Press, London
Keynes JM (1925) The economic consequences of mr. churchill. Hogarth Press, London
Kindleberger CP (1984) A financial history of Western Europe. Allen & Unwin, London
Makinen GE, Thomas Woodward G (1989) A monetary interpretation of the Poincaré stabilization of 1926. South Econ J 56(1):191–211
Marsh DW (2009) The Euro: the politics of the new global currency. Yale University Press, London
Matthews RCO, Feinstein CH, Odling-Smee JC (1982) British economic growth, 1856–1913. Oxford University Press, Oxford
McGouldrick PF (1984) Operations of the German central bank and the rules of the game. In: Bordo MD, Schwarz AJ (eds) A retrospective on the classical gold standard, 1821–1931. University of Chicago Press, Chicago, pp 311–360
Middleton R (2002) Struggling with the impossible: sterling, the balance of payments and British economic policy, 1949-72. In: Arnon A, Young WL (eds) The open economy macromodel: past, present and future. Kluwer Academic Press, Boston
Milward AS (1970) The new order and the French economy. Oxford University Press, Oxford
Moggridge DE (1972) British monetary policy, 1924–31: the Norman conquest of $4.86. Cambridge University Press, Cambridge
Monnet E (2014) Monetary policy without interest rates: evidence from France's golden age (1948 to 1973) using a narrative approach. Am Econ J 6(4):137–169
Monnet E (2018) French monetary and credit policy, 1945–1973. Cambridge University Press, Cambridge
Morgan EV (1952) Studies in British financial policy 1914–1925. Macmillan, London
Mouré K (1991) Managing the Franc Poincaré: economic understanding and political constraint in French monetary policy 1928–1936. Cambridge University Press, New York
Mouré K (2002) The gold standard illusion: France, the Bank of France and the international gold standard, 1914–1939. Oxford University Press, Oxford
Mourlon-Druol E (2012) A Europe made of money: the emergence of the european monetary system. Cornell University Press, London

Mundell RA (1962) The appropriate use of monetary and fiscal policy for internal and external stability. IMF Staff Pap 9(1):70–79
Needham DJ (2014a) UK monetary policy from devaluation to Thatcher, 1967–1982. Palgrave Macmillan, Basingstoke
Needham DJ (2014b) The 1981 budget: a Dunkirk not an Alamein. In: Needham DJ, Hotson AC (eds) Expansionary fiscal contraction: the Thatcher government's 1981 budget in perspective. Cambridge University Press, Cambridge, pp 148–180
Needham DJ (2014c) Goodbye Great Britain? The press, the treasury, and the 1976 IMF crisis. In: Roberts RW, Schifferes S (eds) The media and financial crises: comparative and historical perspectives. Routledge, London, pp 261–276
Needham DJ (2017) Snakes and ladders: navigating European monetary union. In: Cardinale I, Coffman D'M, Scazzieri R (eds) The political economy of the Eurozone. Cambridge University Press, Cambridge, pp 129–152
Neild RR (2014) The 1981 statement by 364 economists. In: Needham DJ, Hotson AC (eds) Expansionary fiscal contraction: the Thatcher government's 1981 budget in perspective. Cambridge University Press, Cambridge, pp 1–9
Northrop MB (1938) Control policies of the Reichsbank, 1924–1933. Columbia University Press, New York
Occhino F, Oosterlinck K, White EN (2006) How occupied France financed its own exploitation in World War II, NBER working paper 12137. National Bureau of Economic Research, Stanford
Officer LH (1996) Between the dollar-sterling gold points: exchange rate, parity and market behaviour. Cambridge University Press, Cambridge
Patat J-P, Lutfalla M (1986) A monetary history of France in the twentieth century. Macmillan, London
Peterson EN (1954) Hjalmar Schacht, for and against Hitler: a political-economic study of Germany, 1923–1945. Christopher Publishing House, Boston
Polak JJ (1997) The IMF model at forty, IMF working paper. 97/49. Internatational Monetary Fund, Washington, DC
Quintyn M (1993) Managing monetary policy reforms. Lessons from the French experience. Banca Nazionale del Lavoro Quartr Rev 186:275–298
Raymond R (1983) The formulation and implementation of monetary policy in France. In: Meek P (ed) Central bank views on monetary targeting. Federal Reserve Bank of New York, New York, pp 105–117
Redish A (2000) Bimetallism: an economic and historical analysis. Cambridge University Press, Cambridge
Revell JRS (1972) A secondary banking system. In: Johnson HG et al (eds) Readings in British monetary economics. Oxford University Press, Oxford
Roberts RW (2014) Saving the City: the great financial crisis of 1914. Oxford University Press, Oxford
Sayers RS (1956) Financial policy, 1939–45. HMSO, London
Sayers RS (1976) The bank of England, 1891–1944, vol 1 and 2. Cambridge University Press, Cambridge
Schneider S (2018) Imperial Germany, Great Britain and the political economy of the gold standard, 1867-1914. In: Hoppit J, Leonard AB, Needham DJ (eds) Money and markets: essays in honour of Martin Daunton. Boydell and Brewer, Woodbridge
Schuker SA (1988) American "reparations" to Germany, 1919-33: implications for the third-world debt crisis. Princeton Stud Int Fin 61:1–170
Schumpeter JA (1952) Essays on economic theory. J.C.B. Mohr, Tübingen
Singleton J (2011) Central banking in the twentieth century. Cambridge University Press, Cambridge
Tempest, Matthew 2005. Treasury papers reveal cost of black wednesday Guardian, 9. February 2005

Trentmann F (2009) Free trade nation: commerce, consumption, and civil society in modern Britain. Oxford University Press, Oxford

United States National Monetary Commission (1910) Interviews on the banking and currency systems of England, Scotland, France, Germany, Switzerland, and Italy. Government Printing Office, Washington DC

von Hagen J (1999) A new approach to monetary policy (1971-78). In: Bundesbank D (ed) Fifty years of the Deutsche mark: central bank and the currency in Germany since 1948. Oxford University Press, Oxford, pp 403–438

Webb SB (1984) The supply of money and Reichsbank financing of government and corporate debt in Germany, 1919–1923. J Econ Hist 44(2):499–507

White HD (1933) The French international accounts, 1880–1913. Harvard University Press, Cambridge, M.A

Wolfe M (1955) The development of Nazi monetary policy. J Econ Hist 15(4):392–402

The Evolution of US Monetary Policy

35

Robert L. Hetzel

Contents

Abstract

Since the creation of the Federal Reserve System, the goal of policymakers has been economic stability. Policymakers' strategies for achieving that goal have evolved with their understanding of how the world works. An overview of that understanding and of its consequences for monetary policy provides an approximation to a laboratory for understanding what constitutes a stabilizing monetary policy. As an institution, when has the Fed been a major contributor to economic stability, and when has it been a major source of instability? This laboratory provides guidance in the construction of a model that allows for identification of the forces that drive prices and the business cycle. A model allows one to go beyond the correlations of monetary and macroeconomic variables in order to

Robert L. Hetzel has retired.

R. L. Hetzel (✉)
Federal Reserve Bank of Richmond, Richmond, VA, USA
e-mail: roberthetzel@comcast.net

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_32

assign causation. It explains how "exogenous" forces, that is, forces emanating from outside the working of the price system, move markets away from stable outcomes. The historical overview here suggests that monetary policymakers still have not settled on a model and a rule for policy that satisfactorily distils the lessons from historical experience. Much work remains in order to achieve consensus on the design of a rule that will make monetary policy into a consistently stabilizing influence.

Keywords
Monetary policy · Federal Reserve System · Monetary regime · History of federal reserve · Central banks

Introduction

Since the creation of the Federal Reserve System, the goal of policymakers has been economic stability. Policymakers' strategies for achieving that goal have evolved with their understanding of how the world works. An overview of that understanding and of its consequences for monetary policy provides an approximation to a laboratory for understanding what constitutes a stabilizing monetary policy. As an institution, when has the Fed been a major contributor to economic stability and when has it been a major source of instability?

This laboratory provides guidance in the construction of a model that allows for identification of the forces that drive prices and the business cycle. A model allows one to go beyond the correlations of monetary and macroeconomic variables in order to assign causation. It explains how "exogenous" forces, that is, forces emanating from outside the working of the price system, move markets away from stable outcomes. The historical overview here suggests that monetary policymakers still have not settled on a model and a rule for policy that satisfactorily distils the lessons from historical experience. Much work remains in order to achieve consensus on the design of a rule that will make monetary policy into a consistently stabilizing influence.

Section "Defining the Monetary Regime" poses the questions, "What is a central bank and how does the systematic behavior of a central bank create the monetary regime?" Section "Pre-World War II Monetary Policy" summarizes the pre-World War II monetary regime, while Section "Post-World War II Monetary Policy and Stop-Go" summarizes the era of stop-go monetary policy. Section "The Great Moderation" reviews the post-disinflation Volcker-Greenspan era and the intellectual sea change that it entailed. Section "What Monetary Regime Did Volcker and Greenspan Create?" reviews the monetary regime created during the Volcker-Greenspan tenures as FOMC chairmen. Section "The Great Recession" reviews monetary policy during the Great Recession. The narrative summarizes Hetzel (2008, 2012). Section "Why Was Raising Inflation so Hard After the Great Recession?" examines the reasons why the Fed regularly missed its inflation target on the downside after the Great Recession. Section "What Is the Monetary Regime?" offers concluding comments.

Defining the Monetary Regime

What makes central banks unique is their control over money creation. The operating procedures of central banks entail the manipulation of bank reserves in order to set a targeted interest rate on those reserves – the federal funds rate. Bank reserves are the instrument for achieving finality of payment. The FOMC can manipulate their supply in order to achieve its interest rate target because reserves possess a well-defined demand. Bank demand for reserves varies with the quantity of the instruments that serve as media of exchange (money), for example, demand deposits.

Starting October 8, 2008, the Fed began paying interest on reserves (IOR). Prior to that date, the New York Desk achieved the FOMC's funds rate target by reducing reserves (selling securities) when the funds rate fell below target and conversely by increasing reserves (buying securities) when the funds rate rose above target. With IOR, the Desk must supply the minimum amount of reserves consistent with the funds rate target. It then effectively sterilizes any additional reserves provided by open market operations as banks voluntarily hold them to receive the interest paid by the Fed.

The nature of the monetary regime depends upon the procedures the central bank employs to discipline money creation through its control over bank reserve creation. In a monetary regime of fiat money, the value in exchange of a paper dollar far exceeds its resource cost of production. In order to limit the quantity of money and give it a well-defined value in exchange for goods (a well-defined price level), the central bank must provide a nominal anchor. Markets cannot perform that function. The nature of the monetary regime then emerges from the interaction of that nominal anchor and the extent to which the central bank either allows the price system to operate to determine the values of real variables or impinges upon its operation in an attempt to trade off between inflation and real variables like unemployment.

In the quantity theory tradition descended from Irving Fisher and Milton Friedman, a credible rule allows the Fed to separate the determination of the price level from the determination of relative prices and real quantities like output and unemployment. Under the assumption that markets are competitive and that they work well in a regime of nominal (price level) expectational stability, it is optimal for the central bank to follow a rule that turns over to the price system the operation of the real economy – a nonactivist monetary regime. The alternative tradition, which descended from John Maynard Keynes, contends that the exercise of monopoly power in markets forces trade-offs on the central bank. The central bank can mitigate extreme fluctuations in inflation through depressing output, and conversely it can mitigate extreme declines in output through increasing inflation – an activist monetary regime. These trade-offs, which are known under the rubric of Phillips curve trade-offs, require the Fed to intervene in the operation of price system.

Given the existence of recessions and inflation variability, the underlying question is whether the economy is inherently stable but subject to periodic episodes of instability arising from mismanagement of the monetary regime by the central bank. Alternatively, is it inherently subject to periodic episodes of instability? In the

former, quantity theory view of the world, periodically the central bank, interferes with the operation of the price system and as a consequence creates destabilizing emissions and absorptions of money that cause the price level to evolve in an unpredictable fashion. In the latter, Keynesian view of world, markets are periodically prone to "herd" behavior that overwhelms the stabilizing properties of the price system. Such behavior could take the form of an inflationary investment boom or of a speculative excess that creates ultimately unsustainable imbalances, the collapse of which entails deflation and recession. Both of these two kinds of destabilizing forces, monetary mismanagement and private excess, can produce economic instability. However, the former are preventable through adoption of an optimal monetary rule, while the latter are subject only to mitigation by the central bank through exploiting a Phillips curve.

Sections "Post-World War II Monetary Policy and Stop-Go" and "The Great Moderation" summarize the evolution of the monetary regime since the early days of the Fed. In assessing the evolution of the monetary regime, the key benchmark will be the attitude of the Fed toward inflation. Does the Fed accept responsibility for inflation and, as an essential prerequisite for its control, the control of inflationary expectations? Because monetary policymakers operate in financial markets, they have always understood the critical role of expectations. But, the question remains, expectations of what?

Pre-World War II Monetary Policy

The conception of a central bank responsible for macroeconomic stabilization is a post-World War II phenomenon. Before modern macroeconomics, there was no understanding of the behavior of the economy determined by the operation of the price system with the real (inflation-adjusted) interest rate the key intertemporal price of resources. With no understanding of the price system, there could be no understanding that the price system worked well in the sense of establishing well-defined values of real variables that equilibrate markets, in particular a "natural" rate of interest. Before the war, there was then no conception of monetary policy understood as the mutual interaction between the setting of a policy instrument (the interest rate on bank reserves or the funds rate) and the behavior of the economy. As a result, there could be no sense that this mutual interaction was a procedure for letting the price system work by causing the real interest rate to track the natural rate of interest.

Before the founding of the Fed, the United States was on the gold standard. After the end of the Second Bank of the United States in 1836, the United States had no central bank, and the gold standard operated automatically. The Bank of England, the central bank for the world on the gold standard, operated with a set of rules that required no understanding of the price system. Nothing prepared the early Fed for the responsibilities of the monetary regime that Congress created in 1912 (Hetzel 2017).

After the creation of the Fed, the United States retained convertibility of gold and Federal Reserve notes. However, the Fed did not follow the rules of the international

gold standard for setting its discount rate. In order to understand the consequences of that departure from gold standard rules, it is useful to survey the operation of the gold standard. The classical gold standard provided a "nominal anchor" in that it imparted a well-defined value to the price level. Given the gold content of a dollar, the real (commodity) value of gold determined the price level. Given the coexistence of paper money (banknotes and greenbacks) and gold dollars, convertibility meant that each had to possess the same purchasing power. Under the gold standard, if the value of gold in exchange for commodities increased, the value in exchange of a paper dollar also had to increase. That increase required a decline in the price level (deflation).

For the countries belonging to the international gold standard that prevailed prior to World War I, the real value of gold set a world price level. At the same time, given the respective gold content of those countries' currencies, the resulting fixed parities in their currencies (exchange rates) could not adjust in order to equilibrate their balance of payments. As a result, the price levels of countries relative to each other had to adjust. For example, if a country had a good harvest at the time of a worldwide poor harvest, its exports of agricultural commodities would cause a balance of payments surplus, gold inflows, and an increase in its price level relative to the world price level. Its terms of trade would increase.

Although the classical gold standard provided a nominal anchor, the nominal anchor was not stable because the real value of gold fluctuated with the demand for gold such as occurred with the move to the international gold standard at the end of the nineteenth century and with the supply of gold such as occurred with gold discoveries in Australia and California. Quantity theorists like Irving Fisher wanted to change the nature of the monetary regime by making stability in an index number of prices (the dollar price of a basket of commodities) the nominal anchor. However, that departure from monetary orthodoxy for a regime of "managed money" was considered heretical by the established financial and political order. In fact, as explained below, with no understanding of what it was doing, the early Fed created the very regime that the established order considered heretical – a regime in which a central bank controlled fiat money creation.

The reform movement that led to the creation of the Fed centered not on a desire to replace the gold standard but rather on the desire to eliminate the financial panics that disrupted trade. The movement combined two traditions: real bills and American populism (Hetzel 2014). Both traditions assumed that the concentration of reserves in New York led to the speculative extension of credit and asset bubbles, the collapse of which produced recession and deflation. The motivating principle behind the federal organization of the Fed then was to eliminate the concentration (pyramiding) of reserves in the large New York banks by dispersing them among independent depositories of member banks' reserves.

Congress passed legislation creating the Federal Reserve in December 1912. America's populist tradition meant that creation of a central bank was unacceptable (Lowenstein 2015). As of 1912, the past experience with central banks in the United States was with the First and Second Banks of the United States, and the contemporaneous experience was especially with the Bank of England. These central banks

had enforced the gold standard. The Bank of England imposed a common interest rate in the money markets based on how gold flows, internal and external, disturbed the adequacy of its gold reserve (Hetzel forthcoming a). That kind of control over the banking system from a central location like New York was not acceptable politically.

Applied to an individual bank, the real bills doctrine was a principle intended to assure that a bank could always meet deposit withdrawals because its assets were liquid. Holding liquid assets meant holding the self-liquidating debt that arose in the process of producing goods and bringing them to market (real bills or in English parlance bills of exchange). The founders of the Fed used this logic to understand bank panics in an American context. That is, they intended to design a system in which bank lending would vary with the demand for short-term productive uses of credit and would eschew long-term capital projects that were inherently speculative. In the National Banking Era, which preceded the Fed, banks that issued banknotes had to collateralize them with US Treasury bonds. However, the limited supply of those bonds made banknote circulation "inelastic." The presumption was that the speculative excess that preceded financial panics arose because this inelastic supply created an excess of credit during periods of recession and that excess then produced subsequent booms based on speculative excess (Hetzel 2014).

Applied to the design of the Fed, when combined with American populism, the logic of real bills led to the creation of regional Reserve Banks that would supply Federal Reserve credit (the deposits that the commercial banks held with it) only through the discount window and only by accepting real bills. In that way, the Reserve Banks would proportion the extension of credit to the legitimate needs of commerce. Discount window administration based on real bills principles would create an "elastic" supply of credit and currency in response to "legitimate" demands for credit and supposedly eliminate the excesses that periodically spilled over into the New York financial markets. With no excess reserves and credit to spill over into speculative excess, there would be no more of the financial panics whose collapse led to recession and deflation (Hetzel 2014).

In the spirit of real bills, the intended role of the Fed was to limit the total credit created by banks to the amount required for productive uses. Its role was to control the amount and nature of financial intermediation. The antithesis of this view was purposeful money creation, which forced money and credit on markets regardless of any productive outlet. These views made it impossible for early policymakers to understand the nature of the monetary standard that they had created.

Equally inimical to such understanding were misperceptions of the nature of the gold standard. As implemented by the Bank of England, the gold standard provided for a nominal anchor and a value of the real interest rate set in the international gold market. It did so by moving its discount rate in a way that maintained over time a steady reserve ratio (the ratio of gold to its liabilities in banknotes). As explained above, the price level varied in a way that kept the real value of the paper pound equal to its gold counterpart in coinage. Although the United States maintained convertibility of the paper dollar into gold, it abandoned these rules of the international gold standard.

Convertibility of the dollar into gold created the misleading presumption that the United States was on the "gold standard." However, the United States was not on the classical international gold standard but rather was pegging the price of gold through a commodity-price stabilization scheme of buying and selling gold at a fixed price. The behavior of the quantity of money was not linked to the behavior of bank reserves through gold flows produced by the international balance of payments. The price level was not then inversely related to the commodity value of gold. Despite gold convertibility, the early Fed operated a regime of fiat money creation. There was never any understanding by early Fed policymakers that the monetary standard they operated determined the behavior of the price level.

By law, the regional Reserve Banks did have to maintain gold reserves against their note issue. However, because of the large influx of gold into the United States from the start of World War I through the early 1920s, this gold cover requirement was never binding. Contrary to the principles of the international gold standard, the Reserve Banks did not lower their discount rates when their holdings of gold exceeded the required gold cover. Even though the regional Fed Banks did not operate according to the principles of the international gold standard, on three occasions when the gold cover requirement threatened to bite due to gold outflows, they raised their discount rates. They did so in late 1919 when gold flowed back to Europe, in fall 1931 when Britain went off the gold standard, and in winter 1933 when newly elected President Roosevelt failed to commit to maintenance of the gold standard. On those three occasions, gold outflows caused the Fed to engineer a severe monetary contraction.

In reality, the actual monetary regime was a fiat money regime. In contemporaneous terms, it was a regime of "managed money." However, the idea of such a regime was anathema to policymakers. They understood such a regime as one in which the Fed would force Federal Reserve credit onto markets rather than proportioning it in response to the legitimate demands for credit. It was a subversion of real bills principles. There was a quantity theory tradition, which related money creation to the price level. However, in the eyes of policymakers, because of its association with "managed money," the theory represented what real bills principles were intended to eliminate.

The historical associations of paper money creation were uniformly negative. They included the French experience with the assignats prior to Napoleon, greenbacks in the Civil War, and, closer to contemporaneous events, the post-World War I hyperinflations in Germany, Austria, and Hungary. Also, the orthodoxy of the gold standard crowded out the quantity theory. There was no perceived need to understand the operation of a monetary standard in which the central bank controlled the price level through its control over money creation. There was then no need for the analytical distinction between the price level and relative prices and the responsibility of the central bank to control an index number measuring the price level (Hetzel 2017).

In order to understand early monetary policy, it is essential to distinguish between "money policy" as understood by early policymakers and "monetary policy" as understood conceptually today with the benefit of decades of subsequent theorizing.

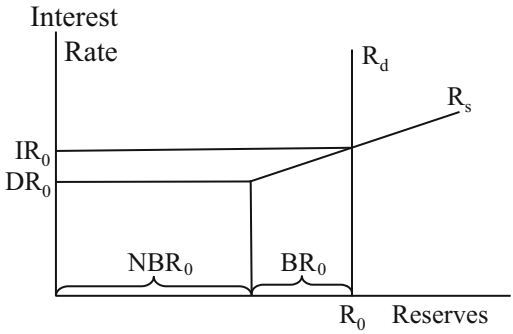


Fig. 1 Notes: R is bank reserves. R_d is the reserves demand schedule of the banking system and R_s the reserves supply schedule of the Fed. IR is the interest rate on bank reserves. DR is the discount rate. NBR and BR are nonborrowed and borrowed reserves, respectively. The 0's denote particular values

Consider first "monetary policy" in its modern sense. The distinguishing characteristic of a central bank is its control over money creation. There is a market for bank reserves, which the Fed creates through a bookkeeping operation and which serve as the media for effecting finality of payment. Bank reserves support a larger superstructure of means of payment (the quantity of money), employed by the public. Through its control over the quantity of bank reserves, the Fed controls the interest on those reserves, the funds rate. The funds rate and, more significantly, the way that markets forecast its future path based on the Fed's reaction function (response to incoming "news" on the economy) shapes the behavior of the term structure of interest rates in credit (money) markets. The Fed exercises that control through its role as money creator not through the role of a financial intermediary supplying funds obtained from the public, which is saving through the banking system.

Figure 1 summarizes the market for reserves that evolved in the 1920s. The reserves demand schedule, R_d, is vertical because banks can only adjust required reserves on deposits over time by adjusting loans and investments. The reserves supply schedule possesses two segments. The vertical segment measures the amount of nonborrowed reserves, NBR_0. The factors beyond the Fed's control that moved the nonborrowed reserves of banks were gold flows, currency in circulation, and Treasury deposits with the Fed. The Fed could manipulate nonborrowed reserves through open market purchases. Because the Fed set total nonborrowed reserves below reserves demand, collectively banks had to borrow from their Reserve Bank. The interest rate on reserves, IR_0, was then determined as the sum of the discount rate, DR_o, plus a nonpecuniary cost that varied positively with the amount of funds obtained from the discount window. Continuous borrowing was considered a sign of lending based not on real bills principles of self-liquidating loans but rather for speculative purposes. The Reserve Banks limited that borrowing through subjecting banks to stringent supervisory oversight if member banks borrowed continuously. Figure 2 shows how the commercial paper rate stayed above the New York Fed discount rate.

In World War I, the Fed kept the discount rate below the rate paid on Treasury securities so that banks would have an incentive to borrow from the window in order to buy the government debt issued to finance the war. Despite the resulting money creation, given its real bills' views, the Fed interpreted the inflation that followed the

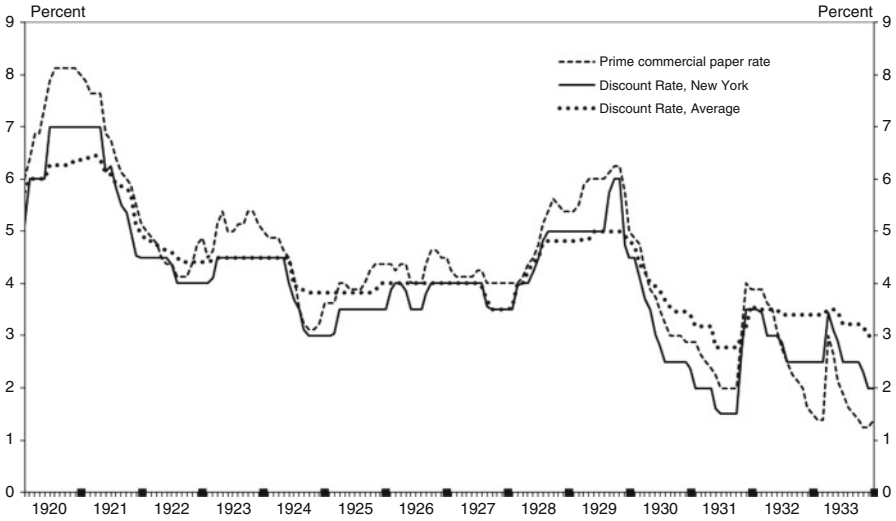


Fig. 2 Money market interest rates and regional Fed Bank discount rates. Notes: Monthly observations of the prime commercial paper rate, the New York Fed discount rate, and the average of the discount rates set by all other regional Reserve Banks. Heavy tick marks indicate fourth quarter. Source: Board of Governors *Banking and Monetary*

war as due to speculation that pushed up commodity prices. In response, it raised the discount rate in late 1919 and again in January 1920 until it reached high of 7%. The country entered a sharp recession in January 1920 but then recovered sharply starting in July 1921. The recovery started even with a discount rate still relatively high, close to 6%. Postwar reconstruction in Europe doubtlessly created strong credit demands and high market interest rates.

The Fed felt vindicated in its real bills views in that a strong recovery followed the recession. Just as had apparently occurred in prior episodes during the gold standard, following the collapse of the presumed speculative excess accompanied by a period of liquidation of speculative credit extension and the unsustainable level of economic activity that it supported, the economy entered the decade of the 1920s with sustained growth and only two minor recessions. However, the Fed was lucky the first time around. It would be unlucky in the Great Depression. In the 1920–1921 recession, because the international gold standard had not yet been reconstructed and foreign currencies floated against the dollar, contractionary monetary policy in the United States did not propagate to the rest of the world. In the Great Depression, with the reconstructed gold standard, it did propagate (Hetzel 2002a).

Prior to creation of the Fed, the operation of the price system determined market interest rates. However, with the creation of the Fed, in reality the United States had a central bank, which controlled short-term market interest rates. With no understanding of the consequences of its actions, and even with presumed "low" rates of interest as the Depression persisted, the Fed kept market interest rates above the level required for economic stability, the "natural" rate of interest.

As a result, the Fed created an unprecedented spiral of monetary contraction, deflation, expected deflation, a rise in the real rate of interest, and economic contraction (Friedman and Schwartz 1963). Nothing in the Fed's real bills framework prepared policymakers for an understanding of the monetary regime of fiat money in which they were operating.

The Great Depression started with the extremely contractionary monetary policy initiated by the Fed in 1928 intended to deflate the boom in equity prices on the New York Stock Exchange, widely considered "a financial debauch." (Insightful narratives of the Great Depression can be found in Chandler 1971; Friedman and Schwartz 1963; Meltzer 2003; Wheelock 1991; Wicker 1966.) As described above, the Fed had adopted procedures in which it set the marginal cost of funds to banks through the cost of borrowing from the discount window. That cost depended upon the discount rate and a nonpecuniary cost of using the discount window for banks that stayed in the window for any extended period. Starting in 1928, the Fed engaged in open market sales that reduced banks nonborrowed reserves and thus forced them into the discount window. In Fig. 1, the R_s schedule shifted leftward. The banks had to borrow extensively. Given the stricture that banks should never be in the window except for short periods, the stringent supervision associated with that borrowing at the window raised the cost of funds. Individually, banks attempted to get out of the window by restricting lending. Collectively, the banking system had to contract and money growth dropped to negative levels (Fig. 3).

At the same time, in the real bills environment of the time, the universally accepted view was that the recession that began in August 1929 had originated in a circumvention of real bills principles. The presumption was that a combination of gold inflows and easy money produced by the Fed in 1927 when it lowered the discount rate as a way of aiding Britain in its return to the gold standard had forced credit onto markets in excess of the legitimate needs of business. Forced credit creation had created a speculative bubble supporting an unsustainable level of economic activity. When, inevitably, the bubble burst, a period of liquidation of

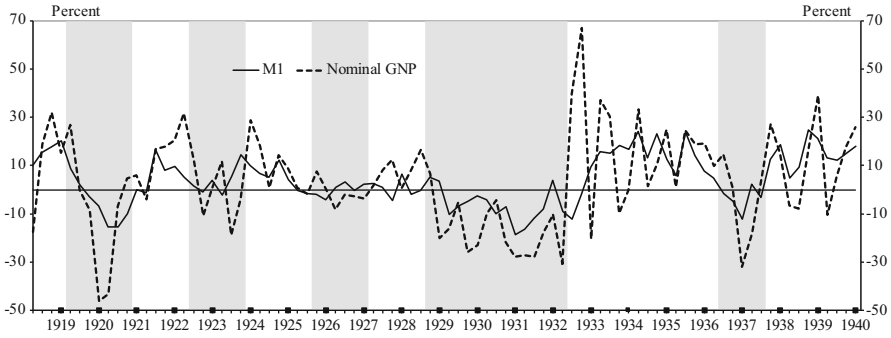


Fig. 3 M1 and nominal GNP quarterly growth rates. Notes: Quarterly observations of annualized quarterly nominal GNP growth and M1 growth. GNP series is from Balke and Gordon (1986, Appendix B). M1 is from Friedman and Schwartz (1970). Heavy tick marks indicate fourth quarter

credit and business had to occur before the economy could grow again on a more solid basis. That perception hobbled monetary policy. Expansionary monetary policy in the modern sense of aggressively buying government securities, forcing down interest rates, and creating money would, it was believed, have recreated the old boom-bust credit cycle that had led to the Depression.

Monetary policy was contractionary in the Depression because the Fed kept the marginal cost of reserves to banks above the natural rate of interest. As shown in Fig. 2, especially at the regional Reserve Banks other than New York, the discount rate, which set a lower bound on the cost of reserves, was maintained at levels well above zero. As the expectation of deflation grew, the real rate of interest rose to levels that were high for any phase of the business cycle. At the time, policymakers understood nothing of this analysis. The regional Reserve Banks considered themselves as depositories of funds whose use allowed them to influence the cost and availability of credit. The prevailing view was that a historically low level of the discount rate meant that the cost of credit was not an impediment to economic recovery. Although the Reserve Banks could increase the supply of credit through open market purchases, they were hesitant to do so. Real bills principles held that funds "forced" into credit markets would be used for speculation not productive purposes.

After 1930, the failure became obvious of the real bills' prediction that the economy would recover after the liquidation of speculative excesses. Because of their quotation in the memoirs of President Hoover (1952, p. 30), Secretary Andrew Mellon's expression of this "liquidationist" view has become famous:

> Mr. Mellon had only one formula: "Liquidate labor, liquidate stocks, liquidate the farmers, liquidate real estate." He insisted that, when the people get an inflation brainstorm, the only way to get it out of their blood is to let it collapse. He held that even a panic was not altogether a bad thing. He said: "It will purge the rottenness out of the system. High costs of living and high living will come down. People will work harder, live a more moral life."

Mellon's views represented economic orthodoxy.

In 1931, policymakers replaced real bills conceptions with actions appropriate for a financial panic as had occurred in the late nineteenth-century international gold standard. In response to an internal drain of currency out of banks into the hands of the public that began in late spring 1931 and in response to an external drain of gold that began after Britain abandoned the gold standard in September 1931, the Reserve Banks raised their discount rates. By 1931, however, without the nominal anchor provided by a gold standard, deflation exacerbated expected deflation, thereby raising real interest rates and sending the economy downward into a deflationary spiral. At the same time, policymakers interpreted their actions as successful because they had restored confidence in the continued convertibility of the dollar into gold. That interpretation of their actions in 1931 as successful reflected the way in which policymakers filled the void created by the failure of real bills views to assure sustainable growth. They viewed the business cycle as reflecting an internally driven alternation in the psychology of businessmen between optimism and pessimism.

In February 1932, the New York Fed, which was the de facto leader of the Reserve Banks, persuaded the Reserve Banks to undertake open market purchases. The issue then became whether to push to the point at which the resulting reserves creation would allow banks to repay their discount window borrowing and create excess reserves. As shown in Fig. 1, the result would have been to replace the existing system of controlling the cost of funds to banks as the sum of the discount rate and a surcharge positively related to discount window borrowing. Many of the regional Reserve Bank governors objected. By forcing banks into the window, the existing procedures offered the Reserve Banks the ability to control the allocation of lending to "legitimate" uses through the "administrative guidance" exercised over banks in the window. Moving from the world of real bills to one that looked like managed money was too great a leap.

In mid-1932, the commercial paper rate dropped below the discount rate (Fig. 2). At this point, because of open market purchases, the New York banks were out of the discount window and had accumulated sufficient excess reserves to be freed from having to fear reserves outflows that would push them back into the window. The resulting decline in money market rates spurred a brief revival in money growth and in economic activity. However, the New York Fed lost the consensus among the Reserve Bank governors it demanded in order to replace the existing real bills procedures with procedures that would allow unlimited reserves creation at a low or zero interest rate.

The New York Fed then backed down from its advocacy of open-ended purchases of government securities and reserves creation. Again, there was simply nothing in the intellectual and political environment of the time that could have pushed Fed policymakers into renouncing the lifetime they had spent understanding the role of the Federal Reserve as controlling financial intermediation on real bills principles. As they understood it, the alternative was to join forces with the populists who wanted to push the United States into a regime of fiat money creation with the social anarchy that entailed. Fed open market purchases ended in August 1932. In the last half of that year, gold inflows replaced open market purchases as a source of reserves creation. However, starting in February 1933, bank panics resumed and banks again became subject to an internal (currency) and external (gold) drain of reserves. Statewide bank holidays only spread the panic. Upon entering office in early March 1933, President Roosevelt declared a national bank holiday. It came with an implicit guarantee that reopened banks would be supported by the Fed and it restored calm.

Economic recovery only became possible with the advent of the Roosevelt administration and the transfer of control over monetary policy from the Fed to the Treasury that occurred in March 1933. By 1933, the public had come to associate the convertibility of the dollar into gold with deflation and deflation with recession. After March 1933, the Roosevelt administration's embargo on the export of gold and its unwillingness to revive the international gold standard produced a dramatic change in expectations from deflation to inflation. With the accompanying reduction in the real interest rate, the economy revived strongly in summer 1933 (Eggertsson 2008; Hetzel 2012).

In 1933, the Roosevelt administration based monetary policy on varying the dollar price of gold with the intention of reviving the price of agricultural exports. In January 1934, it again pegged the price of gold but at a depreciated value of $35 per ounce, up from $20.67 per ounce. Combined with political instability in Europe, gold flowed into the United States. Subject to Treasury control, the Fed kept constant its portfolio of government securities. Unable to sell securities and robbed of its control over market interest rates with banks out of the discount window, the Fed monetized the gold inflows. With the recovery in money growth, the economy recovered (Fig. 3).

With no understanding that monetary policy had been contractionary and had then turned expansionary, the Fed made no connection between the economic recovery and the change in procedures for controlling the access of banks to reserves. On the contrary, it watched with concern as banks accumulated excess reserves. From a real bills perspective, the concern was that excess reserves would again allow banks to expand credit beyond the legitimate needs of business. In summer 1936, the Fed convinced the Roosevelt administration to allow it to increase required reserves in order to immobilize bank use of the reserves. In that way, as business continued to recover, banks would again have to borrow from the Fed. With a revival of its old operating procedures, the Reserve Banks could resume their job of allocating credit to legitimate, productive uses. Banks, however, had desired those excess reserves as a safeguard against a revival of bank runs. After a succession of increases in required reserves that started in August 1936 along with the end of the monetization of gold inflows, money growth ceased, and the economy again entered into recession (Fig. 3).

Post-World War II Monetary Policy and Stop-Go

After 1937 and its ill-fated attempt to again take control of monetary policy, the Fed once more lost control to the Treasury. With the advent of the entry of the United States into World War II, the Fed pegged interest rates on Treasury securities. It pegged the interest rate of 3-month bills at 3/8%. That is, the Fed monetized all the government securities sold by the Treasury in excess of what the market demanded at the pegged rates. After the war, the Truman administration continued to dictate monetary policy, and the peg continued. When the peg did end, the Fed would have to reinvent monetary policy after almost two decades of having lost control to the Treasury and with real bills views intellectually bankrupt. Since 1933, bank credit had expanded based initially on gold inflows and, with the start of the war, with purchases of government securities. Policymakers looking through the lens of real bills would have expected a vast expansion of unsustainable economic activity built on the fragile scaffolding of an artificial expansion in credit. There should have been a huge depression after the war.

The problem after the war, however, was inflation not deflation and depression. Inflation soared with the end of price controls. In wartime, with the vast expansion of military expenditures, the problem was how to limit total aggregate demand

to the physical capacity of the economy to produce goods and services. With the change in focus on aggregate demand rather than on speculative excess as reflected in asset prices, the role of monetary policy changed to one of limiting the overall expansion of bank reserves and bank credit rather than prevention of a misallocation of credit to speculative uses.

The Korean War became the watershed event that led to the creation of the modern central bank. The entry of the Chinese into the war when their armies crossed the Yalu River in November 1950 created the widespread expectation of a World War III. That fear and the associated anticipation of the reimposition of price controls caused commodity-price inflation to soar. Given its rate peg, the FOMC watched painfully as it created reserves that fed the inflation. Clearly, the Fed rather than speculators was the problem. The Fed, not the marketplace subject to real bills restrictions, had to control reserve creation. The Fed forced the issue in early 1951 when it refused to support the price of long-term Treasury bonds. Given the unpopularity of President Truman with Congress, Congress refused to support the administration. The result was the Treasury-Fed Accord of March 1951. When the new appointee as chair of the Board of Governors and the FOMC, William McChesney Martin, set his own course independent of the Treasury, the Fed regained its independence (Hetzel and Leach 2001a, b).

The back-to-back occurrence of the Great Depression with its sustained high unemployment and World War II with its sustained low unemployment changed the intellectual and political environment by placing responsibility for economic stability on the federal government. *The Employment Act of 1946* assigned to government the responsibility "to promote maximum employment, production, and purchasing power." Later, the *Full Employment and Balanced Growth Act of 1978* explicitly assigned responsibility to the Fed "to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates."

The world did everything wrong in the first half of the century but did most things right in the second half. In addition to the creation of a modern central bank rather than a central bank based on real bills principles, the post-World War II monetary regime did not attempt to recreate a gold standard that forced deflation on its adherents. Instead, preparatory to making the Bretton Woods system operational, European countries devalued their currencies relative to the dollar and thus avoided deflation (Hetzel 2002b).

The Keynesian orthodoxy that developed after the war focused on deficit spending as the instrument for controlling aggregate demand and ignored monetary policy. As seen from the Fed, however, in a reversal of real bills, the problem was to control the growth in bank credit by making it available in recession and by limiting it in booms. Moreover, the Martin Fed concentrated on price stability as its measure of the appropriateness of aggregate demand.

Two conceptual developments laid the groundwork for a modern central bank. First, the Fed developed a national market for reserves in New York. The instrument of monetary policy became the interest rate set by the FOMC on those reserves. With the Fed's wartime and postwar peg of the interest rate on short-term government securities, the country did have a national market for bank reserves. Banks could buy

and sell Treasury bills among themselves in order to adjust their reserves positions. Under the general goal of developing a Treasury market with "depth, breadth, and resilience," the Fed strengthened the position of government securities dealers. The Fed developed the incipient procedures of the 1920s and 1930s summarized in Fig. 1 by routinely offsetting factors that drained and supplied reserves in order to stabilize interest rates in the money market (Roosa 1956; Madden 1959, p. 86).

The second conceptual development entailed behaving predictably. That is, monetary policy could be characterized as a reaction function summarizing how the Fed responded to incoming news about the economy. Martin implemented the "bills only" policy under which the Fed only bought and sold Treasury bills, that is, short-term securities. With such a policy, the term structure of interest rates (the entire yield curve) would move in a way that stabilized economic activity only if the FOMC set short-term interest rates in a predictable way in response to incoming information on the economy.

Martin termed Fed policy "lean against the wind," in which the FOMC moved short-term interest rates in a way designed to counter sustained weakness or strength in the economy. (Until the end of 1970, the FOMC looked at a complex of financial measures not just the funds rate to measure ease or tightness in the money market.) In response to the onset of the 1953 recession, the Martin FOMC began to lower market rates. What Hetzel (2008) termed lean against the wind with credibility took shape after inflation began to rise in mid-1956 and then persisted into the 1958 recession. Martin concluded that the Fed had to begin raising rates at the onset of economic recovery to stay on top of inflationary psychology. This policy foreshadowed the Volcker-Greenspan policy of placing nominal expectational stability at the center of monetary policy (Bremner 2004; Hetzel 2008).

Credibility meant that Martin accepted responsibility for inflationary expectations. Martin looked to the bond markets for evidence of inflationary psychology not to stock or commodity markets as had been the prewar case with real bills. He wanted to prevent the introduction of inflation premia in bond rates evidenced by sharp jumps in bond rates. That is, the Fed took control of responsibility for inflationary expectations. In this way, these procedures disciplined lean-against-the-wind changes in interest rates to cumulate to whatever degree necessary to maintain price stability. The FOMC moved toward a rule that created a stable nominal anchor based on the expectation of stable inflation.

Martin (U.S. Cong. 2/6/59, 462 and 467) testified to Congress:

> About this time [summer 1959] inflationary expectations began to spread. The abrupt upward shift of interest levels in central money markets . . . reflected investor demand for an interest premium to cover the risk of a depreciating purchasing power of invested funds. . .. The experience in the government bond market . . . is a vivid example of the influence of inflationary expectations in financial markets. To the extent that such attitudes come to be reflected in decisions on wages, prices, consumption, and investment, they help to bring about their own realization.

In the late 1950s and early 1960s, the United States suffered back-to-back recessions with cyclical peaks in August 1957 and in April 1960. Contractionary

monetary policy evidenced by cyclically low money growth preceded each. The first episode of contractionary policy arose in response to the inflation that emerged starting in 1956. The second arose in response to gold outflows. After the end of World War II, the Bretton Woods system of pegged exchange rates gradually became operational as countries eliminated controls on capital flows. By 1959, the dollar had become overvalued, and the United States ran a current account deficit. Both the Eisenhower Treasury and the Martin FOMC interpreted the gold outflows as a loss of confidence in the dollar. As a consequence of these two recessions, the country entered the 1960s with the expectation of price stability thoroughly embedded in the national psyche.

John Kennedy was elected president in 1960 on a platform of getting the country moving again. Under the leadership of Robert Heller, the Council of Economic Advisers (CEA) persuaded Kennedy to sign on to a national goal of 4% unemployment. However, in the context of the Cuban Missile Crisis, Kennedy held it in abeyance out of fear of a dollar crisis. The Martin FOMC and the conservative Douglas Dillon Treasury overrode the Heller CEA. After the Kennedy assassination in fall 1963, in the Johnson administration, however, Martin found himself isolated. In a stressed fiscal environment of guns and butter due to the Vietnam War and Great Society programs, the political system demanded that the economy grow flat out. In an environment of urban riots, it demanded a low unemployment rate, with 4% only a provisional target on the way to something lower. Keynesian aggregate demand management promised to deliver all this. All that was required of monetary policy was not to raise interest rates and nullify the expansionary effects of the Kennedy tax cut passed in spring 1964.

Under pressure from Congress and the administration, the FOMC had held off on raising the funds rate after passage of the Kennedy tax cut. To have raised rates would have "thwarted the will" of the political system to stimulate the economy and to lower the unemployment rate. As the unemployment rate fell steadily from 5% in early 1965 to 4% at the end of 1965, it became evident that the economy was growing at an unsustainable pace. Starting in late 1965, inflation began to rise. In a warning shot to the administration, the Board of Governors ratified an increase in the discount rate in 1965. Belatedly, in summer 1966, the FOMC began to tighten significantly. Then, in spring 1967, when the economy started to weaken, the Fed reversed course and lowered the funds rate despite the rise in inflation from 1% to 3%. It did so in return for a promise from Treasury Secretary Henry Fowler that, if the Fed would hold off on rate increases, the Johnson administration would submit legislation to Congress for a tax hike.

The Keynesian CEA believed that it was possible to lower the growth of aggregate demand and thus control inflation through restrictive fiscal policy rather than through restrictive monetary policy (the so-called optimal instrument policy). Doing so, it was believed, would avoid the "high" interest rates required by restrictive monetary policy and which hurt housing. Martin bet that if he could not control the fires of inflation with higher interest rates, he would turn down the fire by limiting credit creation. When the tax hike failed to materialize, the FOMC again began to raise the funds rate. However, when a tax surcharge finally materialized in

June 1968 after a run on gold that threatened the Bretton Woods system, the Board of Governors approved a reduction in the discount rate.

Despite continued high rates of money growth, Keynesian economists, including the staff of the Board of Governors, forecast a recession. Instead, the economy continued to grow strongly, and the unemployment rate fell to 3.4% in fall 1968. While interest rates treaded water, inflation rose and reached 6% in 1969. The stimulative effects of rapid money growth trumped the contractionary effect of the tax surcharge. Martin realized his mistake and implemented a highly restrictive monetary policy in 1969. Despite growing recession, he held off on any reduction in rates until inflationary expectations would subside. However, his term as FOMC chairman ran out in January 1970, and he left the Fed before he could return the country to price stability.

As a result of this experience, the Keynesian consensus in favor of deficit spending as the tool of aggregate demand management broke down. Given the evidence that expansionary monetary policy had trumped contractionary fiscal policy, the Keynesian consensus shifted opportunistically to monetary policy as the preferred tool of aggregate demand management. Although the consensus went from viewing monetary policy as impotent to being potent, it did so with the assumption that inflation was a nonmonetary phenomenon. That is, inflation had its origins in phenomena unrelated to monetary policy.

Martin's successor, Arthur Burns, who became FOMC chairman in February 1970, was not a Keynesian, but he shared the view that inflation was a nonmonetary phenomenon. Burns epitomized the "measurement without theory" tradition of the National Bureau of Economic Research (NBER). He understood monetary policy through the eyes of the businessman, whose psychology, he believed, drove the business cycle. Burns also believed that the country needed him to reconcile the low unemployment required for social stability with the low inflation required for business confidence and investment. For Burns, inflationary expectations were crucial. However, they were the expectations of the businessman, and the businessman was concerned about wage inflation. Burns believed that the country could lower inflation and stimulate economic activity by assuaging these expectations through the use of incomes policies and evidence of fiscal discipline in the form of a government budget surplus. In practice, Burns traded monetary policy for influence over incomes and fiscal policy. The exception was during the Gerald Ford administration, which, at the urging of CEA Chairman Alan Greenspan, rejected incomes policies (Hetzel 1998).

Burns believed that union power as well as a permissive welfare state had imparted an inherent inflationary bias to the economy. He also believed that the high unemployment rates required to suppress these cost-push pressures would be politically unacceptable. Because of these beliefs and his intellectual stature, Burns became the foremost proponent and intellectual champion of "incomes policies," which entail government intervention in the price setting of private firms. In order to limit the need for an increase in unemployment in order to restrain inflation, policymakers had to analyze the cause of inflation as cost-push or aggregate demand (demand-pull). They then needed to tailor the remedy to the specific cause of

inflation. Cost-push inflation needed intervention in the particular markets that created it. In a Faustian bargain with President Richard Nixon in August 1971, Burns accepted expansionary monetary policy in return for wage and price controls.

In the 1970s, the general belief that 4% unemployment represented full employment, when combined with an unemployment rate averaging 6%, created the impression that inflation had to be of the cost-push variety. The assumption that powerful monopolies (corporations and labor unions) powered inflation meant that a high rate of unemployment would be required in order to restrain it. In the jargon of the times, the "sacrifice ratio" was high. Incomes policies were the key to relaxing this trade-off. Paul Samuelson, who was the dean of Keynesian economists, expressed these general beliefs view when he opposed lowering inflation through restrictive monetary policy. Samuelson (1979 [1986], p. 972) argued:

> Today's inflation is chronic. Its roots are deep in the very nature of the welfare state. [Establishment of price stability through monetary policy would require] abolishing the humane society [and would] reimpose inequality and suffering not tolerated under democracy. A fascist political state would be required to impose such a regime and preserve it. Short of a military junta that imprisons trade union activists and terrorizes intellectuals, this solution to inflation is unrealistic–and, to most of us, undesirable.

The prevalent belief in the power of cost-push inflation led to the period-by-period discretionary setting of monetary policy in the 1970s. Although the Fed never spoke in terms of trade-offs, Phillips curve trade-offs between inflation and unemployment and the presumption that lowering inflation required large increases in unemployment (a high sacrifice ratio) were at the center of monetary policy. As Burns (1979) argued in his monograph, *The Anguish of Central Banking*, on an ongoing basis, the Fed had to make a judgment about the extent to which to restrain inflation based on a judgmental decision about much unemployment was politically acceptable.

G. William Miller, who succeeded Burns and was FOMC chairman from March 1978 to August 1979, by liberal temperament was unprepared to raise interest rates dramatically to deal with inflation for fear of recession. However, in spring 1979, a rise in inflation exacerbated expected inflation and lowered real interest rates. Monetary policy became expansionary in an environment of already high inflation. It was in that fraught environment that Paul Volcker succeeded Miller in August 1979.

The Great Moderation

As recounted above, at key periods when the character of the monetary regime has been in play, the regime that emerged came out of the way in which the intellectual and political environment of the time influenced the interpretation of economic instability. With the failure of the recession that started in August 1929 to end after a year or so with a strong recovery, monetary policymakers entered into a state of paralysis. Under the assumption that the economic collapse had originated in the

collapse of economic activity artificially inflated by credit forced on the banking system, policymakers found it impossible to adopt modern policies of money and credit expansion. Had not those very policies created the speculative bubble whose collapse led to the Depression?

When this real bills view of the world disappeared in the changed environment following World War II and the Korean War, the economics profession turned to Keynesian ideas. With real bills, there was no conception of the operation of the price system. As economists, Keynesians thought in terms of the operation of the price system but interpreted the prolonged, high unemployment of the Depression as a failure of the price system to provide for full employment. At the same time, Keynesianism sounded a clarion call to the economic profession with its rejection of the real bills belief that recessions represented the inevitable and required adjustment to imbalances created during periods of prior speculative excess.

American economists especially wanted to give content to Keynes' agenda exposited in *The General Theory*. John Maynard Keynes wanted to explain why the market economy failed to assure full employment and how deficit spending could provide the required lack of aggregate demand. Keynes' followers, who by numbers constituted basically all economists in academia until the late 1970s, wanted to fulfill Keynes' agenda through the development of models that would make the predictions required in order to use fiscal policy as an instrument of aggregate demand management.

In the Keynesian model as developed in its IS-LM formulation, equilibration of the economy to shocks to confidence occurred through quantity adjustments, that is, through output, not prices. The working assumption drawn from descriptive observation in the Great Depression was that the replacement of competitive markets with monopolies in the form of large corporations, labor unions, and cartels limited the ability of markets to establish market-clearing prices. The replacement of competitive markets with monopolies rendered a free market economy unstable and riddled it with nominal rigidities. On the one hand, those nominal rigidities meant that aggregate output was subject to the vagaries of an unstable private aggregate demand. On the other hand, they meant that an activist policy based on deficit spending could counter instability in private aggregate demand.

The issue in the 1960s had been how much aggregate demand inflation would emerge from an activist policy of maintaining full employment, universally taken to be at most 4%. Economists sought the answer in empirical estimates of a Phillips curve (the Samuelson-Solow Phillips curve) that offered a long-run inflation-unemployment trade-off. When the combination of 6% inflation and 6% unemployment emerged in 1970, the profession as well as Fed policymakers concluded that inflation could not be aggregate demand inflation but instead had to be cost-push inflation. Cost-push wage and price pressures had pushed up the Phillips curve so that inflation accompanied full employment. Motivated by the political expediency of Richard Nixon, the country turned to wage and price controls in August 1971 to deal with inflation.

In the 1960s and 1970s, the intellectual consensus in favor of Keynesian ideas interacted with a political environment that demanded "low" unemployment as a

salve for rendering tolerable social divisions. However, by the end of the 1970s, the combination of high inflation and high unemployment, termed stagflation, created a receptivity to new ideas. The Nixon wage and price controls and the Carter attempts at incomes policies had failed to restrain inflation.

In the 1960s and 1970s, a few economists, led mainly by Milton Friedman at the University of Chicago, but also including economists at the St. Louis Fed like Homer Jones and Leonall Anderson as well as Karl Brunner and Allan Meltzer mounted a monetarist attack on Keynesianism. Until the end of the 1970s, in the mind of the public, there was no association of inflation with monetary policy. Monetarists challenged the dominant view of inflation as a nonmonetary phenomenon by attributing it to high money growth. Friedman also challenged the Keynesian view by arguing that there was no long-term trade-off between inflation and unemployment. The Fed should return the country to price stability and allow the price system to determine the unemployment rate.

In the 1970s, a group of economists developed the logic of monetarist ideas in a school of thought known as rational expectations. They included Robert Lucas at the University of Chicago and Tom Sargent at the University of Minnesota. Keynesians assumed that inflation, whether cost-push or aggregate demand, propagated through a wage-price spiral that imparted to it a built-in inertia. Rational expectations proponents challenged that view. In doing so, they took aim at popular estimates of the sacrifice ratio, which purported to provide the man-years of unemployment required in order to lower inflation. Proponents of rational expectations argued that a credible monetary policy would attenuate the high unemployment assumed to be required in order to lower inflation (Humphrey 1986).

The rational expectations argument was that markets work because they use information efficiently. If the central bank provides a stable nominal anchor in terms of the domestic price level, firms will set prices in a way that separates the determination of relative prices from the price level. They will coordinate their adjustment of dollar prices on the inflation target of the central bank and then set the relative price of their product at the market-clearing value. Robert Lucas and Tom Sargent espoused the rational expectations view that the Fed could make the setting of individual dollar prices consistent with price stability through the way in which a credible rule would create an environment of nominal expectational stability (Nelson 1977). However, the massive Keynesian consensus, which still held at the end of the 1970s, considered the monetarist/rational expectations views as totally irrelevant to the "real world" in which unions and corporations exercised significant market power. Stated alternatively, the assumption that inflationary expectations tethered to the exercise of market power could instead be tethered to a monetary rule appeared flatly contradicted by observable reality.

In the 1960s, while at the Treasury, Paul Volcker had been the mandarin of the Bretton Woods system. He believed that a stable dollar was a bulwark of the free world against communism. For Volcker, a strong dollar was a moral imperative and a patriotic duty. As a product of the open market desk of the New York Fed, he understood the crucial role of expectations. In the political environment of 1979 with a president, Jimmy Carter, widely seen as weak, Volcker also understood that

the Fed was the only alternative for confronting inflation and inflationary expectations.

The rational-expectations challenge to Keynesianism interacted with Volcker's experiment in the monetary control of inflation to create a new monetary regime. The earlier 1970s experiment in monetary policy was based on the premise that monetary policy could control unemployment at an acceptable cost in terms of inflation guided by estimated Phillips curve trade-offs, especially when aided by incomes policies. Conversely, a policy of price stability, it was assumed, would come at a high cost in terms of unemployment given the way in which cost-push pressures powered inflation. Just as Arthur Burns and his successor G. William Miller delivered the Keynesian experiment in aggregate demand management, Volcker delivered the monetarist/rational expectations experiment of basing monetary policy on the creation of a stable nominal anchor.

In the 1950s, Volcker had worked at the Open Market Desk of the New York Fed. The exposure to money markets inculcated the importance of expectations. At the same time, until the collapse of the Bretton Woods system in 1973, New York Fed economists as epitomized by Charlie Coombs, head of the foreign exchange desk, viewed market speculation as destabilizing. Speculators attacked the exchange rates set by Bretton Woods and had to be fought.

By the time Volcker had moved from president of the New York Fed to FOMC chairman in August 1979, he understood that expectations were indeed destabilizing. However, by the end of the 1970s, they were destabilizing because financial markets had come to expect that in recession an expansionary monetary policy would raise inflation. That expectation pushed up bond rates in recession and limited the downward adjustment of inflationary expectations. Countercyclical monetary policy based on Phillips curve trade-offs had become destabilizing.

Volcker knew that reducing double-digit inflation would require a recession. In the past, the economy had suffered through recessions undertaken in order to lower inflation. The character of monetary policy changed because of Volcker's concern that inflationary expectations should remain subdued during the economic recovery that would follow a disinflationary recession. That is, in order to restore a stable nominal anchor, the Fed not only had to convince markets that it would restore low inflation but also that trend inflation would remain stable rather than increase over the business cycle. In accepting Fed responsibility for inflationary expectations and for making them invariant to the cyclical fluctuations of the business cycle, Volcker had to reject the Keynesian consensus that favored aggregate demand management.

Volcker's acceptance of responsibility for inflationary expectations also rendered irrelevant the endless discussions about the causes of inflation as cost-push or demand-pull and about tailoring the remedy accordingly. Whatever the cause of those inflationary expectations, the Fed had to get them under control, or they would pass into negotiated wage settlements and lock in high inflation. At the beginning, Volcker turned to money targets in order to establish credibility. The commitment to maintain low money growth seemed to provide that credibility.

However, with the *Monetary Control Act of 1980*, money demand became procyclical. The act reduced the cost of transferring funds between money market

instruments used as savings vehicles and bank deposits used for transactions. Given that banks change the rates they pay on their deposits sluggishly as money market rates of interest change, money growth would decline in booms when market rates rose relative to bank deposit rates and conversely. Money growth would then send misleading signals for the desirable behavior of interest rates. As a result, a low, stable target for money growth would not offer a useful guide to monetary policy.

Proponents of monetarism and rational expectations wanted to restore the stability of the nominal anchor lost in the prior stop-go era of monetary policy. Monetarists had advocated either steady money growth or multiyear targets for money. However, after fall 1983, it became evident that the FOMC could not adopt a reserve-aggregate or monetary target. Instead, policy focused on behaving in a consistent way in order to shape the inflationary expectations of financial markets.

The Volcker-Greenspan era became "lean-against-the-wind with credibility" (Hetzel 2008). The Fed followed its long-standing lean-against-the-wind procedures but implemented them in a way disciplined by the bond market vigilantes, who associated the former policy of aggregate demand management with inflation that rose across the business cycle. That is, during the economic recovery from the 1982 to 1983 recession, the FOMC augmented the regular progression of funds rate increases with additional increases when bond rates jumped. The FOMC interpreted these jumps in bond rates as "inflation scares" that challenged its credibility to maintain low, stable inflation (Goodfriend 1993).

One can date the start of the Volcker monetary regime to the May 1983 inflation scare. By raising the funds rate with the unemployment rate at 11% and with sharply falling inflation, it began the work of creating a new nominal anchor in the form of a widespread expectation of low, stable trend inflation. The bond market vigilantes forced the FOMC to raise the funds rate early during the cyclical recovery. The disappearance of the cyclical inertia that had characterized funds rate movements in the stop-go period required an end to the exploitation of Phillips curves inherent in the policy of aggregate demand management. Inflation targeting emerged in the sense that the Fed committed to making expected trend inflation invariant to shocks whether inflation shocks or recession shocks.

The United States moved to a new monetary standard in which the nominal anchor was a rule that stabilized the expectation of the future price level. In doing so, policy had to avoid the attempts to exploit the inflation-unemployment trade-offs that had defined the preceding stop-go monetary policy. The juxtaposition of the Volcker-Greenspan standard with the prior standard in which the Fed had allowed high and rising inflation in an attempt to maintain "low" unemployment could not have been more dramatic. The stop-go monetary policy of the 1970s reflected the Keynesian view of untethered inflationary expectations arising out of a self-reinforcing cycle of market power exercised by large corporations and labor unions. The policy of restoring nominal expectational stability arose out of a belief that inflationary expectations would conform to rule-like monetary policy.

At the time, no one knew whether the Volcker experiment would succeed. There was a foreboding that, if the experiment did not succeed, the United States would live with inflation accompanied by on-and-off price controls and that a free market

economy would survive only in a wounded form. No one knew whether Congress and President Ronald Reagan would support Fed independence through disinflation and recession. And, no one knew with any certainty whether disinflation would be followed by sustained low and stable inflation. The Keynesian consensus held that maintaining low, stable inflation would require periodic bouts of politically unacceptable and socially undesirable high unemployment in order to contain cost-push inflation. That fear failed to materialize.

In the event, the change in the monetary regime replaced the Great Inflation with the Great Moderation. Friedman had offered an explanation for the inverse correlations between inflation and unemployment summarized in the Phillips curve that predicted the disappearance of those correlations in response to sustained inflation. The stagflation of the United States in the 1970s supported that prediction. In reference to the breakdown of this version of the Phillips curve, popularized by Samuelson and Solow (1960 [1966]), Lucas and Sargent (1978 [1981], p. 303) talked about "econometric failure on a grand scale." Similarly, the correlations disappeared with sustained low inflation.

LAW with credibility also shaped the broad contours of the Greenspan tenure as chairman. Like Volcker, Greenspan believed in the fundamental responsibility of monetary policy to shape inflationary expectations. Coming out of the 1990 recession, he lowered the funds rate only cautiously because the inflation premia built into bond rates were inconsistent with his goal of returning to price stability. The strong funds rate increases in 1994 were his decisive response to an inflation scare. The FOMC raised the funds rate from 3% going into the February 1994 meeting to 6% at the February 1995 meeting. Credibility allowed Greenspan to depart from these procedures at the time of the 1997–1998 Asia crisis, but in the context of his entire tenure, the departure was brief.

Although money targets would not be a purposeful instrument for achieving control of prices, money would remain central to the power of central banks as the "stick in the closet." The ability to move nominal money relative to real money demand by moving the real rate of interest relative to the natural rate of interest endows central banks with their ability to enforce a market expectation of inflation consistent with its inflation target. That inflation target then coordinates the nominal price setting of firms that set prices for multiple periods. Moreover, for price stability, nominal money creation must follow the evolution of real money demand. For that to happen, the central bank must avoid the consequences of price fixing by following a rule that causes the real interest rate to track the natural interest rate, that is, the interest rate that would prevail in a world of completely competitive markets. Otherwise, one is back in the world of Milton Friedman (1960) in which erratic changes in money force changes in prices, while "long and variable lags" destabilize expenditure.

The juxtaposition of sustained high unemployment in the Depression with low unemployment in World War II created the impetus for the Keynesian Revolution with its focus on using deficit spending as the tool of aggregate demand management. The juxtaposition of stagflation in the 1970s with disinflation and moderate unemployment in the Volcker-Greenspan era created the a new professional

consensus that economic instability in the past was due to "poor play" by the Fed rather than "poor hand" due to external shocks in the terminology of Francois Velde (2004). Of course, there are "poor hand" interpretations of history also. Blinder (1982), for example, sided with Arthur Burns (1979) in attributing the instability of the 1970s to large supply shocks. Nevertheless, the new professional consensus provided the impetus for the "New Keynesian" models with their focus on the compatibility of price stability and full employment.

In its spare formulation, the New Keynesian (NK) model consists of three sectors of the economy (Woodford 2003). One sector describes how households borrow and lend in order to distribute their consumption evenly over time. In a world in which households are optimistic about their future income prospects, they desire to smooth consumption by bringing it into the present through borrowing. Because of the inability to move production from the future to the present, the real interest rate needs to be relatively high enough in order to counter the desire to borrow and instead to provide an incentive to lend.

The other sector describes how firms price their products. Firms in the "sticky-price sector" can only set their dollar prices periodically. As a consequence, they must set them based on forecasts of both their marginal costs and inflation. The fact that many prices are preset in this way allows the central bank to influence aggregate real demand through its influence on aggregate nominal demand. At the same time, as pointed out by Marvin Goodfriend and Bob King (1997), through following a policy of price stability, the central bank can neutralize this price-stickiness friction. By rendering inoperative that friction, a policy of price stability allows the price system to work to maintain output at its potential (natural) level.

The implication of a policy concentrated on price stability is that the central bank allows the price system to determine real variables like the unemployment rate. Applied to the monetary policy of the Volcker-Greenspan era, the policy of restoring nominal expectational stability, which is the real-world analogue of price stability, prevented the FOMC from attempting to exploit Phillips curve trade-offs. As a consequence, it had to follow a rule for setting the funds rate in a way that tracked the natural rate of interest determined in the competitive core of the economy. In doing so, it allowed the price system free rein to determine real variables.

In an attempt to revive the Keynesian spirit of the 1970s, Blanchard and Gali (2007) pointed out that in a world in which powerful monopolies generate cost-push inflation (represented in the NK model by the addition of markup shocks – shocks to price over marginal cost), the central bank can reduce output variability by increasing inflation variability. In this Phillips curve world, price stability does not guarantee a zero output gap (full employment). In the terminology of Blanchard and Gali, the "divine coincidence" property of the NK model highlighted by Goodfriend and King no longer holds. However, cost-push inflation has failed to materialize. Central banks have been more concerned about low inflation than high inflation.

As a condition for price stability, the central bank must cause nominal money to grow in line with real money demand. As detailed above, the New Keynesian model provides a framework for understanding how the central bank provides for this monetary control. The central bank must follow a rule that provides for a stable

nominal anchor in the form of nominal expectational stability (the expectation of price stability). It must also implement a rule that causes the real interest rate to track the natural rate of interest. Real output follows potential output. As a consequence of its interest rate target, nominal money then grows in line with the real money demand consistent with potential output and random variation in real money demand.

The following section translates this discipline on money creation in terms of the Fed's procedures developed during the Volcker-Greenspan era. The monetary regime created by the Volcker and Greenspan FOMCs led to a long period of economic stability. What was the baseline monetary policy in this period?

What Monetary Regime Did Volcker and Greenspan Create?

The FOMC sets a target for the funds rate, which is an overnight interest rate paid on bank reserves. An overnight rate, however, is meaningless. What is relevant for economic activity is the entire term structure of interest rates – the yield curve. The primary determinant of the shape of the yield curve is the geometric average of the forward (one-period) rates expected by market participants in the future. (The actual yield curve includes term premia that compensate investors for holding a long-term security as opposed to a succession of one-period securities.) It follows that monetary policy consists of how the FOMC communicates to markets its strategy (reaction function) for responding to incoming "news" about the economy, where "news" is new information. In order for monetary policy to be stabilizing, that strategy must cause the yield curve to respond in a stabilizing way to incoming news.

Consider, for example, news that the economy is growing faster than previously anticipated. A stable nominal anchor requires that the yield curve rises with the entire rise occurring in real forward rates and none in inflation term premia. Just as the nominal anchor in the gold standard was the expectation that the central bank would maintain convertibility in the future, the nominal anchor is now the expectation that the Fed will maintain price stability in the future. For that reason, the FOMC continually watches all available measures of expected inflation, for example, the breakevens on TIPS (the difference for a given maturity between the yield on a nominal Treasury security and Treasury securities adjusted for inflation). The difference must compensate holders of nominal securities for inflation.

The FOMC does not articulate an explicit strategy (a rule) to markets but instead communicates in terms of a forecast of the evolution of the economy in a way that promises both price stability and full employment. Underlying the forecast of the economy is a forecast of the path of the funds rate assumed to be consistent with these objectives. In response to news, the FOMC changes the assumed path of the funds rate, but the objectives (price stability and full employment) do not change.

FOMC communication takes several forms. The first two paragraphs of the statement issued following an FOMC meeting contain, respectively, an assessment of the change in the behavior of the economy since the last FOMC meeting and a forecast of its near-term behavior. The Summary of Economic Projections (SEP) included four times a year contains forecasts for the individual FOMC participants of

output, unemployment, and inflation as well as a funds rate forecast contained in the so-called dot plot. Although the SEP does not directly reveal what markets care about, namely, the FOMC consensus, market participants look at median values for such a consensus. The FOMC chair's press conferences, congressional testimony, and speeches also offer information.

There is a two-way dialogue between the FOMC and the financial markets. The yield curve comprises the geometric average of the market's expectation of future forward rates. As time passes and forward rates become today's spot rate, the FOMC's funds rate target dominates. Markets then have to respect the funds rate path the FOMC believes appropriate. At the same time, the yield curve comprises a collective forecast of the evolution of the economy made by participants in financial markets. The FOMC has to be concerned if the market's forecast of the future funds rate implicit in the yield curve differs from its own. Perhaps the market sees incipient strength or weakness that the FOMC is missing.

Because the FOMC does not articulate an explicit strategy, it does not explain how it deals with the fact that there are two targets – inflation and unemployment. Markets learn the strategy through consistency over time in FOMC behavior and from sources such as speeches by the FOMC chair. As noted above, possible strategies fall into two broad categories – nonactivist and activist. Consider first the nonactivist strategy under the assumption that the FOMC has achieved credibility for price stability.

In the period after the 1951 Treasury-Fed Accord, the FOMC developed procedures termed "lean-against-the-wind" by its chairman William McChesney Martin. In the event of unsustainable strength in the economy (sustained increases in the rate of resource utilization or equivalently sustained declines in the unemployment rate), the FOMC raised the funds rate above its prevailing value in a measured way. Credibility required that financial markets make forecasts of the future path of the funds rate based on the assumption that the FOMC would raise the funds rate to whatever extent required in order to maintain price stability.

The situation is complicated when the economy is recovering from recession and the economy is necessarily growing at a rate in excess of potential. The FOMC has to continually assess whether the yield curve possesses a sufficiently steep upward slope so that output will follow a glide path back to and then along the path of potential output. With an activist policy, in contrast, the FOMC limits increases in the funds rate until it starts to observe inflation. If inflation continues at an unacceptably high level, the FOMC raises the funds rate persistently until the economy weakens and in time inflation weakens.

There are then basically two variations in the post-Accord monetary regime. In the nonactivist regime in which the FOMC eschews inflation/output trade-offs, the FOMC provides a stable nominal anchor and then follows a rule that allows market forces (the price system) to determine the real interest rate and by extension real variables like the unemployment rate. In the activist regime, the FOMC attempts inflation/output trade-offs. Those Phillips curve trade-offs at least in principle allow it to mitigate fluctuations in output by offsetting fluctuations in inflation and conversely. After the disinflation of the early 1980s, the Volcker-Greenspan era was nonactivist.

The Great Recession

The real challenge to the Volcker-Greenspan nonactivist policy and its interpretation by Goodfriend/King (1997) in terms of the NK model has come with the Great Recession. It created the widespread perception that the recession arose from a disruption to financial intermediation. Critics of the Fed then argued that "inflation targeting" contributed to the Great Recession through precluding concern for financial stability (Curdia and Woodford 2009). Such explanations often assumed that the low interest rates in the early 2000s initiated a boom-bust cycle in housing whose fallout led to the Great Recession (Taylor 2009).

The intellectual tenor of the times became reminiscent of real bills. In a real bills world, market power appears in the form of the herd behavior of investors. That is, market psychology fluctuates between intervals of excessive euphoria and excessive pessimism. In the periods of optimism about the future, households and firms assume an excessive level of debt. Inevitably, a correction occurs as a consequence of the need to purge these excesses and to eliminate "imbalances." A period of painful deleveraging and recession ensues. According to this view, the prevention of speculative excess (maintenance of financial stability) should be an additional objective of the central bank beyond output and inflation stability.

In any single recession, many factors are at play. In the Great Recession, the financial disruption that followed the bankruptcy filing of Lehman Brothers on September 15, 2008, undoubtedly exacerbated the decline in economic activity. The failure of Lehman shook the belief of cash investors (holders of short-term, liquid debt like repurchase agreements, RPs) that regulators would never allow the failure of a major financial enterprise. Cash investors ceased buying the short-term debt (RPs and commercial paper) of financial institutions holding significant amounts of illiquid assets like mortgage-backed securities (MBS) and placed their funds in the too-big-to-fail banks like J. P. Morgan Chase.

At the same time, in the Great Recession, deflation combined with the decline in economic activity (Fig. 4). In no model can that combination be consistent with an optimal monetary policy. Similarly, the persistent decline in inflation after 2008 must reflect contractionary monetary policy. The severity of the Great Recession is consistent with a monetary policy that kept real interest rates above their natural level. Two major shocks made households pessimistic about the future and would have required a sharp decline in the real interest rate in order to sustain their consumption.

First, in mid-2007, house prices stopped rising and started falling. Second, an enormous commodity-price inflation shock that lasted from mid-2004 until mid-2008 lowered household real disposable income. The price of a barrel of oil (WTI) went from $34 in early 2004 to $134 in June 2008. As shown in Fig. 4, in July 2008, 12-month personal consumption expenditures (PCE) inflation reached 4.2% and core PCE inflation reached 2.3%. Although the FOMC did not set an explicit inflation target until January 2012, implicitly it was aiming for 2% inflation. The inflation overshoot created pressure on the FOMC to impart downward inertia to declines in the funds rate relative to declines in the natural rate of interest.

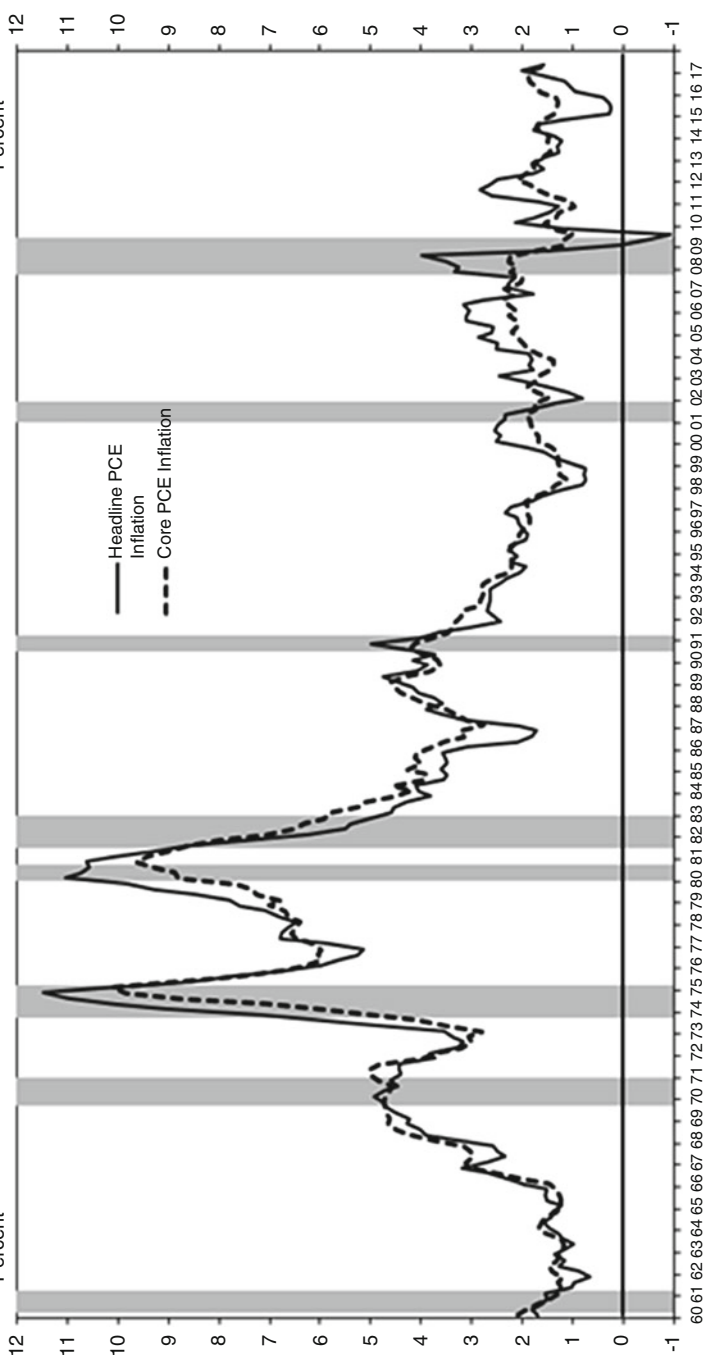


Fig. 4 Headline and core PCE inflation. Notes: Quarterly observations of four-quarter percentage changes in the headline and core PCE deflator. Shaded areas indicate NBER recessions. (Source: Haver Analytics)

One feature of the NK model is particularly relevant for the Great Recession. The counterpart to firms in the sticky-price sector that are constrained to change prices only infrequently is firms in the flexible-price sector. The latter firms operate in auction markets in which prices are determined continuously. In order to allow the price system to determine relative prices, the central bank should target inflation in the sticky-price sector and allow inflation originating in the flexible-price sector to pass through to headline inflation (Aoki 2001). In terms of Fig. 4, assuming an implicit inflation target of 2%, in 2008, the FOMC should have been concerned with the slight overshoot of 2.3% in core inflation not the 4.2% overshoot in headline inflation. However, the Fed was concerned that high headline inflation would push inflationary expectations above its implicit inflation target (Bernanke 2008; Hetzel 2012).

Prior to the cyclical peak in December 2007, the FOMC followed its standard procedures. In cyclical recoveries, it raises the funds rate in steps until the economy weakens. The FOMC raised the funds rate from the 1% prevailing at its June 2003 meeting to 5.25% at its September 2007 meeting. As the economy began to weaken, the FOMC began to lower the funds rate. The economy deteriorated steadily after the cycle peak in December 2007 and entered into a serious recession in late summer. Annualized monthly changes in nonfarm payrolls grew weakly from June 2007 to December 2007 at 0.5%. They basically stopped growing in January 2008 and then deteriorated steadily. By September 2008, the decline was −3.9%. (The September survey date was prior to the Lehman bankruptcy.) The economy declined despite a massive boost to disposable income from the Bush tax cut. Tax rebates produced a jump in real disposable personal income of $562.1 billion in May 2008.

At the same time, the inflation shock intensified headline inflation. Although the FOMC lowered the funds rate from 2.25% to 2% at its April 2008 meeting, it conveyed the message in its post-meeting statement that declines would cease and that the next move was likely upward. As a consequence, the yield curve rose. In retrospect, the natural interest rate continued to fall. At its December 2008 meeting, the FOMC lowered the funds rate to near zero, where it remained until December 2015. With PCE inflation averaging just below 2% over the 5-year period 2009Q2 through 2014Q2, after the fact, it became evident that the economy required a short-term real interest rate of about −2%. Although the 2% funds rate target set at the April 2008 meeting produced a real interest rate near zero using core measures of inflation, it was still too high given the magnitude of the housing and inflation shocks hitting the economy – the shock to wealth from the decline in housing prices and the shock to real disposable income from the inflation shock.

In this respect, the Great Recession evolved similarly to other recessions. Figure 5 shows a measure of the output gap and PCE inflation. A pattern common across recessions is that prior to cycle peaks, the economy begins to weaken. That is, the output gap begins to decline. However, going into recession, inflation is at a cyclical high – a level unacceptable to the FOMC. Figure 6 provides a rough measure of the real rate of interest as the difference between the funds rate and inflation. As shown in Fig. 6, the FOMC puts downward inertia into reductions in the funds rate as the economy weakens and the real rate of interest remains at cyclical highs (Hetzel 2012, Figs. 4 and 5).

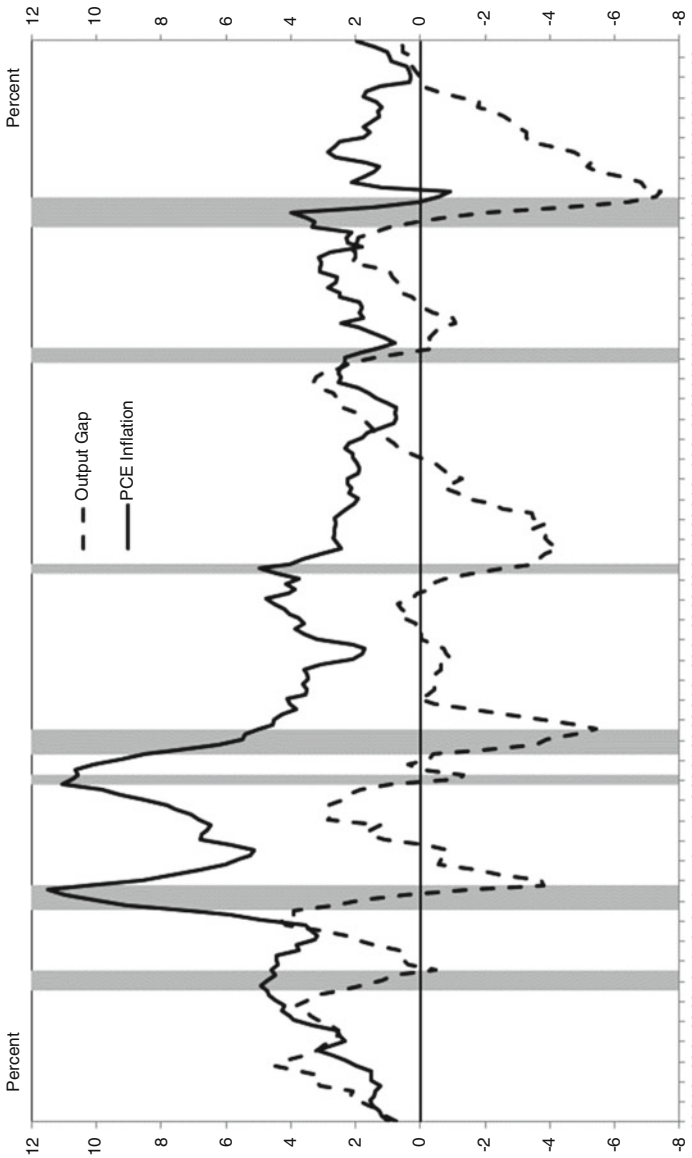


Fig. 5 The output gap and inflation. Notes: PCE inflation is four-quarter percentage changes in the PCE deflator. Shaded areas indicate NBER recessions. The output gap is taken from the econometric model FRB/US. See Board of Governors: Economics Research: FRB/US: model: FRB/US model packages: frbus_package. PCE inflation is from Haver Analytics

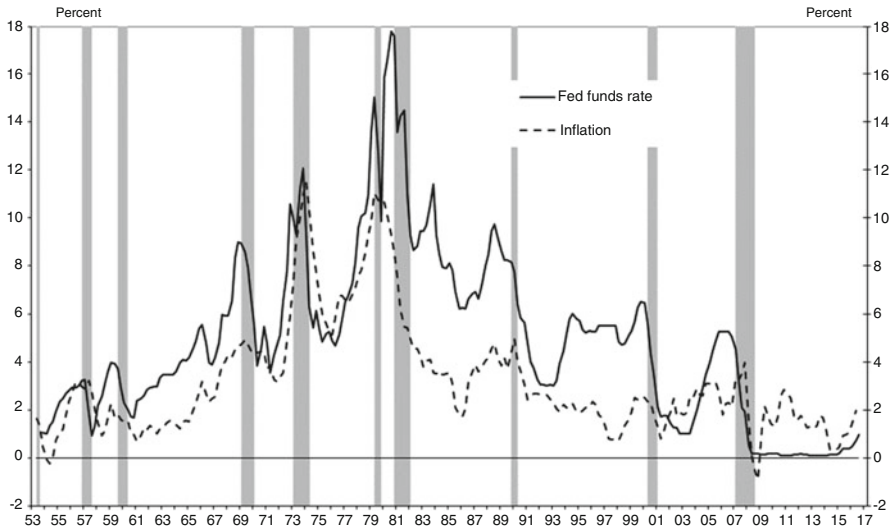


Fig. 6 Fed funds rate and inflation. Notes: Quarterly observations of the effective federal funds rate. Inflation is four-quarter percentage changes in the PCE deflator. Shaded areas indicate NBER recessions. Data from St. Louis FRED data base

This pattern is consistent with contractionary monetary policy at least exacerbating recessions through inertia in the funds rate relative to reductions in the natural rate of interest. The resulting behavior of inflation and the unemployment rate is captured in empirical Phillips curves. In the jargon of economics, however, the resulting Phillips curves are not "structural," that is, invariant to the behavior of monetary policy. The Fed itself creates the correlations between inflation and unemployment. Phillips curves are then not reliable guides to the formulation of monetary policy.

In fall 2008, it became obvious that the economy of the developed world was in a serious recession. When Lehman Brothers declared bankruptcy on September 15, 2008, cash investors stopped funding the investment banks that held large amounts of subprime mortgages. The Fed responded with programs that made it into a conduit for the financial intermediation not undertaken by the market. That is, the Fed undid the flight of cash investors from the institutions with illiquid, hard-to-value asset portfolios. The Fed's term auction facility (TAF), for example, purchased commercial paper. Also, the Fed engaged in swap arrangements with foreign central banks and especially the ECB in order that they could lend dollars to their banks to replace the dollar funding that had dried up (Hetzel 2008, Chaps. 13 and 14). These were credit interventions.

At its December 2008 meeting, the FOMC lowered the funds rate to near zero (a range from 0% to 1/4%), the zero lower bound. Earlier, in October 2008, it had begun to pay interest on the excess reserves of banks (IOER). In that way, the FOMC could create reserves in a way that allowed it to set a funds rate target independently. Starting in fall 2008 and extending until 2014, it engaged in three series of sustained

open market purchases of mortgage-backed securities (MBS) and Treasury securities (the LSAPs or large-scale asset purchase programs). The former had aspects of credit market intervention, while both create bank reserves. The Fed's asset portfolio went from about $900 billion to $4.5 trillion or from 6% of nominal GDP to 23% (Hetzel 2008, Chaps. 13 and 14; Potter 2017).

In terms of monetary policy, when the FOMC could no longer lower the funds rate, it communicated to financial markets its intention to extend the period of a zero funds rate. It used this forward guidance to bend down the yield curve. The difficult part was to avoid communicating to financial markets pessimism about the economy that could exacerbate fears of deflation, raise the real interest rate, and create a downward spiral in the economy. The LSAP program was important in that it communicated a "whatever it takes" policy to markets. That is, the FOMC communicated to markets that it would sustain an aggressively expansionary policy through maintaining the funds rate at zero for however long was needed in order to assure economic recovery. The FOMC was more aggressive than the ECB, and the ECB suffered through back-to-back recessions (Hetzel forthcoming-b).

Why Was Raising Inflation so Hard After the Great Recession?

As of fall 2017, the primary issue confronting the Fed was how to eliminate a persistent shortfall in inflation from its 2% target. Since January 2012, when the Fed established an inflation target, it routinely missed the target on the downside. Over the longer period from August 2008 through June 2017, headline PCE inflation averaged 1.2%, while core PCE inflation averaged 1.5%. The persistent shortfall in inflation from target raised the question, "How do central banks control inflation?"

One often hears the rhetorical question, "Given that the Fed controls money creation, why can't it control inflation?" However, with an interest rate target, nominal money creation follows money demand. Assuming that the Fed is not going to change the monetary standard to one of direct fiat money creation (helicopter money), what does the NK model say about the control of inflation? As noted, the model distinguishes between inflation in the sticky-price sector and in the flexible-price sector.

Figure 7 shows the Atlanta Fed measure of sticky-price and flexible-price CPI inflation (see Bryan and Meyer 2010). In periods in which inflationary expectations are reasonably well anchored, there is considerable stability of sticky-price inflation relative to flexible-price inflation. As a result, there is a significant amount of noise in headline inflation coming from the flexible-price sector. Also, with the exception of the period following the 2001 recession, from 1999 to 2011, flexible-price inflation significantly exceeded sticky-price inflation. Similarly, over this period, as shown in Fig. 4, headline inflation generally exceeded core inflation. The reason is the demand placed upon natural resources by the integration of the BRICS (Brazil, Russia, India, China, and South Africa) into the world economy. Because central banks followed policies intended to restrain headline inflation, monetary policy lowered core inflation.

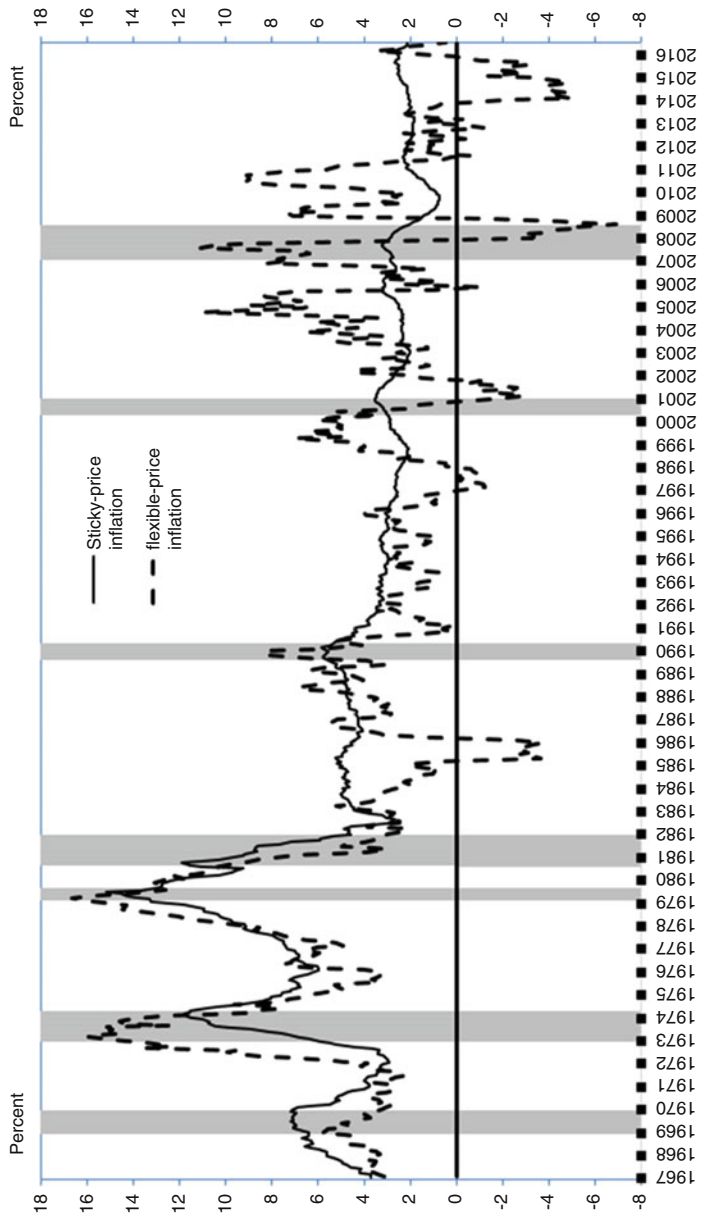


Fig. 7 Sticky-price and flexible-price CPI inflation. Notes: Observations are 12-month percentage changes in sticky-price and flexible-price inflation. Heavy tick marks indicate December. The years are centered below the January number for the year. For construction of the series, see Bryan and Meyer (2010). (Source: Federal Reserve Bank of Atlanta)

Moreover, as shown in Figs. 4 and 7, with an interruption in the last years of the 1980s, in the two decades preceding 1999, the Volcker and Greenspan FOMCs steadily reduced underlying inflation. As noted above, the Fed gained credibility in bond markets for near price stability after raising rates sharply starting in February 1994. Finally, sticky-price inflation declined about a percentage point between the immediate pre- and post-Great Recession periods. In updating the statistical decomposition of inflation into a time-varying trend and transitory changes in Stock and Watson (2007), Cecchetti et al. (2017) found that trend inflation for core PCE inflation declined after the Great Recession to about 1.5%. A weak world economy since 2012 kept inflation in the flexible-price sector low through weak commodity prices, especially oil. This combination of stable but low inflation in the sticky-price sector and low inflation and even deflation in the flexible-price sector kept headline inflation below 2%.

Figure 8 shows PCE inflation in the services and goods sectors. It offers similar insights under the assumption that services inflation is more representative of sticky-price markets and goods inflation is more representative of flexible-price markets. The figure shows the decline in services inflation that occurred after mid-2008 and the high degree of volatility in goods inflation relative to services inflation. Also, the terms of trade move secularly in favor of services. Hopefully, that relationship will remain constant. If it varies, achievement of the FOMC's inflation target will be rendered difficult by the need to alter expectations of firms in the sticky-price sector.

The Fed achieved price stability in the sense defined by Alan Greenspan. When asked by then Gov. Yellen to define price stability, FOMC chairman Greenspan replied that it "is that state in which expected changes in the general price level do not effectively alter. . .decisions" (Board of Governors *Transcripts* July 2, 1996, 50).

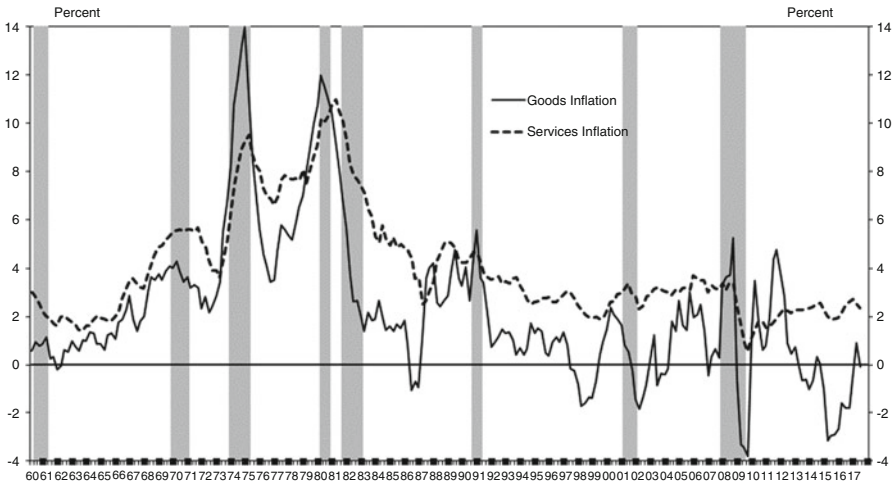


Fig. 8 PCE goods and services inflation. Notes: Quarterly observations of four-quarter percentage changes in the goods and services PCE deflator. Shaded areas indicate NBER recessions. Heavy tick marks indicate the fourth quarter. (Source: Haver Analytics)

Earlier, Greenspan had endorsed this expectational definition of price stability during the hearings in 1989 on Rep. Neal's (D. NC) resolution (H.J. Res. 409) requiring the Fed to achieve zero inflation within 5 years of the resolution's enactment. Greenspan supported the resolution conditional on the language that "inflation be deemed to be eliminated when the expected rate of change of prices ceases to be a factor in individual and business decision making" (citations from Hetzel 2008, p. 197).

In the context of fall 2017, it followed that raising inflation required making firms again taking into account inflation in setting prices. However, as shown in Fig. 4, only once following mid-2008 had the FOMC achieved 2% core PCE inflation. In the first 3 months of 2012, 12-month core PCE inflation reached 2.1%, perhaps because a second inflation shock passed through into core prices. How can the Fed change the inflationary expectations on which firms in the sticky-price sector coordinate? Repetition in speeches of the 2% inflation goal sounds like cheap talk. It will probably take an extended period of inflation in the flexible-price sector from an economy run "hot" to have that effect. What are the dangers of such a strategy?

In a similar earlier debate in 1997, Joseph Stiglitz, chairman of the CEA, answered that the dangers were manageable. The quandary was that the unemployment rate had fallen below its presumed NAIRU (non-accelerating inflation rate of unemployment) value of 6% and was continuing to decline. At the same time, inflation was falling not rising (1997 U.S. *Economic Report of the President*, 48). According to Stiglitz, the Fed could discover the NAIRU by allowing the unemployment rate to decline until inflation rose. In the event of an overshoot of inflation, the cost of lowering inflation would be manageable. The CEA report contained a graph of the Phillips curve (Chart 2.2). As noted in the heading, the Phillips curve implied a "sacrifice ratio" for the control of inflation. Namely, an excess of the unemployment rate over the NAIRU of one percentage point lowers inflation in the subsequent year by 0.6% points. In the event of an overshoot, a moderate but sustained unemployment rate in excess of its NAIRU value would restore inflation to its desired target.

The 1998 CEA *Report* (1998 U.S. *Economic Report of the President*, 60) stated, "The unemployment rate is a useful predictor of future inflation in that it can directly indicate the potential for rising inflationary pressure on the cost side, as excess demand in the labor market tends to raise nominal wages and thus nominal labor costs." The *Report* made reference to Chart 2.3. It showed the unemployment rate falling below the contemporaneously assumed NAIRU value of 6% in 1994Q3 with wage growth starting an upward ascent four quarters later in 1995Q4. However, as shown in Fig. 1, headline PCE inflation did not begin to rise until September 1998, with core PCE inflation beginning to rise only in May 1999. That is, if the NAIRU really was 6%, the lag between pushing the unemployment rate below NAIRU and the increase in core PCE inflation was almost 5 years. Figure 5 shows the output gap turning positive in 1997Q1 and inflation starting to rise in 1999Q2 – a lag of just more than 2 years.

In the late 1990s, the FOMC followed the Stiglitz advice to allow the unemployment rate to decline although motivated by a desire not to allow the dollar to appreciate during the emerging market crisis in 1997, 1998, and early 1999. The

FOMC did not raise the funds rate until June 1999 when the available May unemployment rate was 4.2% (Hetzel 2008, Chap. 17). The unemployment rate fell to 3.8% in April 2000. A recession did follow with a cyclical peak in March 2001. That fact suggests that while there may be an upside to allowing a positive output gap to emerge in order to raise inflation, there is a downside to eliminating it. Stated alternatively, a lowering of the unemployment rate below its natural value requires subsequently raising the unemployment rate by slowing the growth of the economy below potential.

Milton Friedman criticized this sort of aggregate demand management, which at its heart entails making Phillips curve trade-offs. He argued that because of the lags in implementing such policies, they result in destabilizing go-stop monetary policy (Friedman 1960, pp. 86–87). Friedman (1968 [1969], p. 109) wrote:

> The reason for the propensity to overreact seems clear: the failure of monetary authorities to allow for the delay between their actions and the subsequent effects on the economy. They tend to determine their actions by today's conditions–but their actions will affect the economy only six or nine months later. Hence they feel impelled to step on the brake, or the accelerator, as the case may be, too hard.

Figure 9 plots the unemployment rate along with the Congressional Budget Office (CBO) estimate of the natural rate of unemployment. The black squares demarcate increases in the unemployment rate that cumulated to 40 basis points or more and that did *not* immediately precede a business cycle peak or follow a business cycle trough. What is striking is the paucity of episodes of such increases in the unemployment rate, that is, increases not associated with recession.

The four squares shown in Fig. 9 may indicate incipient recessions forestalled by other events. The November 1951 increase demarcated the date when the Korean War intensified due to the entry of the Chinese crossing the Yalu River. It set off fears of World War III with the return of price controls and created an intense demand for any goods that could be stockpiled. The August 1952 increase occurred during the Korean War. The February 1963 increase marked a stalling of the economy. That stalling convinced President Kennedy and his advisors of the need to ask for a tax cut. January 1967 does not meet the criterion for a nontrivial increase in the unemployment rate although the downward movement in the unemployment rate stalled. However, in summer 1966, the Fed tightened in order to snuff out a rise in inflation. When President Johnson agreed to a tax hike, the Martin FOMC backed off tightening. The July 1976 increase remains a puzzle. In that summer, the economy seemed to stall but then picked up again in the fall.

Figure 9 is just a graph, but at a minimum, a nontrivial increase in the unemployment rate is a leading indicator of recession. If monetary policy operated through Phillips curve trade-offs of balancing off inflation and unemployment, especially in an environment in which cost-push inflation was regularly pushing up inflation, one would expect routine moderate increases in the unemployment rate. The CBO

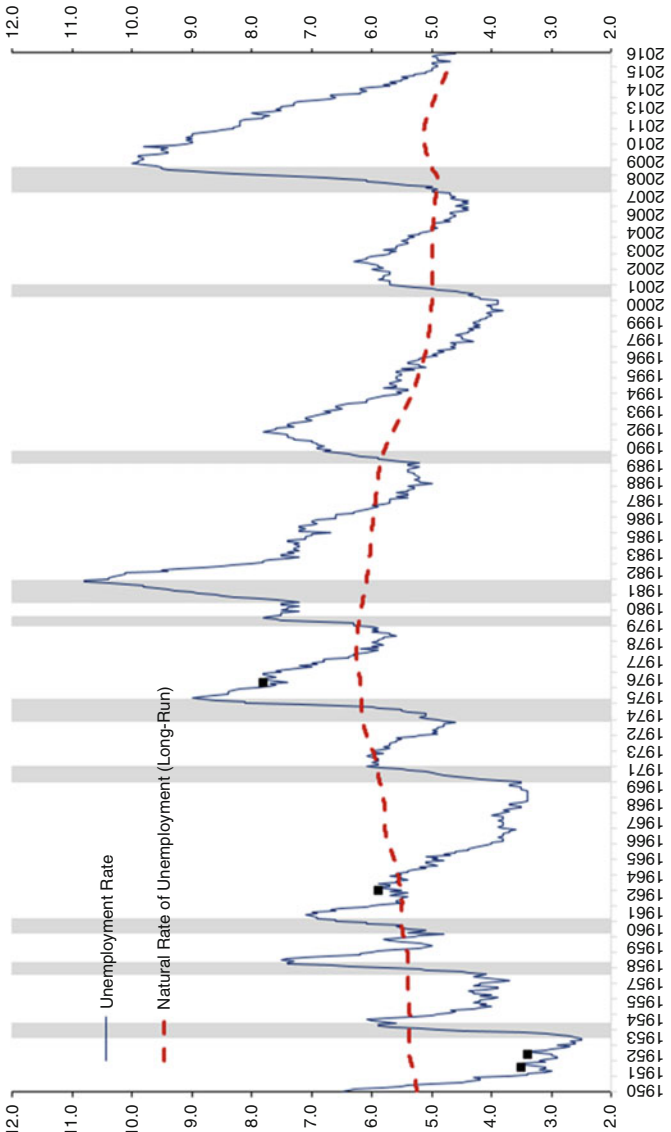


Fig. 9 Unemployment rate and CBO natural rate of unemployment. Notes: Monthly observations of the unemployment rate. The natural rate of unemployment is from the Congressional Budget Office. The black squares demarcate intervals of increases in the unemployment rate that cumulate to 40 basis points or more and that do not immediately precede a recession peak or follow a recession trough. They fall on the following dates: November 1951, August 1952, February 1963, and July 1976. (Source: Haver Analytics)

estimate of the natural rate of unemployment shown in Fig. 9 is perhaps little more than a trend line. Nevertheless, there is a pattern. Declines in the unemployment rate below the natural rate line require subsequent reversals. Those reversals are associated with recessions.

What Is the Monetary Regime?

Since Henry Thornton (1802 [1939]) central banks have been identified with the institution that determines how a country controls money creation. At present, with more than a century of experience with the Federal Reserve System, one can make some general statements about how the Fed controls money creation and about the nature of the monetary regime.

The monetary regime emerges out of the mutual interaction between the systematic behavior of the Fed and the behavior of the economy. Because the Fed operates with an interest rate target, it controls money indirectly. Endowing money with a stable, well-defined value requires that the Fed creates a stable nominal anchor in the form of an expectation of price stability. In order to do so, it must behave in a predictable way, that is, follow a rule.

However, there remains significant disagreement over the nature of the rule and how the rule interacts with the economy in order to define the monetary regime. Does a monetary policy that is stabilizing provide for a stable nominal anchor (nominal expectational stability) and then allow the price system an unfettered ability to determine real variables like the unemployment rate? Alternatively, does a monetary policy that provides for economic stability require trade-offs in such a way that the Fed can mitigate extreme movements in output through offsetting movements in inflation and conversely? Moreover, since the Great Recession, there are many who believe that the Fed should trade off against its output and inflation objectives in order to achieve an additional financial stability objective.

The lack of consensus about how the monetary regime has evolved and what kind of rule will best provide for economic stability leaves the monetary regime fragile. Potentially, the nature of the monetary regime could vary based on changes in the views of newly appointed policymakers. It is then extremely important that there remains an active dialogue between central bankers and economists centered on learning from historical experience.

Cross-References

- ▶ Deflations in History
- ▶ Experiments with Paper Money
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard

References

Aoki K (2001) Optimal monetary policy responses to relative-price changes. J Monet Econ 48:55–80

Bernanke BS (2008) Remarks on the economic outlook. In: Federal Reserve Board. International monetary conference, Barcelona, 3 June

Blanchard O, Gali J (2007) Real wage rigidities and the new Keynesian model. J Money, Credit, Bank 39:35–65

Blinder AS (1982) The anatomy of double-digit inflation in the 1970s. In: Hall RE (ed) Inflation: causes and effects. The University Of Chicago Press, Chicago, pp 261–282

Board of Governors of the Federal Reserve System. Minutes of the federal open market committee

Bremner RP (2004) Chairman of the fed: William McChesney Martin, Jr. and the creation of the American financial system. Yale University Press, New Haven

Bryan MF, Meyer B (2010) Are some prices in the CPI more forward looking than others? We whink so. Federal Reserve Bank of Cleveland Economic Commentary 2010–2, 19 May

Burns AF (1979) The anguish of central banking. Per Jacobsson Foundation, Belgrade

Cecchetti SG, Feroli ME, Hooper P, Kashyap AK, Schoenholtz KL (2017) Deflating inflation expectations: the implications of inflation's simple dynamics, report prepared for the 2017 U.S. Monetary Policy Forum, sponsored by the Initiative on Global Markets at the University of Chicago's Booth School of Business, held in New York, on 3 Mar

Chandler LV (1971) American monetary policy 1928–1941. Harper & Row, New York

Curdia V, Woodford M (2009) Credit spreads and monetary policy. NBER working paper 15289

Eggertsson GB (2008) Great expectations and the end of the depression. Am Econ Rev 98:1476–1516

Friedman M (1960) A Program for Monetary Stability. Fordham University Press, New York

Friedman M (1968) The role of monetary policy. In: Friedman M (ed) The optimum quantity of money and other essays. Aldine Publishing Company, Chicago, 1969

Friedman M, Schwartz AJ (1963) A Monetary History of the United States, 1867–1960. Princeton University Press, Princeton

Goodfriend M (1993) Interest rate policy and the inflation scare problem. Fed Res Bank Richmond Econ Q 79:1–24

Goodfriend M, King RG (1997) The new neoclassical synthesis. In: Bernanke BS, Rotemberg J (eds) NBER macroeconomics annual. MIT Press, Cambridge, MA

Harrison GL. Harrison Papers. Minutes for meeting of the New York board of directors and for the executive committee of the New York board of directors, Columbia University Rare Book and Manuscript Library

Hetzel RL (1998) Arthur burns and inflation. Fed Res Bank Richmond Econ Q 84:21–44

Hetzel RL (2002a) German monetary history in the first half of the twentieth century. Fed Res Bank Richmond Econ Q 88:1–35

Hetzel RL (2002b) German monetary history in the second half of the twentieth century: from the deutsche mark to the euro. Fed Res Bank Richmond Econ Q 88:29–64

Hetzel RL (2008) The monetary policy of the Federal Reserve: a history. Cambridge University Press, Cambridge

Hetzel RL (2012) The great recession: market failure or policy failure? Cambridge University Press, Cambridge

Hetzel RL (2014) The real bills views of the founders of the fed. Fed Res Bank Richmond Econ Q 100(Second Quarter):159–181

Hetzel RL (2017) A proposal to clarify the objectives and strategy of monetary policy. In: Paper prepared for "Monetary rules for a Post-Crisis World," Mercatus-CMFA Academic conference, 7 Sept 2016, J Macroecon, Dec 2017

Hetzel RL (forthcoming-a) The rise and fall of the quantity theory in nineteenth century Britain or why the early federal reserve system did not understand its responsibility to control the price level? Federal Reserve Bank of Richmond Economic Quarterly

Hetzel RL (forthcoming-b) What caused the great recession in the Eurozone? In: Kaufman G (ed) Innovative federal reserve policies during the great financial crises. World Scientific – Now Publishers Series in Business
Hetzel RL, Leach RF (2001a) The treasury-fed accord: a new narrative account. Fed Res Bank Richmond Econ Q 87:33–55
Hetzel RL, Leach RF (2001b) After the accord: reminiscences on the birth of the modern fed. Fed Res Bank Richmond Econ Q 87:57–64
Hoover H (1952) The memoirs of Herbert hoover: the great depression 1929–1941. The Macmillan Company, New York
Humphrey TM (1986) From trade-offs to policy ineffectiveness: a history of the Phillips Curve. Federal Reserve Bank of Richmond, Oct 1986
Lowenstein R (2015) America's Bank: the epic struggle to create the federal reserve. Penguin Press, New York
Lucas RE Jr, Sargent TJ (1978) After Keynesian macroeconomics. In: Lucas RE Jr, Sargent TJ (eds) Rational expectations and econometric practice, vol 1. The University of Minnesota Press, Minneapolis, pp 295–319, 1981
Madden CH (1959) The money side of the 'Street.' Federal Reserve Bank of New York
Meltzer AH (2003) A history of the federal reserve, vol 1. University of Chicago Press, Chicago, pp 1913–1951
Nelson CW (1977) Rational expectations–fresh ideas that challenge some established views of policy making. Federal Reserve Bank of Minneapolis Annual Report
Potter S (2017) Implementing monetary policy with the balance sheet. Keynote Remarks for ECB Workshop: money markets, monetary policy implementation, and central bank balance sheets, Frankfurt am Mein, 6 Nov 2017
Roosa RV (1956) Federal reserve operations in the money and government securities markets. Federal Reserve Bank of New York
Samuelson PA (1986) Living with stagflation (1979). In: Crowley K (ed) The collected scientific papers of Paul A. Samuelson, vol 5, no. 379. MIT Press, Cambridge, MA, p 972
Samuelson P, Solow R (1966) Analytical aspects of anti-inflation policy (1960). In: Stiglitz J (ed) The collected scientific papers of Paul A. Samuelson, vol 2, no. 102. MIT Press, Cambridge, MA, pp 1336–1353
Stock JH, Watson MW (2007) Why has U.S. inflation become harder to forecast? J Money, Credit Bank 39:3–33
Taylor JB (2009) Getting off track: how government actions and interventions caused, prolonged, and worsened the financial crisis. Hoover Institution Press, Stanford
Thornton H (1939) An enquiry into the nature and effects of the paper credit of Great Britain (1802) and Two Speeches (1811)., edited with an Introduction by F. A. v. Hayek. Rinehart and Co., New York
U. S. Congress. Statement of Martin, William McChesney (1959) January 1959 Economic Report of the President. In: Hearings before the Joint Economic Committee, U.S. Congress 86th Congress 1st sess. pp 495–535
U.S. Economic report of the President transmitted to the Congress annually together with the annual report of the council of economic advisers. Washington, DC: Government Printing Office, various issues
Velde F (2004) Poor hand or poor play? The rise and fall of inflation in the U.S. Fed Reserve Bank Chicago Econ Perspect 28(Quarter 1):34–51
Wheelock DC (1991) The strategy and consistency of Federal Reserve Monetary Policy, 1924–1933. Cambridge University Press, Cambridge
Wicker E (1966) Federal reserve monetary policy, 1917–1933. Random House, New York
Woodford M (2003) Interest and prices: foundations of a theory of monetary policy. Princeton University Press, Princeton

The Historical Evolution of Monetary Policy (Goals and Instruments) in Japan: From the Central Bank of an Emerging Economy to the Central Bank of a Mature Economy

36

Masato Shizume

Contents

Abstract

This chapter reviews monetary policy developments in Japan from the late nineteenth century to the early twenty-first century, focusing on its goals and instruments. The goals of monetary policy have been defined based on the goals of the nation as a whole. They have changed over time in accordance with changes in the goals of the nation. On the long road to an industrialized, developed country from the late nineteenth century through the 1970s, the ultimate goal of monetary policy was to provide money for development while maintaining financial and price stability. Once Japan had achieved industrialization and become a leading economic power, the primary goal shifted from

M. Shizume (✉)
Waseda University, Tokyo, Japan
e-mail: masato.shizume@waseda.jp

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_34

development to stability. Instruments of monetary policy have also changed over time while being constrained by the availability of such instruments in the financial markets, even though the central bank kept pursuing policy innovations.

Keywords
Monetary policy · Policy goals · Policy instruments · Price stability · BOJ

Introduction

This chapter reviews developments in the monetary policy of the Bank of Japan (BOJ) from the late nineteenth century to the early twenty-first century, focusing on its goals and instruments. For further reference, the Bank of Japan (1982, 1983a, 1983b, 1984, 1985, 1986a) is available as the official history of the BOJ. Also, Ishii (2001) reviews the BOJ's monetary policy from a historical perspective. Shizume (2018) reviews history of the BOJ focusing on its institutional and legal flamework and its role in the national economy, complementary to this chapter.

Historically, the goals of monetary policy in Japan have been defined based on the goals of the nation as a whole. They have changed over time in accordance with changes in goals of the nation. On the long road to an industrialized, developed country from the late nineteenth century through the early 1970s, the ultimate goal of monetary policy was to provide money for development while maintaining financial and price stability. Once Japan had achieved industrialization and become a leading economic power, the primary goal shifted from development to stability. Meanwhile, instruments of monetary policy have also changed over time while being constrained by the availability of such instruments in the financial markets, even though the central bank kept pursuing innovation in policy tools (Table 1).

When the BOJ was established as the country's central bank in 1882, Japan was under the process of dynamic transformation from a feudal society to a modern one. The founders of the BOJ expected it to mobilize financial resources toward the dual national goal of "wealthy nation, strong military."

In peacetime, the BOJ's balance sheet was around 15 percent of Japan's GDP before WWII and has been around 10 percent since. During and immediately after wars, it often exceeded 20 percent (Graph 1). Inflation accelerated when the expansion of the BOJ's balance sheet was prolonged and associated with a proportionate increase in banknote circulation on the liability side, as seen during the Asia-Pacific War. In contrast, during the Sino-Japanese and Russo-Japanese wars, balance sheet expansions were brief and associated with increases in the government deposit, not with banknote circulation. Also, since the late 1990s, the expansion of the balance sheet has been associated with an increase in private banks' deposits. In these cases, inflation subsided (Graphs 2 and 3).

In this chapter, I use the word "the Asia-Pacific War" for the full-scale war between Japan and its main trading partners, China and the United States, between 1937 and 1945. Japan had been engaged in an undeclared war with China following the Mukden Incident in 1931, but this escalated to a full-scale war after the Marco

Table 1 Chronology of developments in relation to the BOJ and the Japanese economy

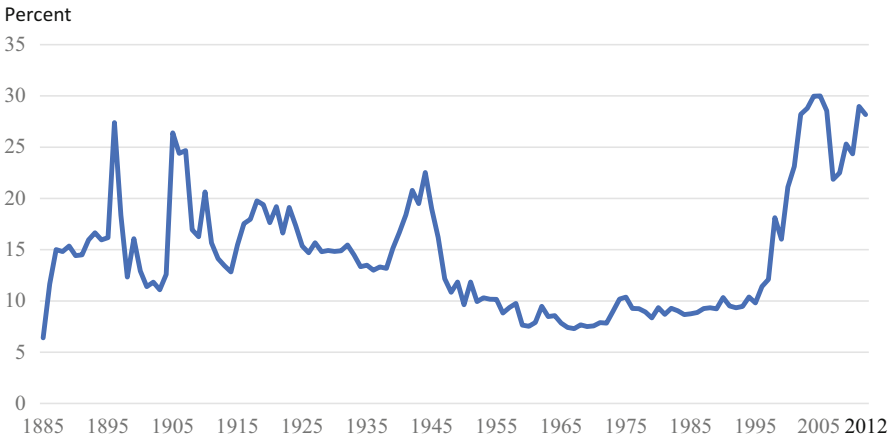
Year	BOJ	Year	Japanese economy/politics
		1859	Opening of the treaty ports
		1868	Meiji Restoration
1882	BOJ established		
1885	First BOJ notes (silver standard)		
1897	Adoption of the gold standard	1894–1895	Sino-Japanese War
		1904–1905	Russo-Japanese War
1917	Suspension of the gold standard	1914–1918	WWI
		1927	Showa Financial Crisis
		1929	Crash of NY stock exchange (the Great Depression began)
1930	Return to the gold standard		
1931	Departure from the gold standard	1931	Mukden Incident
			Takahashi Economic Policy began
		1937	Full-scale war with China began
1942	BOJ Act of 1942	1941	War with USA began
1946	Emergency Financial Measures	1945	Japan surrendered
1949	Amendment of the BOJ Act	1951	San Francisco Peace Treaty
		1960	National Income-Doubling Plan
1973	Move to flexible exchange rates	1973	First oil crisis Plaza Accord Louvre Accord
		1985	
		1987	
1998	BOJ Act of 1998	1997–1998	Heisei Financial Crisis
1999	Unconventional monetary policy began		

Source: Shizume (2018, 329; text)

Polo Bridge Incident in 1937. Subsequently, Japan entered war with the United States when its navy attacked Pearl Harbor in 1941.

The BOJ used lending to private banks (discounting of bills and loans with collateral) as its main policy instrument, especially from the 1880s through the 1920s and from the 1950s through the 1960s. In addition, during its early days, it accepted private equities as well as commercial bills as collateral to fill the gap between its need to achieve policy objectives and the scarcity in existing financial instruments. It also extended credit to the government during wars including the Sino-Japanese War, the Russo-Japanese War, and the Asia-Pacific War.

During and after the major wars with Japan's main trading partners (China and the United States) in the 1930–1940s, monetary policy was overwhelmed by war finance. The BOJ underwrote government bonds and sold them to the private sector,



Graph 1 BOJ balance sheet to GDP ratio. (Source: The Bank of Japan 1986b, 272–304)

calling these "open market operations." During and after WWII, Japan experienced the worst inflation in its modern history (Graph 3).

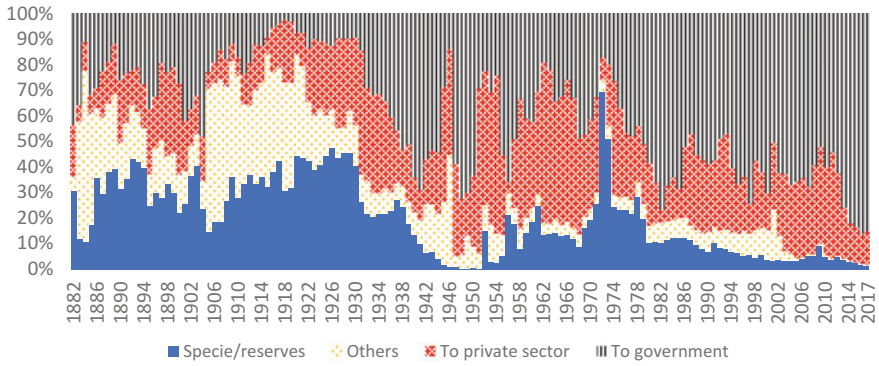
During the high growth period from the late 1950s through the early 1970s, the BOJ conducted monetary policy to facilitate economic growth while ensuring the sustainability of balance of payments. The financial system was heavily regulated, and the BOJ used interest-rate regulations based on the official discount rate (ODR) level as its main policy instrument. It also used "the window guidance" (moral suasion of private banks) and the reserve/deposit ratio as complementary measures to the interest rate policy.

Once high growth was achieved, the primary goal of monetary policy shifted from promoting economic development to maintaining economic stability. At the same time, along with deregulation of the financial system, the BOJ transformed its policy instruments from regulation-based ones to market-based ones. Since the second half of the 1990s, the central bank has faced new challenges with economic stagnation and the zero lower bound (ZLB). In response, it has employed unconventional monetary policy measures, expanding its balance sheet to an unprecedented level.

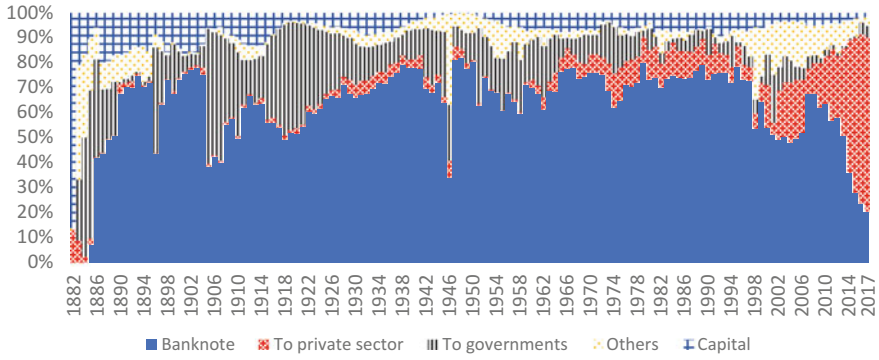
In a nutshell, the relative size of the BOJ's balance sheet compared with the size of the national economy had been stable during peacetime until the early 1990s. The stability largely reflected a stable demand for the central bank's money and the accommodative nature of the BOJ's policy response to such demand. In most periods of the history of the BOJ, banknotes in circulation have accounted for around 60–80 percent of its liability with some exceptions in its early days and since the late 1990s as well as temporary drops during wars (Graph 2).

However, war finance during the Asia-Pacific War disrupted the stability of the central bank's policy stance and caused rampant inflation. As for the recent experience since the 1990s, the "unconventional monetary policy" with a massive expansion of the balance sheet in peacetime marks an exception in terms of the size of the

<Assets>



<Liabilities>

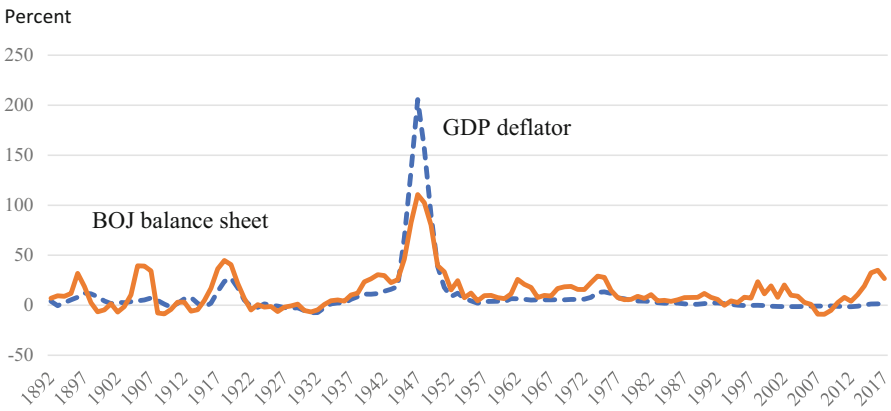


Graph 2 Shares of the components of the BOJ balance sheet. (Source: The Bank of Japan 1986b, 272–304)

central bank's balance sheet in peacetime and of the relationship between the size of the balance sheet and the stability of prices.

Building a Modern Monetary System (1882–1896)

When the Bank of Japan (BOJ) was established in 1882, Japan was in the process of transformation from a feudal to a modern society. The founders of the BOJ expected it to mobilize financial resources toward the dual national goal of "wealthy nation, strong military." To this end, they first needed to establish a credible institution that could ensure stability of the value of the national currency. In terms of policy instruments, the BOJ used lending to private banks (discounting of bills and loans with collateral) as its main policy instrument.



Graph 3 Annual change in BOJ balance sheet and GDP deflator (3-year averages). (Sources: The Bank of Japan (1986b, 272–304), long-term historical statistics, and the Cabinet Office)

It accepted private equities as well as commercial bills as collateral. It extended credit to the government in times of war.

The Establishment of the BOJ

When the BOJ was established in 1882, Japan was in the process of dynamic transformation from a feudal to a modern society. After Japan was forced to open its treaty ports in 1859, the nature of its economy changed from a closed to an open economy. Under the unequal treaties with western powers, Japan had no tariff autonomy, and domestic industries had to compete with their western counterparts without firm protection from the government. Annexations of territories and colonization of neighboring countries by western powers indicated that Japan was no exception. A sense of sovereign crisis mobilized the nationalists in the *samurai* (warriors) class to bring about the Meiji Restoration in 1868. The Meiji government, which replaced the Tokugawa Shogunate (1603–1867), conducted a series of economic, political, and social reforms under the banner of "wealthy nation, strong military." The reforms included the abolishment of clans and introduction of the prefectural system, a new uniform land tax system, national conscription and elimination of stipends to former samurai, debt consolidation of clans, and the integration of the national monetary and financial systems. In 1871, the government declared the introduction of the gold standard and minted gold specie, but it failed to build a domestic monetary system based on gold (the Bank of Japan 1982, 3–12).

The establishment of the BOJ was a crucial element of the national program toward the goal of "wealthy nation, strong military." It was expected to mobilize national financial resources toward economic development and for military requirements. The founder of the bank, Finance minister Masayoshi Matsukata, intended to make it the hub of the national economy. He declared at the opening of the BOJ that he expected "the bank to work as the heart of the national economy, pumping money

to the entire economy, just as the heart pumps blood to the entire body" (the Bank of Japan 1982, 123). According to him, the aims of the BOJ were fivefold: (1) to regulate the regional supply and demand of funds, (2) to increase the financial capacity of private banks and companies, (3) to reduce domestic interest rates by removing financial bottlenecks, (4) to act as the government's treasury management agency, and (5) to settle and adjust the balance of payments in international trade and finance (the Bank of Japan 1982, 123–136).

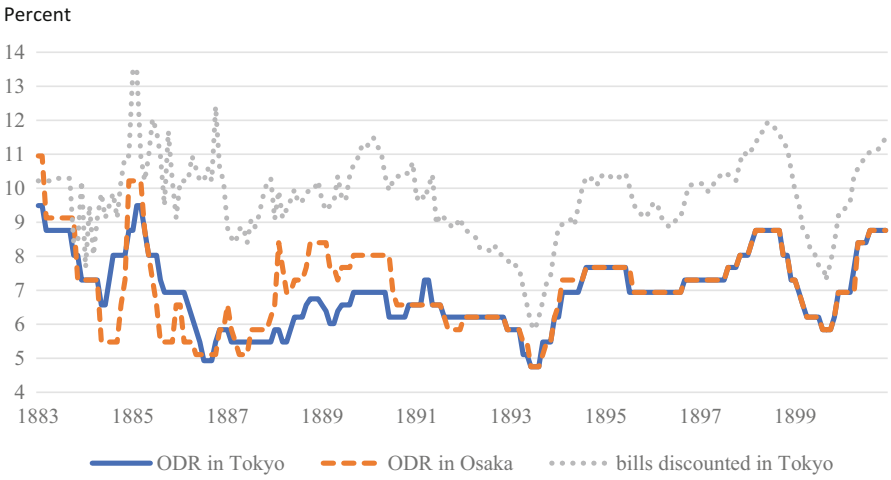
Before the establishment of the BOJ, monetary and fiscal policies were unseparated. Government and private banks (national banks) issued paper money besides specie and subsidiary coins minted by the government. Government notes and national banknotes were inconvertible though denominated in yen, the unit of account introduced in 1871. At the end of 1881, the total amount of currency in circulation was 196 million yen, comprising 119 million yen in government notes, 34 million yen in national banknotes, 14 million yen in gold specie, nine million yen in silver specie, and 19 million yen in subsidiary coins (the Bank of Japan 1986b, 414–415).

National banks were private banks established with national charters. Japan had introduced a national banking system modeled after the United States in 1872, under which national banks had been allowed to issue banknotes. Under the initial National Bank Act of 1872, which imposed strict regulations including the convertibility requirement and a high reserve ratio (2/3rd of banknotes in circulation in specie), only four national banks were established. However, the elimination of the convertibility clause and relaxation of the reserve ratio (1/5th of banknotes in circulation in government notes) in 1876 brought a banking boom, and 153 national banks were established by 1979 (the Bank of Japan 1982, 16–31).

These paper monies depreciated substantially against silver after the 1877 Satsuma Rebellion, the last civil war in Japan. Finance Minister Matsukata implemented an austerity policy during the early 1880s to enable paper monies to circulate at par (*Matsukata Zaisei*, or Matsukata Economic Policy). The BOJ first issued banknotes in 1885. These were convertible to silver, the international currency in Asia-Pacific region at that time (the Bank of Japan 1982, 107–112, 280–293).

Operations of the BOJ During Its Early Days

Though the founder of the BOJ designed it to be the hub of the national financial system, engaging in the business of commercial banking, such as discounting bills, the reality of the Japanese financial markets did not allow it to engage in such activities. At the onset, it had only one branch in Osaka apart from its Head Office in Tokyo. To extend its network, it entered correspondence contracts with private banks. Graph 4 shows that the BOJ's Tokyo Head Office and Osaka Branch ODRs were different until the early 1890s, indicating that the financial markets of these two centers were not fully integrated. In the private sector, the format of bills of exchange had not been standardized yet, and short-term credit was predominantly granted through various kinds of certificates, rather than the standardized form of commercial bills to be discounted. Due to the lack of credible collateral for commodities, private banks often lent without



Graph 4 BOJ official discount rates and short-term market interest rates (1883–1900). (Sources: The Bank of Japan 1986b, 350–365; Asahi Shinbun Sha 1930)

collateral or only with real estate or equities, which were not official eligible collateral for BOJ loans. The interest rate seasonality persisted until around the end of the nineteenth century even as the BOJ attempted to respond to demand for central bank money in an elastic manner (Fukuda 1995, 63–65, 71).

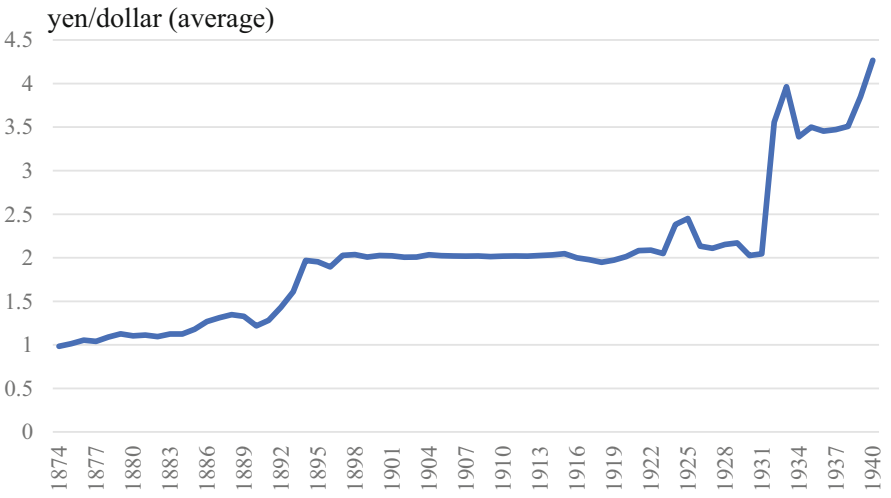
From the onset, the BOJ was heavily involved in industrial policy by providing liquidity in the short run and capital in the long run to the industrial sector. Among others, the BOJ set a large line of credit and generous conditions for lending to the Yokohama Specie Bank (YSB). YSB was established in 1880 and dominated in Japan's international trade finance (the Bank of Japan 1982, 382–388, 415–421). The BOJ kept promoting industrial development, exports, and import substitution (the Bank of Japan 1982, 337–343, 425–426; Ishii 2001, 19–20).

The BOJ conducted a lender of last resort (LOLR) operation for the first time in the wake of the 1980 financial panic. After the panic, it became more aware of its role in the financial markets and played a more active role. It began to set its ODR so as to lead market participants rather than follow them. It formally started accepting equities as collateral when it extended credit to other banks, not only in a panic but also during ordinary times, in an effort to promote industrial development (the Bank of Japan 1982, 428–438, 449).

Guns or Butter (1897–1936)

The Adoption of the Gold Standard

At the turn of the century, Japan successfully joined the club of "first-class nations," having escaped from the risk of colonization. When Japan revised the unequal treaty with Great Britain in 1894, it succeeded in raising tariffs and abolishing British consular



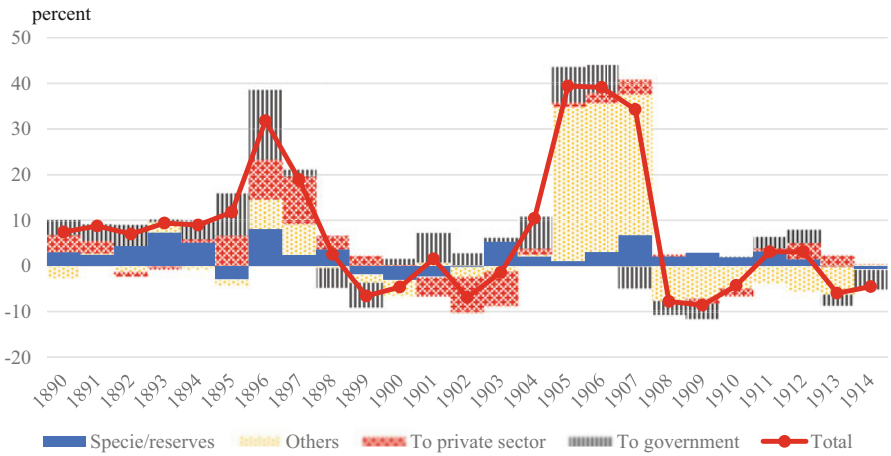
Graph 5 Exchange rates (yen/dollar, 1874–1940). (Sources: The Bank of Japan 1986b, 424–431)

jurisdiction and the unilateral most-favored-nation status. The revision of unequal treaties with other western countries followed, culminating in full tariff autonomy for Japan in 1911. By this point, Japan was recognized as an emerging empire, unrivaled in Asia and a full member of the international community of western powers.

As its external status improved, Japan faced growing conflict between its two national goals of a wealthy nation and a strong military, both needing financial capital. It sought a new source of financial capital, external finance, to deal with the new challenge and with less concern of losing national sovereignty. It adopted the gold standard in 1897, 26 years after its unsuccessful declaration in 1871. It did so using indemnity from China after its victory in the Sino-Japanese War of 1894, following the precedent of Germany in the wake of the Franco-Prussian War of 1870–1871. When Japan set the new gold parity, yen had depreciated to half its value against the US dollar compared with the earlier parity, which was set in the early 1870s (Graph 5).

When Japan adopted the gold standard, the Japanese government expected two effects *ex ante* – that it would stimulate international trade and attract international investors. Empirical evidence shows that investors were not fully satisfied with the country's merely joining the gold standard, although international trade was stimulated as expected. There was no investment boom right after the adoption, and Japanese government bonds still required a high premium in the international capital markets. Japan finally earned a reputation as a "first class nation" in the international capital markets only after defeating the Russian army and navy during the Russo-Japanese War (1904–1905), thereby reducing its geopolitical risk (Matsukata 1899; Mitchener et al. 2010; Smethurst 2007, p.170).

During the Sino-Japanese War and the Russo-Japanese War, the BOJ's balance sheet expanded, reflecting its extension of credit to the government to finance wars on the asset side (Graph 6). During the latter war, especially, it conducted a massive price-keeping operation for government bonds in the domestic market and sent its



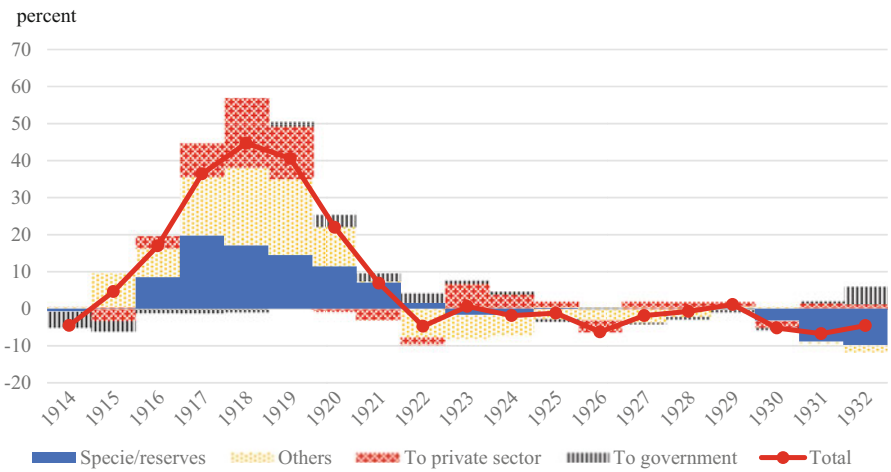
Graph 6 Asset components of the average annual balance sheet change in three years (1890–1914). (Source: The Bank of Japan (1986b, 274–283). Note: Average annual balance sheet changes are three-year moving averages)

Vice Governor, Korekiyo Takahashi, to London, New York, and other financial centers in Europe to issue Japanese bonds in foreign markets. The funds raised in the international financial markets were used to buy weapons and as the main body of foreign reserves during and after the war (the Bank of Japan 1983a, 160–182; Smethurst 2007, 141–187).

After the Russo-Japanese War, monetary and fiscal policies were constrained by a heavy burden of external debt. Japan's current account continued to record deficits due to the mounting burden of debt servicing. In the face of the declining confidence of international investors, Japanese authorities tightened fiscal and monetary policies to restore confidence. Graph 6 shows a contraction in the BOJ's balance sheet after the war. And yet, Japan had difficulties financing its current account deficits in the international financial markets (Shizume 2018, 340).

The First World War and Its Aftermath

WWI changed the external conditions of the Japanese economy. Distant from the main battlefields in Europe, Japan accumulated a huge current account surplus. It obtained German territories in Asia at little military cost. It joined the United States as a creditor nation. Domestic industries enjoyed a long boom period thanks to massive exports to European countries at war, their colonies, and the United States, where the economy was booming. As a result, the balance sheet of the BOJ expanded due to accumulated specie, foreign reserves, and credit to the private sector (Graph 7). In Graph 7, a large portion of "others" consists of the BOJ's holding balances in agents in London and New York, or de facto foreign reserves.



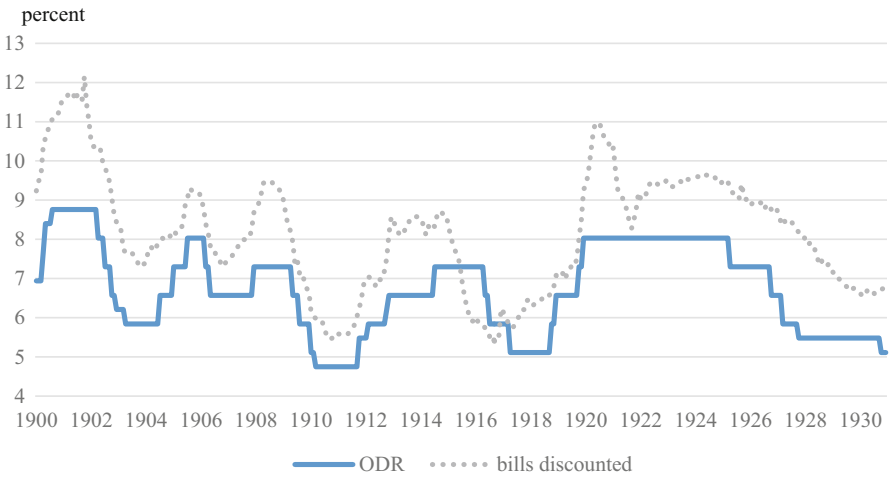
Graph 7 Asset components of the average annual balance sheet change in three years (1914–1932). (Source: The Bank of Japan (1986b, 282–289). Note: Average annual balance sheet changes are three-year moving averages)

Japan experienced a prolonged period of high growth accompanied by massive inflation as commodity prices soared globally during and after WWI. The Japanese economy grew by 40 percent in terms of real output and more than threefold in nominal terms from 1914 to 1920. During the boom, many companies engaged in speculation and/or built excess production capacities with bullish expectations of the prospects for the Japanese economy (Shizume 2017, 133–134).

With growing confidence in Japan's economic and political power, Japanese policymakers expanded the nation's goal from maintaining its own independence to building an empire in the Asia-Pacific region while retaining the slogan of "wealthy nation, strong military." Though the windfall effects of WWI eased the conflict between the two goals temporarily, militaristic expansion into other parts of Asia became, in the long run, the largest potential threat for the sustainability of public finance and the conduct of monetary policy, and a heavy burden on the private sector, impeding its growth (Shizume 2018, 340).

Japan suspended the gold convertibility of its banknotes in September 1917, following after the United States. At that time, these two countries were virtually the only ones remaining on gold, and there was no reason for Japan to continue on gold once the United States exited. Since then, the yen/dollar exchange rate has fluctuated (Graph 5).

After WWI, the Japanese economy suffered from economic stagnation, price deflation, and bad-debt problems. Financial institutions struggled with mounting bad loans, which triggered financial panics in 1920 and 1922. Then, in 1923, the Great Kanto Earthquake hit the Tokyo metropolitan area. The BOJ responded to the tragedy by extending special loans to financial institutions located in the devastated areas. However, contrary to the BOJ's intensions, the special loan facility was



Graph 8 BOJ official discount rates and short-term market interest rates (1900–1930). (Sources: The Bank of Japan 1986b, 350–365; Asahi Shinbun Sha 1930)

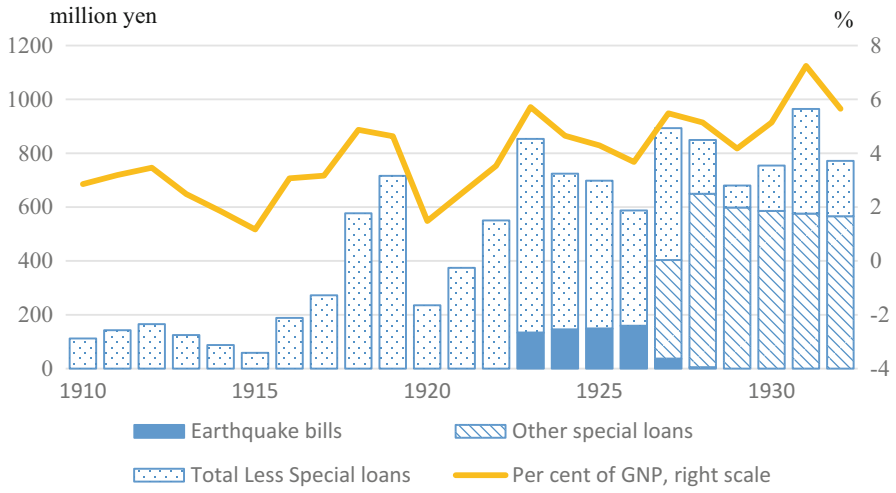
abused by some banks, which were already distressed by bad loans, evincing a moral hazard problem (Shizume 2017, 135–139).

In the spring of 1927, financial panic spread nationwide (the Showa Financial Crisis). The BOJ acted as the LOLR during the crisis, swiftly extending credit to combat bank runs. The crisis accelerated the resolution of the bad-debt problems and banking reforms. The BOJ's potential loss from the special loans related to the Showa Financial Crisis was guaranteed by the government. In 1928, the authorities introduced a new prudential policy scheme – encouraging mergers and acquisitions, setting new regulations, and instituting dual inspections by the Ministry of Finance (MOF) and the BOJ (Shizume 2017, 141–142; Shizume 2018, 341–342).

Thanks to the BOJ's extended credit and banking reforms, domestic financial markets were finally stabilized. Market interest rates, which had stayed high until the spring of 1927, reflecting high risk premium, went down through the rest of the 1920s (Graph 8). However, the BOJ had to bear a new constraint on its monetary policy. Though guaranteed by the government, the special loans accounted for a large part of its private lending and could not be repaid in the short term. As a result, the BOJ lost flexibility in its monetary policy operations (Graph 9; Shizume 2017, 143–145; Shizume 2018, 342–343).

Japan During the Great Depression

In 1929, the government decided to return to the gold standard. Judging that domestic financial stability was finally restored and Japan was ready to return to "normality" as per pre-WWI standards, the government did so in January 1930. However, the timing was bad because the Great Depression was spreading and intensifying globally, and

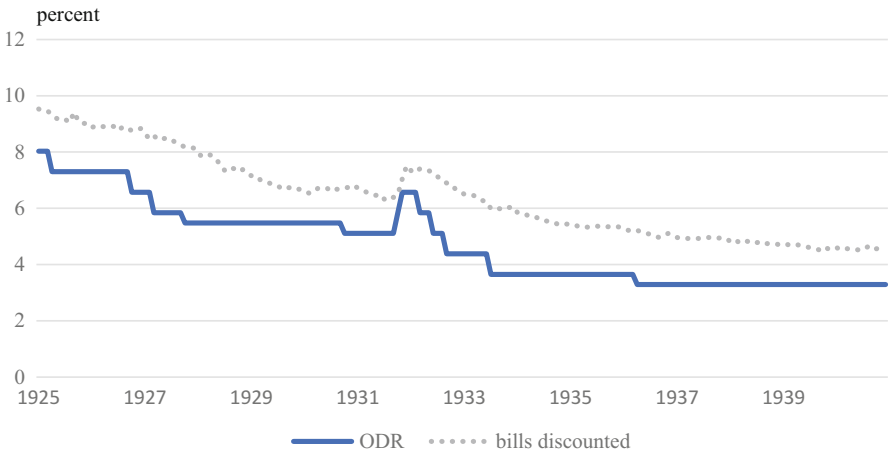


Graph 9 BOJ loans to the private sector. (Source: The Ministry of Finance, *Reference Book for Financial Matters*, annual editions)

the move brought a severe economic contraction and deflation to the Japanese economy. Already depressed by fiscal and monetary austerity policies in preparation for returning to the gold standard, Japan experienced double-digit price deflation in 1930–1931 (Metzler 2006, 199–239; Shizume 2012, 211–228).

When Britain departed from the gold standard in September 1931, the yen was under speculative attack in the international financial markets because investors anticipated that Japan would soon follow suit. In response, the BOJ raised its lending rates twice to protect the yen (Graph 10), which only caused further contraction of the domestic economy and had little effect in terms of preventing capital outflows. In the end, Japan lost almost 60 percent of its reserves (the Bank of Japan 1986b, 334).

In December 1931, veteran finance minister Korekiyo Takahashi returned to office for the fifth time in his career. He led a drastic macroeconomic stimulus package including departure from the gold standard, an increase in fiscal expenditure, and an accommodative monetary policy (*Takahashi Zaisei*, or Takahashi Economic Policy). From 1930 to 1933, the yen depreciated to almost half its value against the US dollar. The departure from the gold standard also freed the government and the BOJ from the constraints of maintaining gold parity. The government increased its expenditure toward public works in rural areas and the war in mainland Asia. The BOJ supported public finance by underwriting government bonds and selling them to financial institutions in accordance with conditions in the financial markets. The BOJ called it an "open market operation." At the same time, the BOJ relaxed its monetary policy by reducing interest rates (Graph 10, Shizume 2012, 211–228; 2018, 343). Some observers argue that the scheme involving the underwriting of government bonds by the central bank dampened fiscal discipline because the government no longer had to worry about financial market conditions when it decided spending in annual budgets (Shizume 2012, 211–228; 2018, 344).



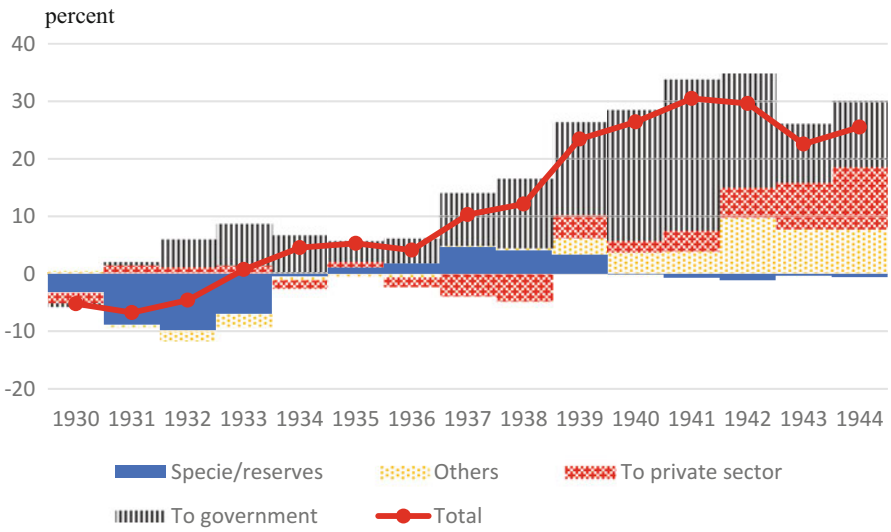
Graph 10 BOJ official discount rates and short-term market interest rates (1925–1940). (Sources: The Bank of Japan 1986b, 350–365; Asahi Shinbun Sha 1930)

Thanks to Takahashi's stimulus package, the Japanese economy recovered from the Great Depression earlier than others. It recorded growth at an annual rate of 6 percent from 1931 to 1936, with low inflation.

Meanwhile, the Mukden Incident of September 1931 marked a critical point for Japan, pushing it toward war with its main trading partners, China and the United States. The Japanese army stationed in Manchuria staged it by falsely claiming an attack by the Chinese army and starting an undeclared war. The Japanese government initially tried to make a deal with the Chinese government for a stand-down but failed because of provocative actions by the Japanese troops and approved extra budget for the actions *ex-post facto*.

The consequence of the Mukden incident was, in retrospect, the government's loss of control over its own troops in China. As a result, fiscal and monetary policies became overwhelmed by military expenses. The BOJ's scheme of underwriting government bonds lacked market-based checks and balances for fiscal discipline. In the budgetary process, the military could submit a request for whatever amount it wanted to the MOF, and if the MOF approved the amount, then the BOJ had to underwrite government bonds to finance it. When Korekiyo Takahashi cut military expenditure for fiscal 1936 on his responsibility and authority as finance minister, a group of soldiers assassinated him in February 1936 (Smethurst 2007, 268–298; Shizume 2012, 211–228).

From 1932 through the early 1940s, the largest increase in the BOJ's balance sheet came from "credit to the government" (Graph 11). The sum of "specie reserves" and "others" fell below half of the BOJ's balance sheet even as the share of "credit to the government" continued to grow. As a large portion of "others" consisted of de facto foreign reserves, the BOJ increased credit to the government instead of holding specie and foreign reserves after its departure from the gold standard (Graph 2).



Graph 11 Asset components of the average annual balance sheet change in three years (1930–1944). (Source: The Bank of Japan (1986b, 278–292). Note: Average annual balance sheet changes are three-year moving averages)

War Finance and Inflation (1937–1951)

The war with China escalated in 1937, when two armies accidentally came face to face at Marco Polo Bridge near Beijing. Once again, the Japanese government tried to call a ceasefire but failed because of provocative actions by the Japanese army.

The breakout of WWII in Europe in 1939, economic sanctions against Japan by the United States and other countries, and Japan's attack on Pearl Harbor in 1941 further exacerbated the external condition of the Japanese economy. Since Japanese industries had relied heavily on overseas supplies of essential materials such as fuel oil and iron, they now faced severe resource shortages. At the same time, the escalation of the war intensified demand for goods and services related to the war and, squeezed domestic physical, human, and financial resources to the extreme (Shizume 2018, 344–345).

To deal with the increasing inflationary pressures owing to external constraints and massive war expenses, the authorities introduced direct control measures, including capital controls, financial repression, and price controls. In 1937, the government passed three wartime control laws, one of which, the Temporary Capital Adjustment Law, required government permission for long-term fundraising for capital formation.

From 1938 to 1945, the government formulated annual Goods Mobilization Plans. In each year's plan, the government estimated the supply of essential materials from imports and domestic production and allocated it to military and civilian use.

The government continued to make these plans until 1945, but never achieved the estimated supply mainly due to essential material imports falling short. Also, in 1938, the National General Mobilization Law was promulgated, stipulating that the government may "mobilize" personnel and material resources for national defense in wartime. Financially speaking, all funds including short-term operations were subject to government permission. In 1939, to deal with general price hikes in the wake of the breakout of WWII, the government issued an imperial order intended to freeze wages and the prices of all commodities (Nakamura 1994, 87–101).

Under the wartime regime, the BOJ devoted itself to the national objective, to fight the war, under the government's control. It underwrote war bonds and then resold them to private financial institutions. It rationed all private credit, which was subject to government control. It provided money to war industries (Graph 11).

In 1942, the Bank of Japan Act of 1882 was replaced by the new Bank of Japan Act, which defined the BOJ as a state entity and strengthened the government's control over its operations. Under the new law, the BOJ was literally to devote itself entirely to achieving the national goal (Shizume 2018, 345–346).

Despite all the efforts by the authorities, wartime control of the economy was poorly managed. Actual production of goods for military and daily necessities fell short of the government's initial plans. Due to a huge gap between supply and demand, black markets prevailed in an open secret, and prices soared (Nakamura 1994, 117–122; Shizume 2018, 346).

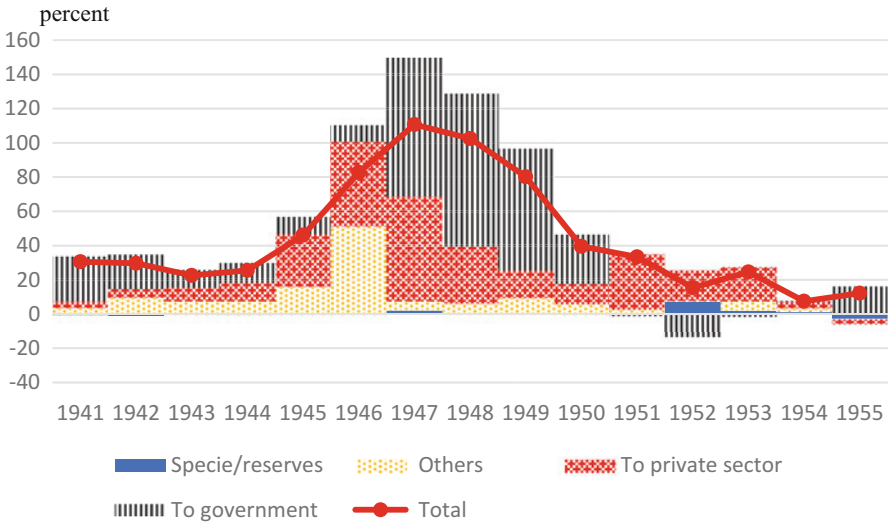
When Japan lost the war in 1945, the economy was in a state of disorder. The labor force badly needed to be restructured from military services and war-related industries to civil industries. One-fourth of total physical assets were lost during the war. Once the war ended, black markets were no longer a secret; brokers sold goods at skyrocketing prices on the streets of cities (Nakamura 1994, 131–132; Shizume 2018, 346–347).

Right after the war, two government measures accelerated inflation. The government paid war expenses to compensate repatriated soldiers and civilians who had worked for the military and the war industries or had lost their houses in air raids. This fueled the sharp rise in inflation. The government also lifted price controls on fresh foods, intending to promote their supply, but only resulting in further price hikes (Graph 3, Shizume 2018, 347).

One piece of good news for the Japanese economy was the demilitarization of Japanese society and economy, which freed fiscal and monetary policies and the public sector from the burden of military spending. Now, the national goal was reset to "wealthy nation without a strong military" even though the legal and institutional setting of the BOJ had not changed much.

The occupation forces led by the United States decided to take the form of an indirect government. The Japanese government and the central bank continued to conduct policies under the directives of the General Headquarters (GHQ) of the Supreme Commander for the Allied Powers (SCAP).

The urgent challenge for the government and the BOJ right after WWII was to balance stabilization and reconstruction. Since inflation was so severe while the need



Graph 12 Asset components of the average annual balance sheet change in three years (1941–1955). (Source: The Bank of Japan (1986b, 290–297). Note: Average annual balance sheet changes are three-year moving averages)

for structural adjustments was huge, monetary policy went back and forth between the two objectives.

In February-March 1946, the government implemented the Emergency Financial Measures. All cash had to be deposited with financial institutions, new banknotes were issued, limits were set on withdrawals, and strict government control over prices was resumed. Inflation in the black markets subsided at least temporarily (Nakamura 1994, 149–150; Shizume 2018, 347).

In 1947, the Public Finance Act was promulgated, banning the BOJ from underwriting government debt. However, prior to the promulgation of the law, the government had initiated the Priority Production System to concentrate physical and financial resources in key industries such as coal and steel, intending to remove bottlenecks in production and power up the whole economy. The Reconstruction Financial Bank (RFB) was established to finance key industries, issuing bonds to be underwritten by the BOJ. Technically, RFB bonds were not government bonds. In effect, the provision of central bank credit through RFB bonds bloated up the BOJ's balance sheet and reignited rampant inflation (Graph 12, the Bank of Japan 1985, 102–104, Shizume 2018, 348).

In 1949, a Detroit banker, Joseph Dodge, who became the adviser to General Douglas MacArthur, the Supreme Commander of the GHQ, initiated a stabilization plan called the "Dodge Line." This involved implementing fiscal austerity and fixing the yen's exchange rate at 360 yen to the US dollar, to prepare for joining the Bretton Woods System. Postwar inflation subsided, but a severe recession hit the economy.

Also, in 1949, the Bank of Japan Act was amended, adding a policy board modeled after the Federal Reserve System into the BOJ's governance structure.

However, this had little effect on the monetary-policy-making process (Shizume 2018, 349).

The BOJ conducted monetary policy in a way as to accommodate the fiscal austerity and the resulting recession, by extending credit to the private sector. It bought government bonds and RFB bonds to support key industries, small and medium sized enterprises (SMEs), and the agricultural sector (Graph 12). The Japanese economy recovered thanks to massive procurement demand during the Korean War (1950–1953; the Bank of Japan 1985, 354–364; Shizume 2018, 348–349).

Japan concluded the San Francisco Peace Treaty in 1951. Japan resumed its sovereignty when the treaty came into effect in 1952 and joined the Bretton Woods System (International Monetary Fund: IMF) in the same year.

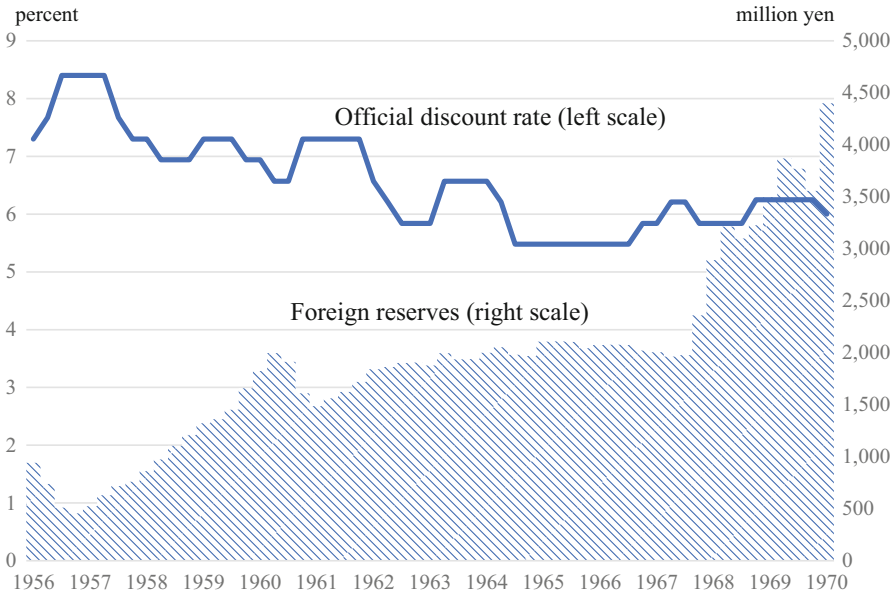
During the reconstruction period, financial intermediation shifted from a combination of banks' balance sheets and securities to mainly through banks. In order to bring down inflation, the whole financial system was kept heavily regulated as under the previous wartime regime. With rampant inflation, long-term debts lost their value in real terms, and equity finance virtually did not function under such heavy regulation. As a result, private companies had to rely on bank lending for fundraising. Meanwhile, following the Public Finance Act of 1947, the government employed the balanced-budget principle until 1965, issuing virtually no bonds (Ito 1992, 165).

As domestic financial intermediation shifted to banks' balance sheet, instruments of monetary policy began to be focused on banks' balance sheet. When allocating credit to private banks, the BOJ prioritized essential industries for national economic reconstruction such as coal, steel, and fertilizers, and then exporting industries such as textiles and machineries. The facilities included reduced-rate lending and the rationing of funds. The facilities for export promotion were retained for years, throughout the following high growth period and until the Smithsonian Agreement in 1971 though the target industries changed over time (the Bank of Japan 1985, 90–92, 123–132, 337–338, 589).

Financing High Growth (1952–1970)

The Japanese economy continued to grow at an annual rate of around 10 percent from the early 1950s through the early 1970s ("the high-growth period"). The growth elevated Japan from a developing country to an industrialized, developed country.

During the high growth period, the BOJ funneled short-term liquidity and long-term capital into the economy while maintaining a fixed exchange rate against the US dollar. The BOJ conducted monetary policy to facilitate economic growth while ensuring sustainability of the balance of payments. As national governments were allowed to regulate international capital flows under the Bretton Woods System, the overall balance of payments was largely determined by the current account balance.

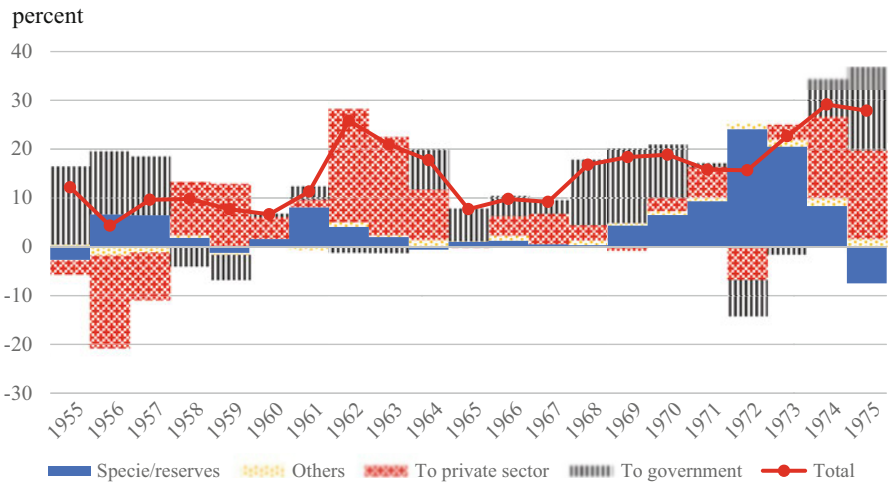


Graph 13 Official discount rates and balance of payments (1956–1970). (Source: Bank of Japan, *Economic Statistics Annual*, annual editions)

The BOJ kept domestic interest rates as low as possible to promote industrial activities and raised the interest rate to curb demand when the balance of payments deteriorated and foreign reserves fell to a certain level (Graph 13, Ito 1992, 114–118).

During the high-growth period, the BOJ's main policy instrument was the ODR. The whole financial system was heavily regulated, and debtors depended mainly on banks for fundraising. Interest rates on deposits and the lending rates of private banks were set according to the ODR (Ito 1992, 114–115, 119). However, *effective* interest rates for banks and their borrowers fluctuated even when contracted interest rates did not change. During this period, banks required their customers to retain a portion of lending as compulsory deposits. By changing the ratio of the compulsory deposit to lending, banks were able to change their lending interest rates effectively (Ikeo 1985, 89–93,109–116).

The BOJ used "the window guidance" as a complementary device to the ODR. Through this device, the BOJ applied moral suasion on private banks regarding the volume of their lending to customers. During the high-growth period, the interest rates were set below equilibrium, and excess demand always existed. Private banks depended on central bank credit for funding. Though the BOJ had no legal authority to set a quantitative limit on the lending of individual banks, it used its bargaining power with private banks as a leverage for moral suasion. The BOJ also used the reserve/deposit ratio as a complementary measure (Hagiwara and Masubuchi 1986, 68, 74; Ito 1992, 114; Itoh et al. 2015, 193–196).



Graph 14 Asset components of the average annual balance sheet change in three years (1955–1975). (Source: The Bank of Japan (1986b, 296–303). Note: Average annual balance sheet changes are three-year moving averages)

In 1962, the BOJ introduced the New Monetary Operation Framework, aimed at reducing private banks' dependence on central bank credit. It replaced its lending to banks by corporate-bond-buying operations to supply new money into the economy corresponding to economic growth (Graph 14). It also set limits on the amount of lending to individual banks (the Bank of Japan 1986a, 97–122; Hagiwara and Masubuchi 1986, 110).

In 1964, Japan gained the IMF Article 8 status and joined the Organization of Economic Cooperation and Development (OECD). The moves marked the full return of Japan into the international community. At the same time, Japan was expected to take on responsibilities as an industrialized, developed country. As part of these responsibilities, it was asked to liberalize its financial markets. It was required to remove all restrictions in current account transactions and to reduce restrictions in capital account transactions.

In 1965, a recession and financial instability hit the Japanese economy after the 1964 Tokyo Olympic boom. To deal with them, the government initiated fiscal stimulus, issuing government bonds. At this point, the BOJ added government bonds to its list of buying operations for providing growth money (Hagiwara and Masubuchi 1986, 112).

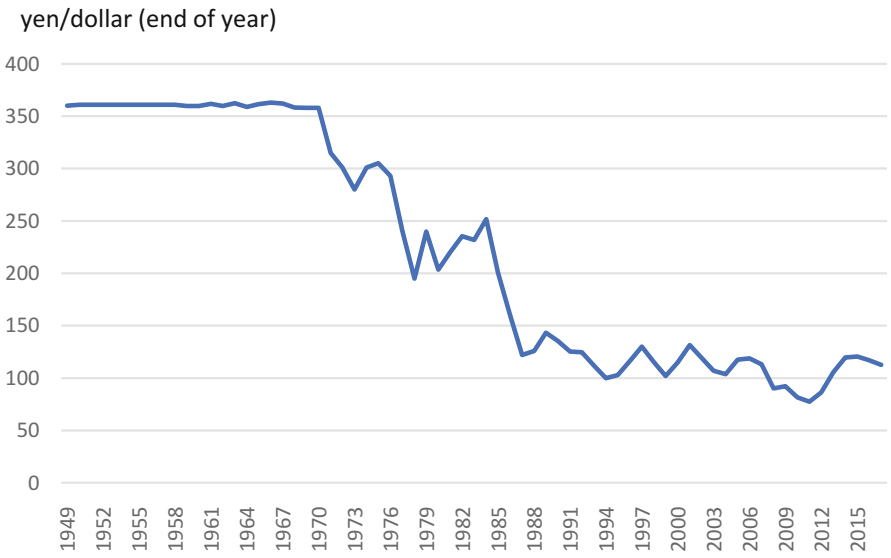
From the late 1960s through the early 1970s, Japan recorded continuous current account surpluses, reflecting improved productivity under the fixed exchange rate. The balance of payments no longer bound monetary policy and did not work as a signal for monetary tightening (Graphs 13 and 14; Hagiwara and Masubuchi 1986, 116).

Deregulation and Decelerated Growth (1971–1990)

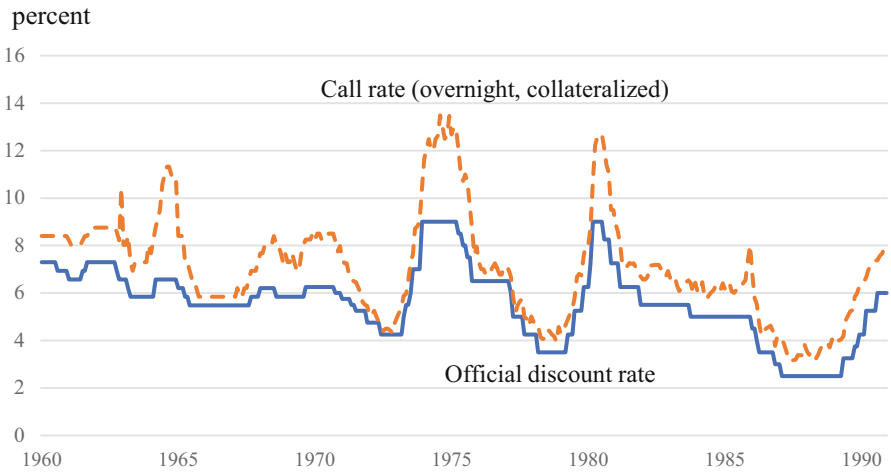
After the breakdown of the Bretton Woods System in August 1971, the Japanese economy went through a turbulent period. Following an unsuccessful attempt to maintain the previous exchange rate (360 yen to US dollar), Japan floated the yen, set a new exchange rate of 308 yen to US dollar at the end of the year, and floated the yen again in February 1973 (Graph 15). Afraid of the deflationary impact of the yen's appreciation, the BOJ maintained an accommodative monetary policy. Then, the first oil crisis hit developed economies. Japan was in the worst condition among the major industrialized countries due to the already overheating economy and a delay in monetary tightening. Consumer prices in Tokyo soared by 12 percent in 1973 and by 23 percent in 1974. The BOJ raised the ODR from 4.25 percent in April 1973 to 9.0 percent in December of the year (Graph 16). The monetary tightening triggered stagflation: for the first time since 1946, Japan recorded negative growth in output in 1974, and the double-digit inflation could not be brought under control until 1976 (the Bank of Japan 1986a, 356–436).

The first oil crisis marked the end of the high-growth period. GDP growth decelerated from an annual rate of around 10 percent to 5 percent.

The government, the central bank, companies, consumers, and labor unions learned lessons from the first oil crisis. In the short run, the monetary authorities became aware that they should have acted promptly, thereby avoiding an increase in inflation expectations in the private sector. Companies and consumers realized that they should not have built up inventories, thereby further accelerating inflation.



Graph 15 Yen/dollar exchange rates (1949–2017). (Source: The Bank of Japan)



Graph 16 BOJ official discount rates and short-term market interest rates (1960–1990). (Source: The Bank of Japan)

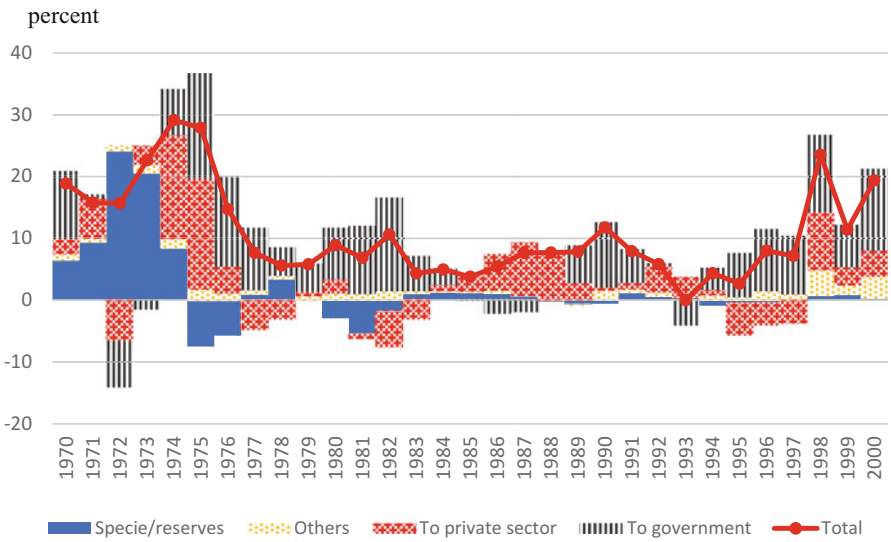
Labor unions reflected that they should not have made such high demands in their annual negotiations with management, thereby causing a wage-price inflation spiral (the Bank of Japan 1986a, 489–522; Itoh et al. 2015, 106). In the long run, the government and the private sector initiated an energy-saving campaign to reduce the consumption of oil. Companies reviewed their products and production processes to reduce energy consumption.

Starting the late 1970s, the BOJ put more emphasis on the monitoring of monetary aggregates in its policy operation. Although it never set a formal target for monetary aggregates, it started reporting them monthly in 1978, in light of the lessons learned in the early 1970s when excess liquidity drove speculation and inflation (Itoh et al. 2015, 197–199).

From the late 1970s through the early 1980s, the BOJ's balance sheet growth remained modest. Also, it accumulated "credits to the government" steadily, reflecting the central bank's stance of providing new money corresponding to economic growth through its buying operations targeting government and corporate bonds (Graph 17).

A by-product of the transformation toward an energy saving economy and society was an enhanced nonprice competitiveness of products made in Japan, which featured the advantages of "light, thin, short and small" in the new global energy environment. When the second oil crisis hit the developed economies in the late 1970s and the early 1980s, the Japanese economy showed great resilience, achieving rigorous growth in exports of automobiles and electronic devices. Things made in Japan were preferred by consumers in Japan and abroad, bringing continuous current account surpluses for Japan (Itoh et al. 2015, 112).

Deregulation in the financial markets proceeded during the late 1970s through the 1990s along with securitization and globalization of the financial/capital markets.



Graph 17 Asset components of the average annual balance sheet change in three years (1970–2000). (Source: The Bank of Japan (1986b, 300–303). Note: Average annual balance sheet changes are three-year moving averages)

Securitization of the financial/capital markets proceeded with increasing issuance of government bonds. To deal with economic stagnation and a decline in tax revenues, the government increasingly depended on public debt. Initially, a syndicate of financial institutions underwrote government bonds and held them outright or sold them to the BOJ. However, as the balance of their holdings of government debt increased steadily, they demanded the government allow them to sell government bonds and got the permission to do so starting 1977. As bond holders became diversified, the secondary market for government bonds emerged and grew (the Bank of Japan 1986a, 526–532; Itoh et al. 2015, 116–117).

Globalization of the financial/capital markets went along with the growing presence of Japan in the global economy. The government prepared for the lifting of capital controls in line with Japan's IMF Article 8 status. In 1979, the new Foreign Exchange Control Law was promulgated and became effective the following year, stipulating foreign exchange transactions be free in principle. This was a drastic change in the legal principle. In the previous flamework, all international transactions were subject to some kind of regulation. In the new flamework, international transactions were free unless otherwise designated to be regulated (the Bank of Japan 1986a, 541–545; Itoh et al. 2015, 110).

Under the Reagan Administration (1981–1989) in the United States, large current account imbalances arose, and the resulting trade dispute became the central political issue in the US-Japan relationship. Reaganomics, a combination of tight monetary policy and aggressive fiscal policy, induced conditions for a

strong US dollar and capital inflows into the United States. The first-term Reagan Administration kept a free-market approach in terms of exchange rates. It blamed Japan for creating external imbalances by shutting down its financial/capital markets from inflows of foreign investments, leading to a cheap yen (Itoh et al. 2015, 119).

To discuss the issues related to the external imbalances, the two governments jointly established the Japan-US Yen-Dollar Committee in 1983. Whether ill-motivated or not, the committee's report published in 1984 had a huge impact on the conduct of monetary policy in later periods. Most notably, the committee set a scheduled roadmap for the elimination of the regulations remaining in the Japanese financial markets. The report delivered a final blow to the regulated scheme for monetary policy, which consisted of a regulated interest rate system based on the ODR, the window guidance, and the reserve/deposit ratio. In 1988, the BOJ initiated a series of reforms in the short-term money markets and the conduct of monetary policy operations. It first revised the procedures of interest rate setting for market participants and introduced new facilities for open market operations. It ended the window guidance in 1993 and started announcing a target for the call rate, the representative short-term interest rate, in 1997 (Miyanoya 2000; Itoh et al. 2015, 121–123).

The second-term Reagan Administration had a different strategy for dealing with the external imbalances. New US Secretary of the Treasury James Baker proposed multilateral adjustments in exchange rates between the United States and its main trading partners. As per his idea, the strengthening of the yen and other currencies against the US dollar, induced by joint interventions in the foreign exchange market, would have a switching effect that would help curb massive exports to the United States from countries such as Japan and Germany. In 1985, the finance ministers and central bank governors of the G5 countries (the United States, Japan, West Germany, United Kingdom and France) gathered at the Plaza Hotel in New York and agreed on joint intervention as part of a concerted effort to correct misalignments among major currencies (the Plaza Accord, Graph 15). In 1987, finance ministers and central bank governors gathered again at the Louvre in Paris and agreed to coordinate macroeconomic policies, including fiscal, monetary, and exchange rate policies (the Louvre Accord; Itoh et al. 2015, 124–137).

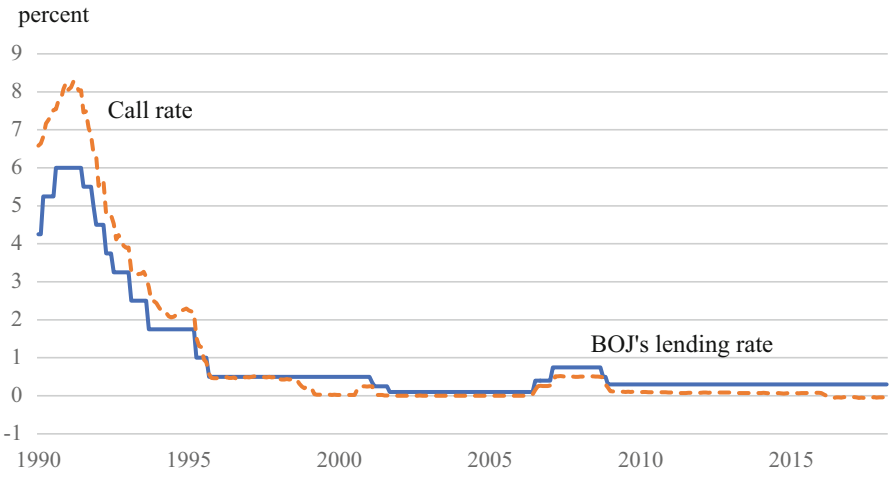
As part of the concerted efforts to reduce external imbalances, the BOJ employed easy monetary policy to stimulate the economy, thereby reducing the current account surplus. While the inflation in general prices was curbed thanks to the stronger yen and stable oil prices, financial bubbles formed as a result of easy money and bullish expectations regarding Japanese economic prospects. Asset prices, such as stock and real estate prices, soared to historical highs. When the bubble collapsed, the BOJ learned its lesson: put your own home in order rather than making an effort to achieve an international goal, and keep a close watch on financial indicators as signs of potential risk in the long run even if headline economic indicators show steady growth with low and stable inflation (Okina et al. 2001, 444–445; Shizume 2018, 353–354).

New Challenges (1991–2017)

In the 1990s, financial bubbles collapsed, asset prices plunged, and Japan experienced financial distress and economic stagnation. The business sector suffered from excess capacity built during the bubble period. The financial sector suffered from nonperforming loans and constraints in new lending (Nakaso 2001, 2–16; Shizume 2018, 355). At the same time, increasing competition with the other emerging Asian economies eroded the made-in-Japan brand and dulled Japan's edge in productivity. Growth fell to an annual rate of around zero percent during the 1990s and beyond.

To tackle economic stagnation and financial distress, the government expanded its fiscal expenditure. The administrative branch of the government tried capital injection into the financial sector but was trapped by public resentment and political backlashes. The delay in financial restructuring paved the way for the financial crisis of 1997–1998 (the Heisei Financial Crisis; Nakaso 2001, 6, 18–19). The BOJ lowered interest rates to stimulate demand and support the restoration of financial stability. When the BOJ reduced the ODR to 0.5 percent in 1995, it virtually hit the zero lower bound (ZLB) (Graph 18; Ueda 2005, 333).

The Bank of Japan Act was amended in 1997 and became effective in 1998. The new act stipulated autonomy and required transparency in the BOJ's policy formation. The roles of the BOJ were defined as: (1) issuing banknotes, carrying out currency and monetary control and (2) ensuring smooth settlement of funds among banks and other institutions, thereby contributing to the stability of the financial system. The currency and monetary controls of the BOJ were to aim to achieve price stability, thereby contributing to the sound development of the national economy.



Graph 18 BOJ lending rates and short-term market interest rates (1990–2017). (Source: The Bank of Japan)

Table 2 Monetary policy innovations during and after the 1990s

Start date	End date	Name/description
Feb. 1999	Aug. 2000	Zero Interest Rate Policy (ZIRP)
Mar. 2001	Mar. 2006	Quantitative Easing Policy (QE)
Oct. 2010	Apr. 2013	Comprehensive Monetary Easing (CME)
Feb. 2012	Apr. 2013	Enhancement of Monetary Easing (Introduction of "the Price Stability Goal," etc.)
Jan. 2013	Apr. 2013	Joint Statement of the Government and the Bank of Japan; "Price Stability Target" and the "Open-Ended Asset Purchasing Method"
Apr. 2013	–	Quantitative and Qualitative Monetary Easing (QQE)
Oct. 2014	–	Expansion of the Quantitative and Qualitative Monetary Easing
Jan. 2016	–	Quantitative and Qualitative Monetary Easing with a Negative Interest Rate
Sept. 2016	–	Quantitative and Qualitative Monetary Easing with Yield Curve Control

Sources: Ueda (2005); the Bank of Japan

The BOJ's autonomy in currency and monetary control was to be respected, while the BOJ was to endeavor to clarify to the citizens the content of its decisions and decision-making process regarding currency and monetary control. The Monetary Policy Meeting was to comprise the Governor, two Vice Governors, and other board members (up to six in number), who would be fully responsible of monetary policy (Shizume 2018, 354).

Because the new Bank of Japan Act was promulgated in the midst of financial distress and economic stagnation, the BOJ with its new Policy Board faced challenges from the onset. It had to deal with the financial crisis. It had to innovate new policy tools to conduct monetary policy under the ZLB. In response, it introduced a series of policy innovations called "unconventional monetary policy" measures, pioneering a new paradigm of monetary policy ahead of other central banks (Table 2; Ueda 2005, 332).

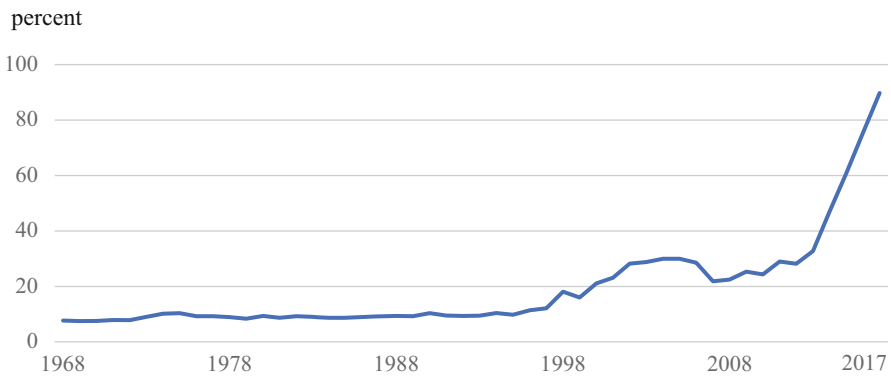
In April 1999, the BOJ introduced the Zero Interest Rate Policy (ZIRP). As part of the ZIRP scheme, the BOJ not only reduced its interest rate target to zero but also committed itself to maintaining the target until a preannounced condition was fulfilled. Governor Masaru Hayami declared that the bank would continue the ZIRP "until deflationary concerns were dispelled." This was the first time anywhere in the world that a central bank had used the length of time during which a zero rate is maintained as an instrument of monetary policy. The BOJ, facing the ZLB, ended up at the forefront of monetary policy innovation in trying to affect expectations of future monetary policy actions and, thus, short-term interest rates. Other central banks followed suit after the collapse of Lehman Brothers in 2008, calling such a policy instrument "forward guidance" (Ueda 2005, 334).

In August 2000, the BOJ lifted the ZIRP, as the economy was recovering with signs of overcoming deflation. In late 2000, however, the economy began to

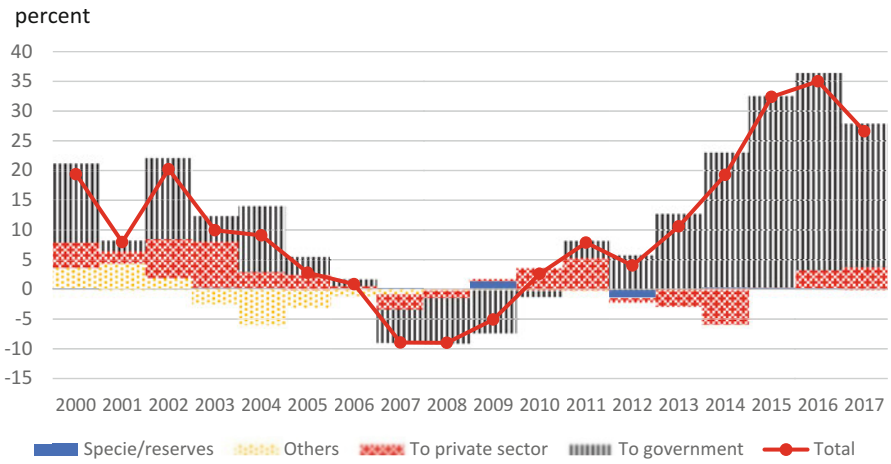
deteriorate again in the face of a global decline in the demand for high-tech goods. The BOJ adopted the Quantitative Easing (QE) policy in March 2001 (Ueda 2005, 334).

Ueda (2005) divides the unconventional monetary policy measures under the scheme of the QE (2001–2006) into three elements: (1) shaping or managing interest-rate expectations, (2) altering the composition of the central bank's balance sheet to change the relative supply of securities in the market, and (3) expanding the size of the central bank's balance sheet beyond the level required to set the short-term policy rate at zero (Ueda 2005, 335–336). Under the schemes of Comprehensive Monetary Easing (CME: 2010–2013) and Quantitative and Qualitative Monetary Easing (QQE: 2013-present), the BOJ has purchased or accepted risky assets such as CPs, exchange traded funds (ETFs), and real estate investment funds (REITs) as collateral. This type of policy measure is also called credit easing because it aims to reduce risk premiums on private assets. Under the scheme of QQE, the BOJ purchases government debts in longer maturities, intending to reduce the term premiums of these bonds.

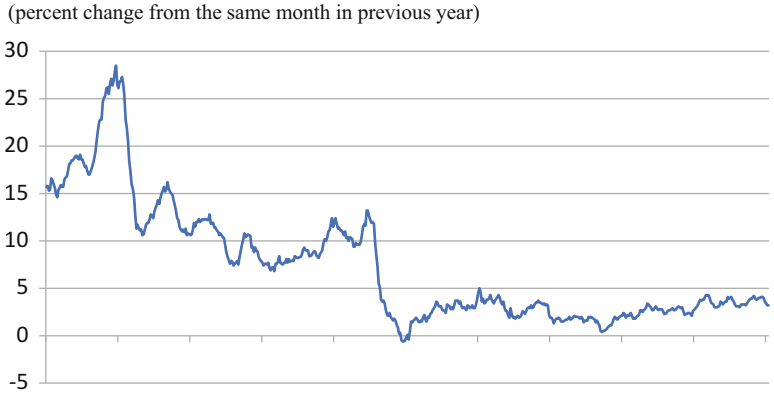
QQE starting April 2013 has been the most unconventional policy in terms of size. Declaring a clear annual inflation target of two percent, the BOJ has been pursuing aggressive monetary easing. As a result, the QQE regime is literally unprecedented in terms of the central bank's balance-sheet size compared with the national economy. The BOJ's balance sheet reached 90 percent of GDP in 2017 (Graph 19). Most of the recent growth in the balance sheet has been due to increase in credit to the government by means of the purchase of government bonds (Graph 20). Despite the massive increase in the central bank's balance sheet over the past five years, however, the inflation target has not been achieved while monetary aggregates have stayed stable since the 1990s even with recent policy innovations (Graph 21; Shizume 2018, 355).



Graph 19 BOJ balance sheet/GDP ratio (1968–2017). (Sources: The Bank of Japan 1986b, 272–304)



Graph 20 Asset components of the average annual balance sheet change in three years (2000–2017). (Source: The Bank of Japan. Note: Average annual balance sheet changes are three-year moving averages)



Graph 21 Money stock (M2). (Source: The Bank of Japan)

Conclusion

Historically, the goals of monetary policy have been defined based on the goals of the nation as a whole. They changed over time in accordance with changes in the goals of the nation. On the long path to an industrialized, developed country from the late nineteenth century through the early 1970s, the ultimate goal of monetary policy was to provide money for development. In the pre-WWII period, the central bank faced a conflict between two national goals – a wealthy nation and a strong military. Especially, during the big wars with Japan's main trading partners in the

1930s–1940s, monetary policy was overwhelmed by war finance and resulted in rampant inflation. In the post-WWII high-growth period, monetary policy was focused on growth without interference from war finance. Once Japan had achieved industrialization and become one of the leading economic powers, the primary goal shifted from development to stability. Meanwhile, the country faced new challenges owing to the maturing of its economy and society.

Instruments of monetary policy have also changed over time. They were constrained by the availability of instruments in the financial markets while the central bank pursued policy innovations. In the early days, the central bank faced lack of standardized financial instruments and began accepting equities as collateral for lending. During wars, the central bank supported war financing. When the government issued a large volume of bonds in the 1930s–1940s, the central bank underwrote the bonds and sold them to the private sector, calling those operations "open market operations." During the high-growth period in post-WWII Japan, the main policy instrument was interest-rate regulation accompanied by window guidance and the reserve/deposit ratio. When the Japanese economy, having achieved complete the high growth, became a major economic power, instruments of monetary policy shifted from regulation-based ones to market-based ones. In the 1990s, Japanese economy hit the ZLB, and the central bank has since been trying to deal with it by inventing unconventional monetary policy measures.

Cross-References

- ▶ Currency Blocs: the Yen
- ▶ Deflations in History
- ▶ International Monetary Regimes: The Bretton Woods system
- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ Money in Wars
- ▶ The Historical Evolution of Central Banking
- ▶ The Monetary System of Japan in the Tokugawa Period

References

Asahi Shinbun Sha (1930) Nihon Keizai Toukei Souran (Comprehensive volume of Japanese economic statistics). Asahi Shinbun Sha, Osaka (in Japanese)

Fukuda S (1995) The founding of the bank of Japan and the changed behavior of interest rates and inflation rates. J Jpn Int Econ 9:56–74

Hagiwara K, Masubuchi M (1986) Zusetsu Nippon Ginkou (An illustration of the Bank of Japan). Zaikei Shouhou Sha, Tokyo (in Japanese)

Ikeo K (1985) Nihon no Kin'yuu Shijou to Soshiki (Financial markets and institutions in Japan). Toyo Keizai Shinposha, Tokyo (in Japanese)

Ishii K (ed) (2001) Nippon Ginko Kin'yuu Seisaku Shi (Historical studies on the Bank of Japan). University of Tokyo Press, Tokyo (in Japanese)
Ito T (1992) The Japanese Economy. MIT Press, Cambridge, MA
Itoh M, Koike R, Shizume M (2015) Bank of Japan's monetary policy in the 1980s: A view perceived from archived and other materials. Mon Econ Stud 33:97–199
Matsukata M (1899) Report on the adoption of the gold standard in Japan. The Government Press, Tokyo
Metzler M (2006) Lever of empire: the international gold standard and the crisis of liberalism in Prewar Japan. University of California Press, Berkeley
Mitchener KJ, Shizume M, Weidenmier MD (2010) Why did countries adopt the gold standard? Lessons from Japan. J Econ Hist 70(1):27–56
Miyanoya A (2000) A guide to bank of Japan's market operations. Financial Markets Department Working Paper, 00-E-3, the Bank of Japan
Nakamura T (1994) Lectures on modern Japanese economic history, 1926–1994. LTCB International Library Foundation, Tokyo
Nakaso H (2001) The financial crisis in Japan during the 1990s: how the Bank of Japan responded and the lessons learnt. BIS Paper No.6, October. Basel: Bank for International Settlements
Okina, K, Shirakawa M, Shiratsuka S (2001) The asset bubble and monetary policy: Japan's experience in the late 1980s and the lessons. Monetary and economic studies, Special Edition, February, pp 395–450
Shizume, M (2012) The Japanese economy during the interwar period: instability in the financial system and the impact of the world depression. Bank of Japan Review, 2009-E-2. Reprinted in Ögren, Anders and Lars Fredrik Øksendal, eds, The gold standard peripheries: monetary policy, adjustment and flexibility in a global setting, Palgrave Macmillan, New York, pp 211–228
Shizume M (2017) Financial crises and the central bank: lessons from Japan during the 1920s. In: Suto I, Rockoff H (eds) Coping with financial panics: some lessons from financial history, Studies in economic history. Springer, Singapore
Shizume M (2018) A history of the Bank of Japan, 1882–2016. In: Jacobson T, Waldenström D, Edvinsson R (eds) Sveriges Riksbank 350 years: a central bank in a world of central banks. Cambridge University Press, Cambridge, pp 328–360
Smethurst RJ (2007) From foot soldier to finance minister: Takahashi Korekiyo, Japan's Keynes, Harvard East Asian monographs, vol 292. Harvard University Press, Cambridge, MA
The Bank of Japan (1982) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 1, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1983a) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 2, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1983b) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 3, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1984) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 4, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1985) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 5, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1986a) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years), vol 6, The Bank of Japan, Tokyo (in Japanese)
The Bank of Japan (1986b) Nippon Ginko Hyakunen-Shi (Bank of Japan: The First Hundred Years). Related materials and statistics, The Bank of Japan, Tokyo (in Japanese)
Ueda K (2005) The Bank of Japan's struggle with the zero lower bound on nominal interest rates: exercises in expectations management. Int Finance 8(2):329–350

The Historical Evolution of Monetary Policy in Latin America

37

Esteban Pérez Caldentey and Matías Vernengo

Contents

Abstract

Recent monetary policy in Latin America has evolved in five different phases. The first one revolves around the monetary doctoring missions, mainly but not exclusively, from the United States that established the first central banks in the region in the 1920s and 1930s adhering to a gold standard regime. The second phase comprises the first attempts to adopt a proactive countercyclical monetary stance and the granting of ample discretionary powers to the central bank to protect economies from the effects of the international business cycle. The third phase consists in the generalized adoption of developmental and inward

E. Pérez Caldentey (✉)
Economic Commission for Latin America and the Caribbean ECLAC, Santiago, Chile
e-mail: esteban.perez@un.org

M. Vernengo
Bucknell University, Lewisburg, PA, USA
e-mail: mv012@bucknell.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_44

industrialization goals by central banks. The fourth phase is characterized by central bank's abandonment of developmental objectives placing the focus of their actions on price stability. The last phase in the evolution of monetary policy centers on inflation targeting. Currently, the most important and pressing challenges of monetary policy relate to its impact and pass-through on the real economy and to the relationships between monetary policy, fiscal policy, the financial system, and financial stability.

Keywords

Gold standard · Countercyclical policy · Exchange controls · Monetary approach to the balance of payments · Inflation targeting

Introduction

Monetary policy in Latin America has evolved in five different phases. The first one revolves around the adoption of the gold standard during the first globalization period at the end of nineteenth century. During the colonial period, the monetary system was tied to the metropolis, and the monetary system was underdeveloped. Metallic currency was scarce, and no colonial banks existed. The first banking experiment in Latin America was associated to the Bourbonic reforms of Carlos III, the so-called Banco de Avío y Minas, created to fund mining activities. The first bank in Brazil was the Banco do Brasil, founded in 1808 with the arrival of the Portuguese royal family to Rio de Janeiro. In Argentina the Banco de la Provincia de Buenos Aires, founded in 1822, is often cited as the first bank. In Mexico the first bank was the Banco de Avío Industrial created by Lucas Alamán in 1830. Banking activities remained limited in the region during the nineteenth century (Marichal and Gambi 2017).

Toward its end, the period is marked by the monetary doctoring missions, mainly but not exclusively from the United States and the United Kingdom, that established the first central banks in the region in the 1920s and 1930s. Monetary policy followed the pro-cyclical rules of the gold standard. Monetary stability, viewed as essential to attract foreign investment, also involved trade measures and fiscal reforms to balance the budget. The gains from the reforms were for the most part temporary and countries experienced increased macroeconomic disequilibria and rising debt. The gold standard experience in Latin American came to an end with the Great Depression, which set the stage for the second phase in monetary policy.

This phase comprises the first attempts to adopt a proactive and, more precisely, countercyclical monetary stance and the granting of ample discretionary powers to the central bank to protect economies from the effects of the international business cycle. The countercyclical aspect of monetary policy marks the beginning of modern central banking in Latin America. During this phase, some central banks also pursued developmental objectives. The instruments used to achieve these objectives included rediscounting, exchange controls, subsidies, and the allocation of credit.

The third phase consists in the generalized adoption of developmental and inward industrialization goals by central banks which, in some instances, prevailed over

short-run price stability objectives. The two main instruments used included capital controls and reserve requirements. Capital controls included differentiated and multiple exchange rates for foreign current and capital exchange transactions, between public and private sector, and also between residents and nonresidents. Reserve requirements were used to direct credit to specific economic activities and finance government operations.

During this phase Latin America registered in the 1960s and 1970s the highest average growth on record. At the same time, the rise in inflation in most economies became a growing concern for the monetary authorities. The third phase ends with the 1980s Debt Crisis, which led to a radical shift in economic and monetary policy under the implementation of Washington Consensus policies consisting in liberalization, stabilization, and privatization.

The fourth phase is characterized by central bank's abandonment of developmental objectives placing the focus of their actions on price stability. Stabilization policies were guided by the monetary approach to the balance of payments (MABP) which views both inflation and the balance of payments position as monetary phenomena. While the MABP recommends focusing on credit rather than money growth relative to output to control prices, countries still adhered to monetary targets and exchange rate management. Countries also introduced changes to central bank legislation addressing the need for political and operational independence and greater transparency and accountability. During this phase, as part of a worldwide trend since the middle of the 1990s, most Latin American countries saw a significant reduction in their inflation rates bringing these to one-digit levels. Monetary policy had to confront financial and banking crises, which drove home the message that price stability cannot be equated with financial stability. As well, the weight placed (at least in some economies) on exchange rates management created a potential conflict with the objective of price stability.

The last phase in the evolution of monetary policy centers on inflation targeting, which since the 2000s has become the most prevalent monetary framework in Latin America. Monetary policy operates through the public announcement of numerical targets for the inflation rate to guarantee transparency and accountability and the countercyclical management of the short-term policy interest rate according to a monetary rule, often associated to some variation of the Taylor rule. The recommendation is to adjust the monetary policy rate in a flexible manner to avoid output and employment costs. In an open economy, inflation targeting is compatible with a freely floating exchange rate that acts as an absorber and dampener of external shocks. The available evidence for Latin America shows that inflation targeting countries do not necessarily pursue a countercyclical monetary policy and that these often intervene in the foreign exchange market in order to control inflation.

Currently Latin America economies have been able, with a few exceptions, to maintain single-digit inflation rates, so that high inflation and its disruptive effects on the real economy are not an impending threat. The most important and pressing challenges for monetary policy relate to its impact and pass-through on the real economy and to the relationships between monetary policy, the financial system, and financial stability.

Antecedents to Central Banking and Monetary Policy

The first initiatives to create central banks in Latin America date back to the nineteenth century as governments, either at the national, regional, or provincial level, took on the power of issuing domestic currency and credit (an activity confined until then to the private sector) for developmental purposes and, also, to finance government operations. National banks were established in several Latin American countries including in Argentina (1826, 1836, and 1891), Bolivia (1911), Colombia (1880 and 1905–1909), Ecuador (1890), and Peru (1821) (Posso Ordoñez 2016; Marichal 2008).

Argentina established the Banco Nacional (1826) to provide credit to the government and reduce its dependency on external inflows. This was followed by the creation of the Casa de la Moneda (1836) to fund public sector deficits and the Banco de la Nación (1891) to stabilize the country's financial system after the start of the Baring Crisis and promote the agro-export development model. Bolivia created the Banco de la Nación Boliviana (1911). Brazil created the Banco do Brasil in 1808, as noted before, which continued to function for a while after independence in 1822 but was eventually liquidated. The new Banco do Brasil was created in 1853, with the merging of two banks, and given the monopoly on emission. Colombia created the National Bank (1880) to foster the development of the country and authorized to issue currency, receive deposits, and carry out discount operations. Later on, the Central Bank of Colombia operated from 1905 to 1909. Ecuador planned unsuccessfully for the creation of its central bank in 1890. Peru created the Banco Auxiliar de Emision (1821) to issue currency and grant credit. Chile opted for having the Treasury take the role of the central bank.

These initiatives were generally short-lived as these resulted in currency depreciation and inflation (which were in some cases brought about to benefit debtors' and exporters' interests) and, in general, monetary disorder and inconvertibility. These initiatives were replaced by exchange offices. One of the most illustrative and well-documented examples is the Argentine Currency Board established in 1899, following a period of monetary instability reflected in the sharp rise and fall of the gold premium due in part to the Baring Crisis (1890–1897).

The exchange office functioned like a currency board. It maintained fixed convertibility of the peso to gold, and through the establishment of a conversion fund, guaranteed that any addition to the money supply should have a 100% gold backing. Similarly, any withdrawal of gold would translate into the withdrawal of an equal amount of paper pesos. Foreign exchange intervention was also allowed to maintain the peso-gold parity. The Currency Board functioned smoothly between its entry into force in 1900 and 1912–1913, helped by fair weather conditions and favorable terms of trade for Argentina.

Contrarily in 1914, the Currency Board faced poor climatic conditions with devastating effects on agricultural production and exports, the balance of trade position, and a depressing impact on land values and capital flows. These effects were reinforced by the needs to finance the European war efforts, which dried up

Argentina's external sources of finance. The resulting contraction of liquidity led to pro-cyclical response with negative effect on income and employment and asset prices (land values). It also aggravated the financial situation of the farmers affected by the weather shock, thus impinging on the recuperation of exports and land values. This provided a further blow to the confidence of foreign investors.

The Creation of Central Banks

The failure of national banks and exchange offices to provide monetary and financial stability, along with a growing international consensus on the need of countries to have a central monetary authority and a gold standard regime, led to the establishment of central banks in Latin America. The first central banks were established in the 1920s by US foreign missions to Colombia (1923), Chile (1925), Ecuador (1927), Bolivia (1928), and Peru (1931). The missions were headed by Edwin Kemmerer (1875–1945), a Princeton University professor and one of the leading developers of the quantity theory of money. Kemmerer was also a financial advisor to the Commission of Administrative Financial Reorganizing as well as to the governments of Mexico (1917) and Guatemala (1919, 1924).

The significance of these missions is explained by the US outward economic and commercial expansion and also by its growing importance as an international creditor. Between 1913 and 1929, US investments in South America increased from US$ 72 to more than US$ 900 million which was above the level reached by British investments. As explained by Kemmerer (1927, p. 4): "It is the desire on the part of the foreign governments through setting their financial houses in order, and through making a favorable impression upon American bankers and investors, to facilitate the borrowing of money by the government in the American market and to encourage the flow of American capital to their shores for private enterprises." Kemmerer also believed that foreign investment directed trade flows (Kemmerer 1916).

The Kemmerer missions sought to maintain monetary and financial stability through the adherence to the gold standard. Kemmerer was a staunch supporter of the gold standard. Earlier on in his career, he had proposed a Pan-American monetary unit based on the dollar, which was tied to gold, and throughout his career, even after the Great Depression, he remained committed to gold convertibility in part, to protect the interests of the creditors. He identified several benefits of the gold standard including its simplicity, strong confidence by the public, an automatic mechanism needing little management, the homogeneity in the monetary standard (i.e., gold), and its stability. Kemmerer recognized some of its weaknesses such as its rigidity or cost but did not think these undermined or offset in any way its benefits (Kemmerer 1944).

The central banks' mandates included the exclusive right of note issue, the centralization of the gold reserve, maintaining the functioning and stability in the payment systems, and, also, providing limited finance to the government (see Table 1). Kemmerer also thought that a gold standard regime required an improved

Table 1 Year of creation and functions of selected central banks in Latin America

Argentina	Ecuador	Chile	El Salvador	Colombia	Peru
Law 12.155, (1935):	Organic Law of the Central Bank of Ecuador (2927):	Law 486 (1925):	Law that Creates the Reserve Central Bank of El Salvador (1934):	Law 25 (1923):	Law 4500 (1922):
Issue the domestic currency	Issue the national currency	Issue the national currency	Issue the domestic currency	Issue the national currency	Monopoly of currency issue
Regulate the amount of money and credit in line with the needs of the economy	Act as lender of last resort (LOLR) to the banking system	Conduct rediscount and discount operations with banks and the general public	Control the volume of credit and the money supply	Conduct rediscount and discount operations with banks and the general public	Conduct rediscount and discount operations with banks and the general public
Accumulate sufficient international reserves to moderate the adverse effects of exports and foreign investments and preserve the value of the currency	Conduct rediscount and discount operations with banks and the general	Provide financial assistance to the public sector on a limited basis	Preserve the external value of the currency	Provide financial assistance to the public sector on a limited basis	Discount and rediscount commercial paper, treasury bonds, and other financial instruments
Preserve appropriate conditions of liquidity and credit and apply the legislation on the banking system	Provide financial assistance to the public sector on a limited basis	Work as a fiscal agent	Act as a fiscal agent and receive deposits from the government	Work as a fiscal agent	Define discount rates
Act as fiscal agent and advisor for the management of public debt	Work as a fiscal agent	Receive deposits from banks, the public sector, and the general public		Receive deposits from banks, the public sector, and the general public	Provide clearing for payments

(*continued*)

Table 1 (continued)

Argentina	Ecuador	Chile	El Salvador	Colombia	Peru
	Receive deposits from banks, the public sector, and the general public	Provide clearing for payments		Provide clearing for payments	Provide financing to the government on a limited basis
	Define the rediscount rate	Define the rediscount rate		Define the rediscount rate	
	Provide clearing for payments				

Source: Jácome 2015; Flandreau 2003

control and performance of public finances. To this end, the missions also introduced fiscal measures alongside currency and banking reforms. These included, among others, initiatives and laws to strengthen the supervision of the government budget and restrain the scope for discretionary spending, increase the efficiency of the tax and tariff system and the introduction of income and property taxes, and oversee the railroad administration. Some of these measures were aimed also at protecting American economic interests. In the case of Colombia, Kemmerer proposed, in 1930, a tax on the export of bananas benefiting the United Fruit Company that held the monopoly of Colombian banana trade (Seidel 1972; Drake 1989).

The main result of the Kemmerer mission was the increase in foreign investment and lending. Available data in the case of Colombia show that, while in 1914 foreign investment had reached US$ 60 million, it increased to US$ 236 million for the period 1926 to mid-1928 (Daalgard 1980). In the case of Chile, between 1927 and 1928 foreign loans increased from US$ 228 to 513 million (Glaser 2003, p. 173).

At the same time the missions did not necessarily lead to improved economic performance or macroeconomic management. Evidence available in the case of Chile shows that after the implementation of the gold standard in 1925 until the start of the Great Depression in 1930, GDP, mineral production, and industrial production registered positive growth only in 1928 and 1929. Equally, public debt expanded from 2,767 to 4,105 million pesos and the country also experienced a current account deficit (350 million pesos at the end of 1929). The country eventually defaulted on its debt in 1931 and, as other countries with Kemmerer-style central banks, went off the gold standard.

The Beginnings of Countercyclical Monetary Policy

The Great Depression and its negative effects led central bank to shift course toward a more proactive, and in some cases explicitly countercyclical, monetary policy stance. This change in policy orientation is reflected in the introduction of new legislation expanding the scope of action and discretionary power of central banks. This marked the beginning of modern central banking in Latin America.

The reform of the laws of the Mexican and Chilean central banks in the early 1930s provide partial illustrations of this trend. In April 1932, the Bank of Mexico began using rediscounting as a policy to manage liquidity and promote economic development, becoming de facto a Central Rediscount Bank. In 1931–1932 the Bank of Chile adopted exchange controls and multiple exchange rates, introduced the use of the rediscount, and financed government development initiatives.

The Central Reserve Bank of El Salvador (1934) was created as an orthodox institution in similarity with the Kemmerer banks but eventually was allowed to undertake open-market operations. Another case in point is the Central Bank of Venezuela (1939) that was given discretionary powers to grant credit to the banking system and the public and allowed to use rediscounting to finance temporary budget deficits (Triffin 1944).

Apart from these examples, the story of the creation of the Central Bank of Argentina (BCRA), its distancing from the orthodox position advocated by the British monetary doctor Otto Niemeyer (1883–1971), and its evolution toward a definite countercyclical stance provide the most representative cases of this change in monetary policy. The BCRA was created in May 1935, and Raúl Prebisch, who was also responsible for drafting the bank's project proposal, was appointed its first general manager. The BCRA took on the functions of the Currency Board, the Treasury, and those of the Office for the Control of Exchanges, with the exception of the dealings with exporters and importers.

The BCRA was conceived as an institution independent of the government to avoid the creation of excess money supply due to fiscal deficits and to permit a more rational distribution of monetary functions and more efficient management of its policy of reserve accumulation, whose main objective was monetary stability, along conventional lines. The evolution of the BCRA's monetary policy is marked by three stages, namely, (i) 1935 to the first half of 1937, (ii) second half of 1937–1941, and (iii) 1941–1942 (Prebisch 1991; Pérez Caldentey and Vernengo 2018).

The first stage (1935 to the first half of 1937) consists in central bank interventions (including open-market operations, the creation of a foreign exchange fund and its use to repay external debt, and moral suasion) to limit excessive credit growth and the overheating of the economy in the boom phase of cycle. This period was characterized by a favorable international context which translated into increased exports and in the stock of international reserves increasing the overall liquidity in the economy.

The second stage (second half of 1937–1941) was triggered by the decline in economic activity accompanied by a worsening of the balance of payments position. This stage consists in the realization that central banks must intervene not only in the

boom phase but also during the recession through the expansion of credit. To this end, the BCRA resorted to exchange controls to dampen import growth and protect the balance of payments position.

In the third stage, the BCRA prioritized the full employment of resources while insulating the economy from external shocks. It was devised in 1941 as a result of Argentina's loss of its export markets during World War II, which translated into a severe balance of payments restriction and the fear of general economic prostration. Exchange controls were perfected and became a key monetary policy instrument.

This third stage provided the setting to devise a nationally autonomous monetary policy with the aim of providing a more stable level of economic activity, reduce the vulnerability to external shocks, and ensure the most favorable conditions to fulfill the growth potential of the economy. This monetary policy consisted, on the one hand, in ensuring the provision of sufficient purchasing power through the extension of domestic credit to offset the impact of a fall in exports or decline in foreign financial flows on the economy. On the other hand, it contemplated the application of exchange controls to ensure that the expansion of credit would not lead to a disequilibrium in the balance of payments. Exchange controls also mitigated the potential effects of exchange rate depreciation which included a rise in the cost of living, the extraordinary benefits accruing to some sectors of the economy, and the protection without discrimination of the national industry.

The exchange control proposal sought to differentiate between the categories of imports that should adapt to the business cycle from those that should be isolated from business cycle fluctuations. To this end, the government would establish a hierarchical order of the different import categories and, according to the circumstances, would prioritize the imports most needed to fulfill essential needs and carry out production. The exchange control scheme sought to control imports by varying the exchange rate rather than by quantitative controls seen as too complicated to implement and economically inefficient. The exchange control scheme would be implemented through a process of auctions.

Another Round of Monetary Doctoring in Latin America

The adoption of a countercyclical monetary policy became a generalized practice in the 1940s as the State took on a leading role in intervening and guiding the direction and pace of economic development until the Debt Crisis of the 1980s. This development strategy was underpinned by the Good Neighbor policy adopted by the United States toward Latin America, based upon nonintervention and noninterference with its domestic affairs and which evolved into one of active financial and economic cooperation (Helleiner 1999).

The Federal Reserve Board (FED) provided technical assistance on monetary and financial issues to several Latin American countries including Cuba (1942) and under the lead of Robert Triffin (1911–1993) to the Dominican Republic (1946), Ecuador and Guatemala (1945–1946), Honduras (1943), and Paraguay (1943–1944). Triffin broke with the monetary doctoring of the 1920s whose

principles, he thought, had been "artificially transplanted and implemented to an entirely different environment from that of the great financial centers where it developed historically," thus resulting in a failure (Triffin 1947).

Triffin argued that the problem of monetary management in Latin America was determined by a dominating influence of trade and financial flows on the domestic economy, the erratic nature of fluctuations in the balance of payments, the conflicts between fiscal and monetary policy derived from the dependence of the government on central bank credit, and the absence of well-developed financial markets which made useless and inefficient the use of the traditional monetary policy instruments. The imposition of rigid monetary standards deprived these countries of their monetary management capacity. Thus, they were forced to adapt pro-cyclically to external shocks which aggravated their intensity and effects on monetary and real variables (Triffin 1944, 1947).

He considered his advisory expert missions to Latin America to be truly revolutionary as these placed the central bank and the financial system at the service of economic and social development, rather than trying merely to imitate the Bank of England or the US Federal Reserve model (Wallich and Triffin 1953; Triffin 1981). According to the specific circumstances at hand, his monetary proposals included the use of reserve accumulation, capital regulation measures, the establishment of foreign exchange controls, and an active policy of rediscount and advances as means to confront external shocks and dampen the fluctuations of the business cycle. In some cases, Triffin recommended that central banks should allocate medium- and long-term lending to productive sectors.

The most emblematic case of this new approach to monetary doctoring is that of Paraguay. The reform proposal included adapting the functions of the Central Bank to the needs of the domestic economy, breaking with the traditional rules of linking the monetary issue to international reserves, and ensuring an adequate distribution of credit according to the best interests of the country. Accordingly, the reform gave ample and indeed unprecedented powers to the Central Bank to carry any type of operation in monetary, exchange rate, and credit. This included the provision of medium-term and long-term credit to agriculture and industry and also long-term mortgage loans.

To isolate the domestic economy from external shocks, the reform proposed the use of several instruments, including the buildup of a stock of international reserves, foreign exchange sterilization operations, capital controls, and also exchange controls. The exchange control legislation was drafted by Raúl Prebisch, who was invited by Triffin to join in the spirit of the legislation prepared for Argentina.

Another well-documented case of banking and monetary policy reform is that of the Dominican Republic that gave ample powers to the central bank including the adequate distribution of credit among the different economic activities. The legislation also contemplated the use of capital controls and exchange controls through a system of licensing of credit and investment operations and the transfer of funds abroad.

Monetary Policy During the Developmental Period

In the 1940s other central banks in Latin America – including those of Chile, Colombia, Peru (originally created by the Kemmerer missions), Guatemala, and Mexico – also put economic development (along with price stability) as a central policy objective. Two of the main monetary policy instruments to fulfill this goal were capital controls and reserve requirement. These tools were also used to weather external shocks. In addition, in some cases (in particular in Colombia, Mexico, and Peru), the government through the Minister of Finance was actively involved in determining monetary policy guidelines.

Capital controls, which took the form of differentiated and multiple exchange rates for foreign current and capital exchange transactions, as well as between public and private sector or between residents and nonresidents, allowed greater policy autonomy to limit shocks in the foreign exchange market, favor investment in some sectors, and promote the domestic ownership of specific activities. Capital controls were also used to promote a process of inward industrialization.

Capital controls were combined with the use of unremunerated reserve requirements (in some cases established or modified by the government) to direct credit (through the purchase of government securities) to specific economic activities and, also, to finance the government deficit. Data show that in Argentina loans to the government represented 5% of total domestic credit in 1950; this had increased to 70% by 1970. Similarly, in the case of Peru loans to the government fluctuated between 50% and 60% in the same period.

Central Bank legislation established a minimum and maximum limit for the rate of reserve requirements that applied to different bank liabilities, left open the possibility of introducing marginal reserve requirements rates, and specified the type of assets (not only cash and deposits but also securities) that qualified as reserves. Quantitative controls were used to avoid excessive credit growth to control inflationary pressures. Qualitative controls were used to allocate credit to productive activities through the rediscount of commercial bank paper at the central bank. Credit quotas for commercial banks were established on the basis of their capital, and differential interest rates were granted depending on the loan maturity and the type of economic activity (Jácome 2015).

During this period the Central Banks of Costa Rica (1950), Honduras (1950), Nicaragua (1960), Brazil (1964), and Uruguay (1967) were established. Prior to the establishment of their respective central banks, Costa Rica, Uruguay, and Nicaragua had state banks that mixed central bank with commercial, agricultural, and mortgage banking activities (Triffin 1944).

In this period, the process of state-led industrialization was to a certain extent stimulated by American cooperation within the context of World War II, first, and then later during the Cold War. As a consequence, many Latin American countries also created development banks. The Brazilian Banco Nacional de Desenvolvimento Econômico e Social (BNDES) is emblematic of that period. As noted in Tavares et al. (2010), the bank was one of the suggestions made by the report of a joint

US-Brazil mission, which was concerned with increased rates of inflation. The mission recommended credit restrictions, but supported the creation of an investment bank to promote the infrastructure projects included in the report. This developmental period also coincided with an increase in inflation rates, mainly in South America as exemplified by the cases of Argentina, Bolivia, Brazil, Chile, and Uruguay. Between 1945–1955 and 1975–1979, the rates of inflation for Argentina and Uruguay increased from 20% to 204% and from 9% to 60.3%, respectively. In the case of Chile, inflation expanded from 28.4% to 227.1% in 1970–1975. For Brazil, the available data shows a rise in inflation from 14.7% in 1945–1955 to 101% in 1981. Finally, for Bolivia inflation creeped up from 40.4% in 1945–1955 to 123% in 1982 (Table 2).

However, with the exception of Bolivia, the highest inflations on record were experienced in the 1990s after the Debt Crisis of the 1980s as regional GDP per capita growth contracted in the next 3 years (by −1.8%, −3.6%, and −4.7% in 1981, 1982, and 1983, respectively). In the case of Bolivia, inflation kept on increasing after 1983 to reach 11,750% in 1985. The median annual rate of inflation for main countries (Argentina, Bolivia, Brazil, Chile, and Uruguay) reached 9.8% for the period 1935–1945, increasing to 34.8% between 1955 and 1965, reaching a peak of 594.8% in the period 1979–1992. Within this grouping the most renowned cases include Brazil (1,021.4% in 1993–1996), Bolivia (993% in 1979–1992), Peru (917.5% in 1979–1992), and Argentina (594.8% in 1979–1992). These high inflation cases were in general termed hyperinflation cases (a monthly rate of inflation ranging above 50 for at least a year), in similarity with the rates of inflation experienced by some European countries after World War I and World War II, although the historical similarity is highly questionable.

The rise in the prices of some the hyperinflation European countries or African countries is much higher than those in Latin American countries. The record rate of inflation is attributed to Hungary, which between August 1945 and July 1946 recorded an average rate of inflation of 19,800% per month with an extreme of $4.2 \cdot 10^{16}$% in July 1946. More recently, in November 2008, Zimbabwe nearly reached these levels ($7.96 \cdot 10^{10}$%). In Latin America, the record monthly inflation rate stands at 397% (Peru, August 1990). Also, European hyperinflations lasted in general a year, much longer than the so-called hyperinflations in Latin America which were observed only for a few months. Moreover, because of their duration, some European hyperinflation affected the level and composition of GDP. Germany's hyperinflation (Sept. 1922 to Nov. 1923) was accompanied by capital accumulation, very low unemployment levels, and the vertical integration of industry. At a closer inspection, the Latin American inflations turn out to be nothing more than high and extremely volatile case of inflation. The "hyperinflationary" misnomer served to justify drastic adjustment programs and a change in the orientation of monetary policy.

The Causes of Inflation and the Monetary Policy Prescriptions

The increasing concern with inflation as a disruptive force gave rise to two different and opposing interpretations regarding its origins and persistence: the structuralist and monetarist explanations. According to the former, inflation has its

Table 2 Annual inflation rates for Latin American countries, 1935–2016. In percentages

	1935–1945	1945–1955	1955–1965	1965–1970	1970–1975	1975–1979	1979–1992	1993–1996	1996–2000	2000–2010	2010–2016
Argentina	2.2	20.0	30.6	13.3	62.4	227.6	**594.8**	4.6	−0.1	8.8	10.2
Bolivia	21.4	40.4	39.5	13.2	19.4	10.1	**993.0**	9.8	6.3	4.8	5.2
Brazil	9.8	14.7	40.7	28.1	. . .	. . .	626.9	**1021.4**	7.6	6.7	6.8
Chile	13.1	28.4	34.8	135.7	**227.1**	150.4	22.2	9.9	5.2	3.2	3.2
Colombia	7.9	9.3	10.1	13.1	16.8	23.8	24.7	21.7	15.6	5.9	3.8
Costa Rica	7.7	5.2	2.0	2.6	12.5	8.0	25.0	16.0	12.7	10.4	3.7
Dominican Republic	16.3	2.8	1.3	1.3	9.8	9.6	22.8	7.9	6.5	12.5	4.1
Ecuador	15.7	12.1	1.9	8.6	12.2	12.2	35.8	29.9	47.9	16.5	3.6
Guatemala	9.3	7.6	0.2	1.4	7.6	11.2	15.5	10.5	7.6	6.7	4.1
Honduras	4.3	4.2	1.6	1.8	5.9	7.9	10.9	21.4	16.1	7.9	4.8
Nicaragua	34.3	6.3	1.4	1.7	16.0	14.9	**1960.1**	12.4	11.0	8.4	5.9
Peru	8.9	13.1	8.3	11.8	11.5	43.9	**917.5**	23.7	6.9	2.5	3.1
Paraguay	11.6	41.6	10.3	6.3	9.7	11.9	21.9	15.5	8.8	7.9	4.5
El Salvador	6.7	6.5	0.7	1.1	7.9	13.1	17.8	12.2	3.9	3.3	1.4
Uruguay	5.0	9.1	25.4	62.8	62.1	60.3	66.1	42.4	13.9	8.4	8.4
Venezuela, RB	3.7	3.6	0.5	1.6	5.2	9.0	24.9	64.7	45.1	21.7	79.3

Source: On the basis of World Bank (2018)

roots in the real rather than the monetary sphere (Furtado 1952; Noyola Vásquez 1956; Sunkel 1958). The canonical model of structural inflation separates inflationary pressures into basic, circumstantial, and cumulative pressures. Inflationary pressures are the initial causes of inflationary processes, while the propagation mechanisms maintain inertia or give strength (in the physical sense) to that process.

Basic pressures include the external constraint, deteriorating productivity, mediocre investment performance, an inadequate infrastructure system, and a tax system that is unable to cope with the needs of a modern society. Circumstantial pressures refer to exogenous events or decisions, while cumulative pressures are endogenous to inflation. The endogeneity of inflation allows incorporating the feedback from inflation to other variables while giving a dynamic character to the structural approach. The mechanism of propagation par excellence is the distributive struggle between the different agents/sectors of the economy, often associated to some form of formal wage indexation mechanism.

The opposite view, which was the one adopted by the central banks in the region, centered on the quantity theory of money and attributed the rise in process to excess supply over money demand. The excessive money was generally attributed to fiscal deficits. As is well known, the budget constraint is defined as:

$$G - T + rB = \frac{dB}{dt} + \frac{dM}{dt} \tag{1}$$

where G is the level of government spending (excluding interest on government debt), T is the income from taxes, r is the interest rate of government debt (B), and M is the level of high-powered money (monetary base).

When governments are constrained in their ability to increase taxes or access the capital market, government deficits are financed by increasing high-powered money $\left(\frac{dM}{dt}\right)$. The rise in high-powered money and money supply places a tax – the inflation tax – on cash balances by depreciating the value of money. The government revenue from inflation (i.e., GRI) is equal to the inflation rate (π) times the demand for real cash balances $\left(\frac{M}{P}\right)$. At zero inflation rate, the inflation tax revenue is also zero. As the rates of inflation become positive, the inflation tax revenues increase provided $\frac{M}{P}$ remains constant or as long as the increase in inflation offsets the decrease in real cash balances. During inflation real cash balances tend to decrease as agents try to avoid the tax on cash balances (the inflation tax).

At a certain point, the tendency for the demand for real cash balances to decrease will overwhelm the effect of the inflation tax, and it will cease to bring in revenue. The point of inflection of the inflation tax revenue curve is the point at which the inflation tax revenue is maximum, the point corresponding to the optimal inflation rate. Once the optimal rate is trespassed, inflation tax revenues decrease. As the inflation tax revenues decrease, prices will continue to increase, and the government will move to cover its financial needs by printing more money, thus increasing the inflation rate. This, in turn, through its effect on real cash balances will decrease the inflation tax revenues further.

Other explanations developed in the aftermath of the structuralist-monetarist debate on inflation traced its rise to balance-of-payments crises brought about by capital outflows, terms-of-trade shocks, and/or currency overvaluation. During the period several Latin American countries experienced either a currency or sovereign debt crisis. Laeven and Valencia (2012) register 9 and 24 currency crises in the 1980s and 1990s (Table 3). In some instances, balance-of-payments crises were accompanied with banking crises.

Balance-of-payments crises led to devaluation of the currency which led inevitably to inflation through different channels, including increase in the imported cost components of goods, expectations of further devaluation, the rise in the cost of servicing external debt, and loss of capital gains in the domestic bond market. These set the stage for further increase in prices through increases in nominal wages, further exchange rate devaluations and through the increase in the budget deficit and in the money supply. Other variables that also lend themselves to explain inflation dynamics include dollarization, capacity constraints, and indexation.

The onset of the 1980s Debt Crisis, which is considered the worst crisis episode to hit the region since the 1930s Great Depression, produced a Copernican turn in economic and monetary policy. Latin American countries adopted Washington Consensus policies which consisted mainly in maintaining fiscal discipline, liberalizing trade and financial markets, deregulating the economy, and privatizing state-owned industrial firms and public utilities (Williamson 1990).

Within this context the preferred approach to deal with inflation and to formulate stabilization policies was the monetary approach to the balance of payments (MABP). This viewed inflation and balance of payments deficits as a result of a disequilibrium in the money market (mainly generated by budget deficits) and hence as a monetary phenomenon (Polak 1957; Frenkel and Johnson 1976; International Monetary Fund 1977; Kreinin and Officer 1978; Mundell 1968). The chain of causality ran from budget deficits to credit expansion to increasing real cash balances and a rise in absorption above domestic output, inducing both inflation and a deficit in the balance of payments. The adjustment is brought about by changes in the stock of foreign reserves which drives the supply of money. A deficit (surplus) in the balance of payments produces a decline (increase) in the stock of foreign reserves leading to a reduction (rise) in the money supply resulting in a reduction (increase) in absorption until it conforms to the level of output. With all its assumptions (the world is an integrated capital and commodity market; domestic price levels conform through competition to a one-world price level; the price level is pegged to the world price level moving rigidly in line with it; full employment), the MABP has important implications for monetary policy.

First, the relevant control variable is the rate of growth of credit rather than the rate of growth of the money supply. Second, the world rate of inflation is determined by the world money supply. As long as countries maintain pegged or fixed exchange rates, they cannot avoid conforming over time to the world price level. Third, fiscal and monetary policies do not have any effect on real aggregate output. Fiscal policy can only change the composition of government expenditure and the proportion financed by debt. For its part monetary policy has no control over the money supply;

Table 3 Systemic banking, currency, and sovereign debt crises for Latin American and Caribbean countries 1970–2010

Country	Systemic banking crisis (starting date)	Currency crisis (year)	Sovereign debt crisis (default date)	Sovereign debt restructuring (year)	Total crises per country
	By country				
Argentina	1980, 1989, 1995, 2001	1975, 1981, 1987, 2002	1982, 2001	1993, 2005	12
Bolivia	1986, 1994	1973, 1981	1980	1992	6
Brazil	1990, 1994	1976, 1982, 1987, 1992, 1999	1983	1994	9
Chile	1976, 1981	1972, 1982	1983	1990	6
Colombia	1982, 1998	1985	–	–	3
Costa Rica	1987, 1994	1981, 1991	1981	1990	6
Dominica	–	–	2002	n.a.	1
Dominican Republic	2003	1985, 1990, 2003	1982, 2003	1994, 2005	8
Ecuador	1982, 1998	1982, 1999	1982, 1999, 2008	1995, 2000, 2009	10
El Salvador	1989	1986	–	–	2
Guatemala	–	1986	–	–	1
Guyana	1993	1987	1982	1992	4
Haiti	1994	1992, 2003	–	–	3
Honduras	–	1990	1981	1992	3
Jamaica	1996	1978, 1983, 1991	1978, 2010	1990, 2010	8
Mexico	1981, 1994	1977, 1982, 1995	1982	1990	7
Nicaragua	1990, 2000	1979, 1985, 1990	1980	1995	7
Panama	1988	–	1983	1996	3
Paraguay	1995	1984, 1989, 2002	1982	1992	6
Peru	1983	1976, 1981, 1988	1978	1996	6
Suriname	–	1990, 1995, 2001	–	–	3
Trinidad and Tobago	–	1986	1989	1989	3
Uruguay	1981, 2002	1972, 1983, 1990, 2002	1983, 2002	1991, 2003	10
Venezuela	1994	1984, 1989, 1994, 2002, 2010	1982	1990	8

(*continued*)

Table 3 (continued)

Country	Systemic banking crisis (starting date)	Currency crisis (year)	Sovereign debt crisis (default date)	Sovereign debt restructuring (year)	Total crises per country
	By country				
Decade	Averages				
1970s	1	9	2	0	–
1980s	12	24	16	1	–
1990s	14	14	1	17	–
2000s	4	8	6	6	–

Source: Data file to accompany the IMF Working Paper "Systemic Banking Crises Database: An Update" Laeven and Valencia 2012. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Systemic-Banking-Crises-Database-An-Update-26015

it is an endogenous variable (the public adjusts the nominal supply of money to its demand by exporting or importing money via the balance of payments). Monetary policy can only determine the composition of the money supply between net domestic assets and international reserves. Fourth, international reserve growth is positively related to domestic growth and the income elasticity of demand for money. Finally, the effects of a devaluation are akin to those of a deflationary policy stance. A devaluation of the currency increases domestic prices, ipso facto, through the law of one price and translates into a decrease in real cash balances. Economic agents' restoration of their desired level of real cash balances leads to an increase in lacking (to use the Robertsonian term) and a lower level of expenditure.

The Change of Orientation in Monetary Policy

In line with these changes, central banks in the region reformed their legislation placing price stability as the primary objective of monetary policy excluding any reference to the promotion of economic development or growth through monetary means. To this end central banks had to be free from political interference in the design and implementation of monetary policy including from the constraints derived by lending to the government to cover fiscal deficits. This implied both political and operational independence.

Political independence materialized in granting tenure periods to the board of directors coinciding or exceeding those of governments. Also, the board of directors does not include representatives from the government or private businesses. A government representative, generally the minister of finance, is permitted to sit in the meetings of the board of directors but does not have voting power. Furthermore, the legislative branch plays an important role in either appointing or confirming the members of the board of directors. Finally, the reformed legislation sought to provide

greater control of central bank's actions by the periodic reporting of central bank actions, aims, policy implementation and results, and financial statements (for the most part certified by an external source to the central bank). In some cases, the legislation also included hearings before the legislative branch to explain monetary policy reports (Table 4).

During the 1990s, the three-digit (or higher) inflation rates disappeared in the region, and the majority of Latin American and Caribbean economies witnessed a decline in their inflation, in line with global developments. The average rate of inflation fell from 336% (1979–1992) to 82.8% (1993–1996) and finally to 13.4 (1996–2000). In the period 1979 to 1992, no country in Latin America had a one-digit inflation rate. In 1993–1996, four countries had a one-digit inflation rate, and in the period 1996–2000, half of the countries in Latin America reported one-digit inflation rates. These became the norm in the 2000s. The decline in inflation was worldwide rather than a phenomenon specific to Latin American countries.

In the particular case of Latin America, both countries that introduced changes in the bank legislation in the 1990s and those that did not witnessed a decline in the rate of inflation. In the 1990s central banks in the region faced two significant challenges. First, central banks often targeted the nominal exchange, leading them to have two nominal anchors: the nominal exchange rate and the price level. The second challenge was that, in the 1990s, central banks continued to deal with financial and banking crisis.

The existence of two policy targets could lead to important contradictions and inconsistencies in the formulation and implementation of monetary policy. This can be justified by the fact that in small open economies, the exchange rate plays a key nominal and real role and that the monetary authority or central bank cannot be oblivious to its fluctuations. Nonetheless, maintaining two nominal anchors can pose significant inconsistencies in monetary policy objectives. For one thing, stabilizing the exchange rate can, under certain conditions, e.g., in the case of external shocks, take precedence over inflation. Also, seeking to stabilize the exchange rate or giving priority to the external sector (maintaining external competitiveness) over domestic conditions can lead to a pro-cyclical policy amplifying real and nominal fluctuations.

The importance of the external sector and the way in which it can enter into conflict with inflation objectives is illustrated by the Chilean experience in the 1990s. During this time the authorities implemented several discretionary and ad hoc modifications to the exchange rate regime in an effort to control exchange rate fluctuations. These weakened some of the very same elements that form part of the core of the new framework in monetary policy adopted in the 1990s, namely, transparency and communication.

Chile adopted in 1991–1992 a crawling peg in relation to the dollar accompanied by daily devaluations according to the internal-external inflation differential. Thereafter, the exchange rate regime switched to a target zone around a basket peg.

Table 4 Key features of central banks following the reforms of the 1990s

Main objective			
Price stability as the sole or primary objective	Price stability plus other objectives, with no indication of priority		
	Operation of the payment system	Stability of the financial system	Growth or economic development
Argentina (1992), Bolivia (1995), Colombia (1992), Costa Rica, Dominican Republic, Mexico (1993), Peru (1993), Venezuela	Chile (1989), Honduras, Nicaragua	Guatemala, Paraguay, Uruguay	Brazil
Credit to the government by the central bank			
No direct or indirect credit, or credit extended on the secondary market with limits	Credit to cope with seasonal liquidity shortages, or credit extended on the secondary market without limits		Direct credit
Argentina, Chile, Costa Rica, Brazil, Dominican Republic, Guatemala, Peru, Uruguay, Venezuela	Bolivia, Honduras, Mexico, Nicaragua, Paraguay		Colombia
Independence in the use of monetary policy instruments			
Full independence in monetary and exchange rate policy	Restrictions on the conduct of monetary or exchange rate policy		Monetary and exchange rate policy set by the government
Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Honduras, Nicaragua, Peru, Uruguay	Guatemala, Mexico, Paraguay, Venezuela		
Financial independence			
Government is required to maintain central bank's capital	Government is authorized but is not required to capitalize the central bank, or capitalizes it with nonnegotiable bonds		No legal basis exists for the government capitalizing the central bank
Brazil, Colombia, Dominican Republic, Guatemala, Nicaragua, Peru, Venezuela	Bolivia, Chile, Mexico		Argentina, Costa Rica, Honduras, Paraguay, Uruguay
Reporting by the central bank			
Formal appearance before the legislature	Submission of a report to the executive or legislative branch, or publication of a report in the news media		
Argentina, Brazil, Chile, Colombia, Dominican Republic, Guatemala, Mexico, Paraguay, Venezuela	Bolivia, Costa Rica, Honduras, Nicaragua, Peru, Uruguay, Mexico		

(*continued*)

Table 4 (continued)

Publication and transparency of information		
Financial statements certified by external auditors	Financial statements certified by a public agency separated from the central bank	Financial statements signed by an auditor appointed by the central bank's executive board
Argentina, Chile, Guatemala, Mexico, Nicaragua	Brazil, Colombia, Honduras, Paraguay, Uruguay, Venezuela	Bolivia, Costa Rica, Dominican Republic, Peru

Source: Latin American national constitutions and central bank legislation as of 2003, and Fiscal Responsibility Law in Brazil; Carstens and Jácome 2005; Schmidt-Hebbel 2011

The central parity was tied to a basket of currencies including the US dollar, the Deutsche mark, and the Japanese yen. Thereafter until 1999, there were several and frequent modifications to the central parity, to the currency basket weighs, and to the bands. In fact, overall from 1992 until 1999, there were ten different changes to the parameters of the exchange rate target zone regime. The main difficulty with the exchange rate regime and the different changes was the fact that it was very hard on the basis of the information provided for market participants to know or understand the type of exchange rate regime that prevailed at the time.

Also, as noted above, financial and banking crisis were prevalent in the 1990s. The available data for 24 Latin American and Caribbean countries show only one sovereign debt crisis corresponding to Ecuador, which as a result dollarized its economy. El Salvador and Panama are the other two countries that have a dollarized economy, while Argentina reinstituted a Currency Board and kept a fixed exchange rate to the dollar. This is a sharp drop from the 16 cases registered in the 1980s. However, in both the 1980s and 1990s decades, currency crises persisted (24 and 14 for the 1980s and 1990s, respectively), and the number of systemic banking crisis increased (12 and 14 cases for the same periods, respectively).

The most significant crises episodes in the region include the Mexican "Tequila" (1994–1995), the Brazilian (1999), and Argentinean (2001–2002) crises. The Tequila crisis, whose epicenter was Mexico, had contagion effects mainly in Argentina and to a lesser extent in Brazil. Mexico and Argentina's GDP per capita growth fell by −8.0 and −4.1%, respectively. The contagion effects of the Asian crisis and Russian-Brazilian were felt throughout emerging market economies and in Latin America affected Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, and Venezuela. Finally, the Argentinean crisis (Argentina recorded a contraction of 5.5% and 11.9% in real GDP per capita in 2001 and 2002) impacted on Brazil, Paraguay Peru, Uruguay, and Venezuela.

The comparative analysis of these three episodes shows important common features across country experiences prior and during to crises/contagion episodes. The evidence indicates that prior to the year in which the crises materialized or the year the contagion effects were felt, the government accounts registered declining imbalances. Also, the year prior to the crisis or the year in which contagion effects were felt, the government had either a mild fiscal deficit or a surplus. Contrarily, the

private sector recorded a widening imbalance, and this was a major source of vulnerability. The private sector deficit was reflected in the current account. In the majority of cases analyzed, the rest of the world constitutes an important source of liquidity for the economy and is an important source of finance to cover the private sector's deficit.

The crises/contagion transmission mechanisms include in general a decline in external provision of finance, the contraction in domestic credit, and the deleveraging by the private sector. The government sector does not show, across the different episodes analyzed, a common pattern of behavior, although in some cases the fiscal accounts tend to deteriorate as a result of lower growth of the economy and thus of government revenue. For its part, the policy reaction of the monetary authority is pro-cyclical and illustrates the standard approach to confront crises/contagion episodes. The evidence shows that the central bank restraining rather than providing liquidity to the private sector during the crisis episodes, to counteract the fall in spending. The central bank did provide extra financing to the financial sector in order to maintain financial stability. But in no case is the extra financing to the financial sector a source for the provision of countercyclical liquidity for the economy (Pérez Caldentey and Cruz 2017).

The Adoption of Inflation Targeting

Since the 2000s Latin American countries have turned to inflation targeting strategies (Table 4). Inflation targeting is traditionally defined as a monetary policy strategy framework consisting in the public announcement of numerical targets for the inflation rate, acknowledging that price stability (a situation where the inflation rate at which inflation is no longer a public concern, Clarida et al. 1999) is the fundamental goal of monetary policy and a firm commitment to transparency and accountability (Bernanke et al. 1999; Svensson 2010). The main instrument of monetary policy is the management of the short-term interest policy rate through a Taylor type policy rule (Eq. 4 below).

Within the context of this definition, numerical targets can refer to a point inflation rate, a range or a point with a tolerance range. The inflation rate can refer to the consumer price index (CPI) as is the standard case for most developing economies or to the core CPI. In an open economy, inflation targeting favors a free-floating exchange rate regime to avoid having two nominal anchors which can lead to the type of contradictions that surfaced in some Latin American countries in the 1990s (described in the section above). Within an inflation targeting regime, external shocks, which are one of the main sources of volatility and cyclical fluctuations, are absorbed by movements in the exchange rate thus avoiding the direct transmission from these impulses to the domestic economy.

Transparency means that the monetary authorities must communicate their targets, forecasts of inflation, decisions on monetary policy, and the motivation for their decisions. Finally, accountability here means that the monetary authorities are

responsible for attaining the announced objectives and subject to "public scrutiny for changes in their policy or deviations from their targets."

The canonical model of inflation can be stated as follows:

$$y_t^g = f\left(y_{t-i}^g, y_{t+i}^g\right) + f(i_t - E\pi_{t+i}) + u_t \tag{2}$$

$$\pi_t = f\left(y_t^g, \pi_{t-i}, , E\pi_{t+i}\right) + v_t \tag{3}$$

$$i_t = f(r^*, E\pi_t, y_t^g) + f(\pi_t - \pi^*) \tag{4}$$

where y^g is the output gap (i.e., the difference between actual and potential output), π_t,π^* are actual and target inflation rates, i_t is the actual short-term nominal interest rates (i.e., the monetary policy rates), E is the mathematical expectation, and u_t,v_t are random errors.

The inflation targeting framework makes clear that low and stable inflation should not be pursued at any cost. Moreover, monetary authorities do not necessarily have to pursue a rigid inflation targeting regime. Rather they can opt for a "flexible" inflation targeting. The latter implies that the monetary authorities or the central bank do not have only a monetary objective (stabilizing inflation) but have also a real objective (stabilizing real output). The adoption of flexible inflation targeting entails pursuing a gradualist approach to the achievement of monetary policy objectives. Flexible inflation targeting and hence a gradualist approach to monetary policy are conceptually justified mainly on the grounds of uncertainty regarding (i) the workings and current state of the economy, (ii) the transmission mechanisms and policy parameters, and (iii) the nature of external shocks as well. A gradualist policy can also contribute to buffer the effects on real variables caused by external shocks.

Inflation targeting strategies permit monetary authorities to pursue a "lean against the wind" policy while avoiding large fluctuations, and indeed lowering volatility, of output. Within the international setting of the early twenty-first century, this monetary practice must be accompanied by a consistent and coherent fiscal policy without endangering stability or development objectives. This consists in formulating fiscal policy objectives and instruments allowing it to play at the same time a stabilization role while recapturing its distributive role.

Several Latin American countries have tended in the past decade to show their preference for inflation targeting regimes including Brazil (1999), Colombia (1999), Chile (1999), Guatemala (2005), Mexico (2001), and Peru (2002). To those countries that already have in operation a full regime of explicit inflation targets must be added others that are in the process of implementation of the scheme (Costa Rica, the Dominican Republic, Paraguay, and Argentina) (Table 5).

Costa Rica decided to orient its monetary policy toward an inflation targeting regime in 2005. Paraguay chose to initiate the formal start-up of this scheme in 2011. Meanwhile, the Dominican Republic began its transition to an inflation targeting regime scheme in the 2012. Argentina adopted inflation targeting in 2016, but it seems to have abandoned it, in the midst of the 2018 crisis. Uruguay is an atypical

Table 5 Inflation targeting regimes in Latin America

Country	Adoption of target	Inflation measure	Target (2006)	Target horizon	Inflation report	Published forecast
Brazil	June 1999	CPI	4.5% (+/−2%)	12 months	Yes	Yes
Chile	January 1991	CPI	2%–4% centered at 3%	12–24 months	Yes	Yes
Colombia	September 1999	CPI	4%–5%	12 months	Yes	Yes
Mexico	January 1999	CPI	3%(+/−1%)	12 months	Yes	Yes
Peru	January 2002	CPI	2.5% (+/−1%)	12 months	Yes	Yes

Source: On the basis of official information

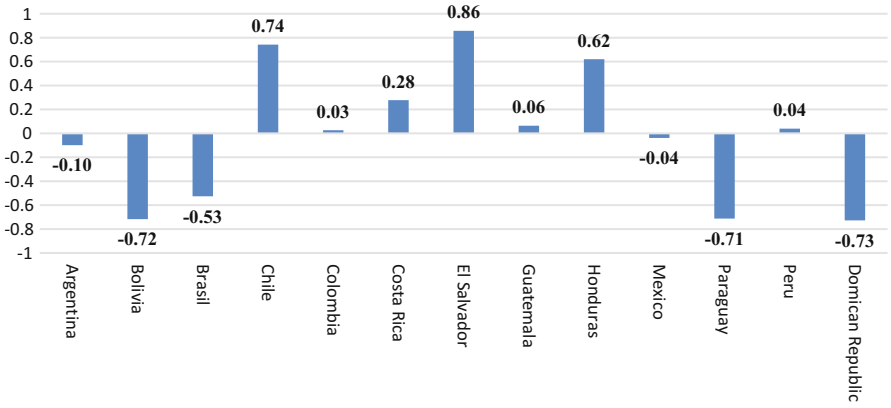


Fig. 1 Correlation coefficients between the output gap and the inflation rate for selected Latin American economies (circa 1995–2016). Note: The output gap is the difference between actual and potential GDP. (Source: On the basis of Aravena et al. (2018) and World Bank Development Indicators (2018))

case as much as this country began to implement an inflation targeting scheme in 2007 and decided to reverse toward a quantitative target regime in 2012.

Some Evidence on Latin American Inflation Targeting Regimes

The available evidence shows that inflation targeting regimes do not necessarily practice countercyclicality and do not adhere in general to floating exchange rates. Data for 13 Latin American countries show that the correlation coefficients between the output gap and the monetary policy rate is negative as expected in less than half of the cases considered (Fig. 1).

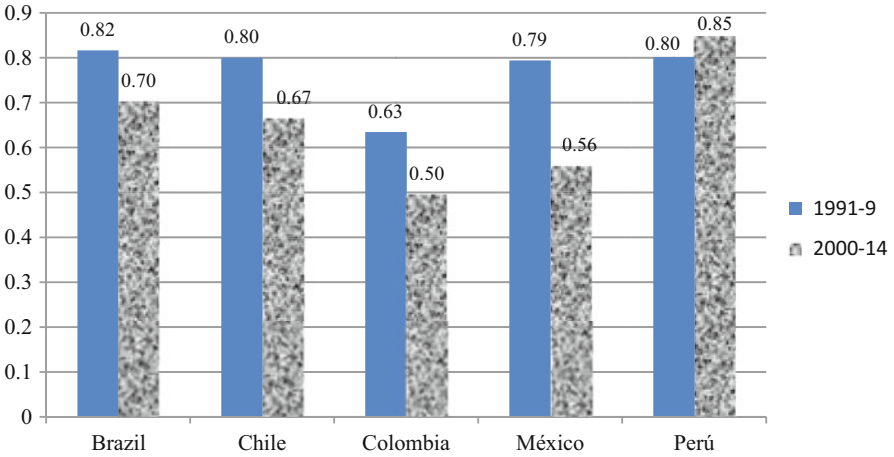


Fig. 2 Exchange rate intervention index for Latin American countries with inflation targeting regimes. 1991–1999 and 2000–2014 (averages). Note: $IIC = \frac{\sigma_{\Delta international\ reserves}}{\sigma_{\Delta International\ reserves} + \sigma_{\Delta Real\ exchange\ rate}}$, where σ= standard deviation and $\Delta = x_t - x_{t-4}$. (Source: On the basis of IMF (2013–2018))

At the same time, countries that have adopted inflation targeting regimes show "fear of floating" and frequently intervene in the foreign exchange market. This is illustrated in Fig. 2 for the case of Latin American countries that have in force inflation targeting scheme. The metric used is an index of exchange rate intervention (IIC) (see note below Fig. 2) before and after the adoption of this scheme. The IIC takes values between 0, which reflects a situation of pure flotation, and 1 indicating that the authorities intervene systematically to soften variations in the exchange rate. Although in all countries the degree of exchange rate intervention decreased in the twenty-first century compared to the 1990s (with the exception of Peru), none of the countries considered adheres to a nonintervention scheme. Also, the IIC is for some of these inflation targeting economies close to the 0.73 obtained by Ostry et al. 2012, for developing countries that do not adhere to inflation targeting.

Foreign exchange intervention is due to the fact that in a context of financial openness and globalization, the nominal exchange rate is a price that responds to expected returns (whether profit or potential losses) in future markets. In this sense the exchange rate behaves like the price of an asset. The lack of systematic effectiveness of arbitrage results in a high instability in the exchange market that can only be tackled with a systematic policy of accumulation of reserves. As shown in Table 6, full-fledged inflation targeting countries have increased their stock of reserves as a percentage of GDP.

According to the logic of the inflation target scheme, in the absence of sterilization operations, the accumulation of international reserves translates into an increase in the monetary base and liquidity of the economy, which can pose a threat to the achievement of the inflationary target. In the worst case, it can jeopardize monetary stability and even financial stability. To avoid these risks, central banks tend to sterilize the expansive effect of the increase in net international reserves. The

Table 6 International gross reserves as percentage of GDP for full-fledged inflation targeting countries 1990–2000 and 2001–2017 (averages)

Period	Brazil	Chile	Colombia	Mexico	Peru
1990–1995	5.5	20.9	9.8	–	12.5
1996–2000	5.9	19.6	8.1	5.0	17.8
2001–2005	7.3	18.2	11.0	7.4	17.9
2006–2009	11.6	12.4	10.1	8.9	25.0
2010–2017	16.2	15.2	12.7	14.2	31.0

Source: On the basis of official information

Table 7 Sterilization coefficient (EC) for Latin American countries with inflation targets 1996–2000, 2000–2006, and 2006–2013 (averages)

	1996–2000	2000–2006	2006–2013
País	CE	CE	CE
Brazil	0.57	0.58	0.84
Chile	0.92	0.19	0.66
Colombia	n.d.	−0.50	0.45
Mexico	0.85	0.59	0.67
Peru	0.86	0.99	1.00

Note: n.d. = not available
Source: On the basis of IMF (2013–2018)

intensity of the sterilization operations can be assessed from the quotient, $\frac{\Delta AEN - \Delta CC}{\Delta AEN}$, where AEN = NET external assets and CC = Cash.

It reflects the extent to which monetary authorities apply sterilization policies that include not only the extent to which they use bond market interventions to neutralize the effect of stockpile accumulation on the monetary basis but also integrate the variations in the reserves of commercial banks in the central bank to adjust the monetary multiplier (Lavigne 2008). The sterilization ratio fluctuates between 0 and 1. The lower value of the sterilization coefficient (0) reflects a non-sterilization situation, while the higher value (1) refers to a condition of complete sterilization. Table 7 shows the respective sterilization coefficient for Brazil, Colombia, Chile, Guatemala, and Peru for periods 1996–2000, 2000–2006, and 2007–2013.

The evidence shows that in the inflation targeting period, all countries have tended to increase the intensity of their sterilized interventions. Sterilized interventions can put upward pressure on interest rates. The increase in the interest rate is self-corrective in a context of currency depreciation but produces a cumulative effect under a situation of appreciation, which is when most of the exchange interventions take place.

In addition, sterilization policies may have significant quasi-fiscal consequences for the interest rate differential between domestic rates and external rates, which is large for Latin American countries with inflation targets due in part to the low interest rates prevailing in the United States in the decade of the 2000s. In fact, on average in the period 2001–2013, the differentials between the yields of the treasury

bonds of the United States and its equivalent for the countries with inflation targets amounted to 10, 6, 1, 3, and 1 percentage points in the case of Brazil, Colombia, Chile, Mexico, and Peru, respectively.

Conclusion

Monetary policy is considered the central pillar of economic policy and the main instrument for economic stabilization and to approximate the actual to the potential level of GDP. As things stand, inflation rates continue to hover, with a few exceptions, notably Argentina and Venezuela, at single digits pushing aside inflation as a macroeconomic threat. Yet in some cases, inflation stabilization has been obtained at the expense of output stabilization. Monetary policy needs to gain an improved understanding of the transmission mechanism between monetary policy and the real economy in a more financially complex context and especially when monetary policy is contractionary.

The existing monetary dominance has been reflected in the adoption of rules in the form of inflation targeting regimes. These schemes that are conceptualized within the framework of the new consensus on macroeconomics have been increasingly imposed in the developing and developed world and have relegated fiscal policy to a secondary level and more specifically, in some cases, to the status of a stochastic error. In turn to avoid the destabilizing effect of this stochastic error, monetary rules have been supplemented with fiscal rules.

This macroeconomic policy poses important costs and risks in both financial and real terms. Fiscal policy has been transformed into a subsidiary policy to the objectives guided by maintaining macroeconomic balances and has paid little attention to fiscal expenditure and its composition, taking for granted the effect of human capital expenditure and physical in long-term growth. In fact, fiscal policy has simply become a social welfare policy. The lack of attention paid to private indebtedness and dependence on external financing exposes the performance and growth trajectory to fluctuations in international financial flows. More importantly, it can generate its own volatility to boom-and-bust cycles, which is exactly what was intended to avoid by subordinating fiscal policy to monetary dominance. Indeed, one of the most important challenges to monetary management in Latin America are the dominating influence of international trade and financial flows on economic activity and the constraints these impose on central bank's policies. In this sense, financial stability including macroprudential policy which has become a broader and more relevant theme that central banks must be incorporated in central banks' mandates.

Cross-References

- ▶ Currency Boards
- ▶ International Monetary Regimes: The Bretton Woods System
- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ The Anatomy of Inflation: An Economic History Perspective

References

Aravena C, Lopez G, Pineda R (2018) Producto Potencial de Mediano Plazo en América Latina. Mimeo, Santiago. Chile

Bernanke B, Laubach T, Mishkin FS, Posen AS (1999) Inflation targeting. Lessons from the international experience. Princeton University Press, Princeton

Carstens A, Jácome HLI (2005) Latin American Central Bank reform: progress and challenges. IMF Work Pap 05(114):1

Clarida R, Galí J, Gertler M (1999) The science of monetary policy: a new Keynesian perspective. J Econ Lit XXXVII:1661–1707

Dalgaard BR (1980) Monetary reform, 1923–30: a Prelude to Colombia's economic development. J Econ Hist 40(1):98–104

Drake P (1989) The money doctor in the Andes: the Kemmerer missions, 1923–1933. Duke University Press, Durham

Flandreau M (2003) The money doctors: the experience of international financial advising 1850–2000. Routledge, New York

Frenkel JA, Johnson HG, (eds) (1976) The monetary approach to the balance of payments. London/Toronto: Allen & Unwin/University of Toronto Press.

Furtado C (1952, 1954) Capital formation and economic development. International economic papers. In: Agarwala AN, Singh SP (eds) Reproduced in The Economics of Underdevelopment, 1958. 4(3):124–144. Oxford University Press, Oxford.

Glaser E (2003) Chile's monetarist money doctors, 1850–1988. In: Flandreau M (ed) The money doctors: the experience of international financial advising 1850–2000. Routledge, New York, pp 166–185

Helleiner E (1999) Central bankers as good neighbours: US money doctors in Latin America during the 1940s. Financ Hist Rev 16(1):5–25

IMF International Monetary Fund (2013–2018) International financial statistics. IMF, Washington, DC

International Monetary Fund. (1977) The monetary approach to the balance of payments. Washington, DC

Jácome HLI (2015) Central banking in Latin America; from the gold standard to the golden years. IMF Work Pap 15(60):1

Kemmerer EW (1916) The theory of foreign investments. Ann Am Acad Pol Soc Sci 68:1–9

Kemmerer EW (1927) Economic advisory work for governments. Am Econ Rev 17(1):1–12

Kemmerer EW (1944) Gold and the gold standard. McGraw-Hill Book Company, New York

Kreinin M, Officer L (1978) The monetary approach to the balance of payments: A survey, Princeton Studies in International Finance No. 43. Princeton: Princeton University Press

Laeven L, Valencia F (2012) Systemic banking crises database: an update. IMF Work Pap 12(163):1

Lavigne R (2008) "Sterilized intervention in emerging market economies: trends, costs and risks", Discussion Papers, N° 08-4, Banco de Canadá

Marichal C (2008) Banking history and archives in Latin America. Bus Hist Rev 82(3):585–602

Marichal C, Gambi T (2017) Historia Bancaria y Monetaria de América Latina (siglos XIX y XX). Nuevas Perspectivas. Ediciones de la Universidad de Cantabria, Santander

Mundell RA (1968) International economics. Macmillan, New York

Noyola Vásquez J (1956) El Desarrollo Económico y la inflación en México y otros países Latinoamericanos. Investig Econ XVI(4):603–648

Ostry J, Ghosh A, Chamon M (2012) Two targets, two instruments: monetary and exchange rate policies in emerging market economies. IMF staff discussion note. International Monetary Fund, Washington, DC

Pérez Caldentey E, Vernengo M (2018) Heterodox central banking in the periphery. Res Hist Econ Thought Methodol 36C:83–102

Polak J (1957) Monetary Analysis of Income Formation and Payments Problems, https://en.wikipedia.org/wiki/IMF_Staff_Papers. IMF Staff Papers 6(1):1–50

Posso Ordóñez RV (2016) Historia de la Creación de la Banca Central Latinoamericana -El Pretérito Es la Base de un Presente prominente. Tendencias 17(2):166–187

Prebisch R (1991) Obras 1919–1949, Tomos I-III, Fundación Raúl Prebisch, Buenos Aires
Schmidt-Hebbel K (2011) Central banking in Latin America: changes, achievements, challenges. Bank of Spain occasional paper, vol 1102. Banco de España, Madrid
Seidel RN (1972) The Kemmerer missions in South America, 1923–1931. J Econ Hist 32(2):520–545
Sunkel O (1958) La inflación – un enfoque heterodoxo. El Trimestre Económico Oct.-Dic:570–599
Svensson LEO (2010) Inflation targeting. In: Durlauf SN, Blume LE (eds) Monetary economics. The New Palgrave economics collection. Palgrave Macmillan, London
Tavares M d C, de Melo HP, Caputo AC, Costa GM, de Araújo VL (2010) As Origens do Banco Nacional de Desenvolvimento Econômico (BNDE) 1952–1955. Memorias do Desenvolvimento 4(4):13–44
Triffin R (1944) Central banking and the monetary management in Latin America. March 4, Mimeo, Washington, DC. Published as "Central banking and monetary management in Latin America". In: Harris SE (ed) Economic Problems of Latin America. McGraw-Hill, New York, pp 93–116
Triffin R (1947) Monetary and Banking Reform in Paraguay. Board of Governors of the Federal Reserve System. Washington D.C
Triffin R (1981) An economist's career: what? why? how. Banca Nazionale del Lavoro Quarterly Review 138(3):239–259
Wallich HC, Triffin R (1953) Monetary and banking legislation of the Dominican Republic 1947 with reports and letter of transmittal. Federal Reserve Bank of New York, New York
Willliamson J (ed) (1990) Latin American adjustment: how much has it happened? Institute for International Economics, Washington, DC
World Bank (2018) World Bank development indicators. World Bank, Washington, DC

Part X

Aggregate Price Shocks

Bullionism

38

Joshua R. Hendrickson

Contents

Abstract

The term "bullionism" describes a belief that the price of gold, measured in domestic currency, is an appropriate indicator of inflation and the depreciation of the currency caused by an excess supply of money. The term bullionism arose from what has come to be known as the Bullionist Controversy in the United Kingdom. In 1797, while at war with France and in the midst of declining reserves at the Bank of England, Britain suspended the convertibility of Bank of England notes into gold. The British also pressured Ireland to suspend the convertibility of Bank of Ireland notes. Following the suspension of convertibility, both countries experienced a rising price of gold and a depreciation of their exchange rate. Since price indexes were not available at that time, the debate was about the cause of the rising price of gold and depreciation of the exchange rate. The Bullionists argued that the rising price of gold was an indicative of rising prices in general and that the rising prices and depreciation of the exchange rate were due to an excess supply of bank notes. The Antibullionists denied that an excess supply of bank notes was to blame for depreciation and instead argued that

J. R. Hendrickson (✉)
Department of Economics, University of Mississippi, Oxford, MS, USA
e-mail: jrhendr1@olemiss.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_22

the cause was a combination of foreign remittances, bad harvests, and an adverse balance of trade. A similar debate took place in Sweden over a half of a century earlier when elected officials used what is now known as the Sveriges Riksbank to finance an expansion of commercial activity with new bank notes. These debates carry important lessons about the determination of the price level and the exchange rate under the gold standard and under a paper money standard. The empirical literature on these periods is somewhat limited. This chapter summarizes the existing literature and points to areas for future research.

Keywords
Bullionism · Quantity theory of money · Bullionist controversy

JEL Classification
B12 · N13 · E42

Introduction and Background

Bullionism is the idea that a rising price of gold in terms of the domestic currency is an indicator of an excess supply of money under a paper money standard. The term finds its origins in what have come to be known as the bullionist controversies in Great Britain, Ireland, and Sweden. While the bullionist controversies have received their name due to the emphasis on the price of gold in terms of the domestic currency, the controversies were actually larger debates about the determination of the price level and the exchange rate under a paper money standard. What made such arguments difficult was the lack of appropriate empirical estimates of the price level or inflation since "the notion of an index number was still in its infancy" (Viner 1937, p. 126). As a result, the rising paper price of gold was used as an indicator of inflation and depreciation of the exchange rate. Regardless of the focus on the paper price of gold, these debates contributed to our understanding of the differences in the determination of the price level and exchange rate under different monetary regimes. The focus here will be on the empirical evidence used to evaluate the claims made on each side of these debates and the lessons that evidence provides for our understanding of different monetary regimes. Nonetheless, it is appropriate to start with some background in order to understand the circumstances surrounding each debate and the interpretation of the evidence.

The most famous of the bullionist controversies took place in Great Britain. During the 1790s, the British technically had a bimetallic system, but were effectively on a gold standard. The Bank of England had a monopoly over note issuance in London and could be considered at that time to be the center of the British monetary system (Laidler 2000). The Irish monetary system was centered in Dublin and remained the case following the Acts of Union that created the United Kingdom of Great Britain and Ireland.

In 1792, Great Britain went to war with France. They would remain at war throughout both the French Revolutionary Wars and the Napoleonic Wars that

followed. By 1795 and 1796, the Bank of England was experiencing a significant outflow of specie. Rumors of a French invasion of Wales in early 1797 caused a run on banks in Britain and a further drain on reserves at the Bank of England (Laidler 2000). Finally, in February of 1797, the British government suspended the convertibility of bank notes at the Bank of England. Convertibility was not officially restored until 1821. Over this period of time, known as the Bank Restriction Period, the British experienced a dramatic increase in the price level. From 1797 to 1801, the price level increased 50%. After a significant reduction in the price level from 1801 to 1803, inflation resumed with the price level following an uninterrupted upward trend until 1813. The cause of this increase in prices was subject to considerable debate. This debate is commonly referred to as the Bullionist Controversy. At the time, since data on the price level was not available, participants in the debate focused on the paper price of gold or the exchange rate. These debates were complicated by the fact that the British lacked a theory of flexible exchange rates or central banking, more generally (Laidler 2000). Nonetheless, participants such as Henry Thornton proved to be sufficiently up to the task (Laidler 2000, Hendrickson 2018).

The two groups in the debate were the Bullionists and the Antibullionists. The Bullionists argued that the rising price of gold bullion was a sign of overall inflation that was caused by an excess issuance of bank notes on the part of the Bank of England following the suspension of convertibility. The Antibullionists argued that the rising prices could be blamed on foreign remittances, poor harvests, war expenditures, and foreign subsidies. In fact, some Antibullionists adopted what we would now refer to as the real bills doctrine, arguing that the central bank was incapable of producing an excess supply of bank notes.

When Britain suspended convertibility, Ireland was still an independent state. Since Ireland was independently governed, the suspension of convertibility did not apply to the Bank of Ireland. Nonetheless, the British pressured Ireland to suspend convertibility at the Bank of Ireland (Fetter 1955). After some debate, the Irish agreed to suspend convertibility at the Bank of Ireland. Just as the restriction in Britain had only applied to the Bank of England, the suspension of convertibility in Ireland only applied to the Bank of Ireland. In 1801, the Act of Union brought Ireland into the United Kingdom. The Bank of Ireland was able to maintain its own monetary policy, but the suspension of convertibility in Britain and Ireland effectively became one and the same. What made the Irish experience especially interesting was that Great Britain was Ireland's largest trading partner and that the Bank of Ireland was able to maintain its own monetary policy. Thus, the Irish pound was not only free to float relative to other countries, but was also free to float relative to British pound. As a result, the focus of this debate was primarily on the behavior of the exchange rate between the Irish and the British pound. As in Britain, the depreciation of the Irish pound was blamed on an excess supply of bank notes. The Bank of Ireland, however, argued that its policies did not affect the exchange rate and that any depreciation in the Irish pound was due to changes in the remittances between Ireland and Britain.

The Swedish bullionist controversy actually predated the British and Irish controversies, but has received considerably less attention. Unlike the British case, the

Swedish example has an important and overtly political element. The controversy took place during the period known as "Frihetstiden," or "Age of Freedom." During this time in Sweden the monarchy was so weak that the political power was centered in the Riksdag. The Riksdag was a governing body that was composed of four groups: peasants, clergy, nobles, and burghers. The business class tended to dominate among the four groups in terms of the exercise of power. Two main political parties, known as the Hats and the Caps, battled for control of the government during this period. It is the political differences between these groups that gave rise to the Swedish bullionist controversy.

The Hats were the majority party in the Riksdag from 1738 to 1765. The Hats were largely the "commercial capitalists whose wealth was derived from foreign trade," whereas the Caps were "small industrialists, merchants, and importers" (Eagly 1969, p. 752). The Hats primarily advocated mercantilist policies aimed at increasing domestic production and economic growth. As part of this plan, the Hats used the Riksens Ständers Bank (RSB), a bank established by parliament in 1668. The name meant "Bank of the Estates to the Realm." The bank remains the central bank of Sweden but is now known as the Sveriges Riksbank. In 1739, the Hats began to use the bank to provide loans to Swedish firms in an effort to promote commerce. When this process began, Sweden was on a copper standard. While the bank initially maintained convertibility, the war with Russian in the early 1740s as well as the growing provision of bank notes from commercial loans ultimately led to a suspension of convertibility in 1745.

From 1745 to 1764, there was a significant increase in the supply of bank notes, and the price level more than doubled (Eagly 1969). The Caps opposed the mercantilist policies of the Hats and argued that the increase in the price level and the corresponding depreciation in the exchange rate were the result of an excess supply of bank notes. The Caps advocated a deflationary policy to return the exchange rate and the price level to the levels prior to the increase in the supply of inconvertible paper money. The Hats, however, thought that there was a (seemingly permanent) causal relationship between the money supply and real output. In addition, they blamed inflation on the balance of payments deficits and argued that increasing the money supply would both increase output and correct the balance of payments (Eagly 1969). The Caps were therefore akin to the Bullionists and the Hats to the Antibullionists.

By 1765, rising prices and a depreciation of the exchange rate ultimately led to the Caps becoming the majority party and the Hats being the lead opposition party. When the Caps took power, they adopted a deflationary policy in which they sought to reduce the supply of bank notes in order to restore the price level and the exchange rate to the pre-suspension levels. In 1766 they started reducing the supply of bank notes, and the price level declined and the exchange rate appreciated. However, the policy was costly because it significantly reduced output and increased unemployment. The Caps were once again driven from power. Ultimately, a coup d'etat restored the monarchy in 1772 and a silver standard was adopted in 1776 (Edvinsson and Ögren 2014).

Although the British and Irish controversy and the Swedish controversy arose under different circumstances, each of these controversies represent some of the earliest debates about the determination of the price level and the exchange rate under a paper money standard. In each case, one side took a distinctly monetarist view, arguing that inflation and exchange rate depreciation was due to an excess supply of money. Also, in each case, the opposing side denied that the money supply was an important factor in the rising prices and the depreciation of the exchange rate. Examining the empirical evidence is of some importance. Despite the significance of these debates, there is a relatively small empirical literature on these controversies, especially those that use modern multivariate time series analysis. In the next section, I summarize the existing evidence.

Evidence

Much of the early literature on the bullionist controversies is limited to subjective interpretation of available data in tables and charts. Among the subjective interpretations, the earliest of this literature is the contemporaneous literature itself. In most cases, there is a considerable gap between those writing contemporaneously and the subsequent literature. Regardless, knowing what we now know about time series analysis, this early literature is limited and subject to bias, especially among the contemporary writers. More recently, economists have used modern multivariate time series analysis to reexamine these questions. While this literature remains small, this more recent research provides a better assessment of the evidence in light of each debate. I summarize the literature for each country below.

Great Britain

In the very early stages of the Bank Restriction Period, Lord King (1803, p. 33) wrote that "variations in the price of bullion and in the state of exchange since the suspension have corresponded in a very remarkable degree with the variations in the quantities of Bank notes." This finding was corroborated most notably David Ricardo (1810–1811), and also by Galton (1813).

The literature on the British controversy was subsequently revived by Silberling (1924a, b). He does not find support for the Bullionist position. In addition, he argues that the evidence seems to suggest that Bank of England reduced private discounting to offset advances to the government. However, he makes this claim from time series evidence and therefore lacks a proper counterfactual. Angell (1926) suggests that Silberling's variable measuring foreign payments seems to predict the exchange rate. Viner (1937) similarly argues that, based on Silberling's data, there is a reason to believe that large foreign payments at least played some role in the depreciation of the exchange rate. These authors view this as a refutation of the Bullionist positions, but

this is only true of the most dogmatic of the Bullionists, like Ricardo. Henry Thornton (1802) acknowledged in his book that foreign remittances could be a factor.

Nachane and Hatekar (1995) are the first to examine the British controversy using modern multivariate time series analysis. They use annual data from 1802 to 1838. They find evidence that the data have unit roots. As a result, they test for cointegration using a variety of bivariate models. They only find evidence of cointegration between wholesale prices and the ratio of the money supply to income. This actually seems to support the Bullionist position, but they do not acknowledge it as such. The authors then perform Granger causality tests to determine whether the lags of a particular variable help to predict another variable. They find evidence that the balance of trade and the exchange rate both Granger-cause the wholesale price index. These findings support the Antibullionist position with little support for the Bullionist position.

There is reason to be skeptical of these results. First, if there are differences in the determination of the price level across monetary regimes, then it would seem to be of some benefit to only consider data from the Bank Restriction Period. For example, Glasner (1985) argues that, contrary to conventional wisdom, there is a distinction between classical monetary theory and the quantity theory of money. One lesson that we can learn from many of the controversies – including the bullionist controversies – is that it is important to know when to apply the classical theory and the quantity theory. If the classical theory is appropriate under convertibility and the quantity theory is appropriate for a paper money standard, then the failure to limit the analysis to a sample in which one or the other regime is in place might lead to inconclusive results. The choice of sample size seems to have been motivated by their use of annual data. However, quarterly data is available. Also, the price index that Nachane and Hatekar use comes from Silberling (1923), which Gayer et al. (1953, p. 463–467) found to be flawed. Finally, the lack of cointegation might be due to a lack of a long-run equilibrium relationship or it might be due to incorrectly concluding the data have a unit root.

Officer (2000) improves on the prior research in a few important respects. First, he limits his analysis to the Bank Restriction Period. Second, he uses quarterly data in his analysis. And third, he uses the Gayer et al. (1953) price index that corrects for the deficiencies in Silberling's index. Officer uses Granger causality tests and a structural vector autoregression model to test the hypotheses of the Bullionists and Antibullionists. The Granger causality tests indicate that the supply of bank notes predict the price level and the exchange rate. However, the tests also indicate that the price level and the price of wheat (used to proxy for a supply shock) Granger-cause the supply of bank notes. Officer concludes that this evidence is mixed because it supports both the Bullionist and Antibullionist positions. Officer uses Granger causality tests to inform the causal ordering in the vector autoregression model. He finds that shocks to the supply of bank notes do not have a statistically significant effect on the price level. In fact the only shock that has a statistically significant effect on the price level is the supply shock. Officer therefore concludes that the evidence is mixed, but favors the Antibullionist position.

Hendrickson (2018) constructs a structural model that attempts to capture the key features of the British monetary system. He then estimates the structural model using Bayesian techniques. A key feature of his model is that he allows for the supply of bank notes to respond to government spending through a feedback rule. In estimating the parameters of the model, he finds that government spending had an effect on the supply of bank notes thereby implying that at least some fraction of the government expenditures were paid for through money creation. This contrasts with earlier work by Silberling (1924a, b) who claims that the Bank of England showed remarkable constraint in its issuances. In addition, Hendrickson estimates impulse response functions and finds that a shock to the supply of bank notes does lead to an increase in the price level and a depreciation of the exchange rate. Finally, he simulates the model to generate artificial time series data. He estimates Granger causality tests using this data and the actual British data and finds that one cannot reject the hypothesis that bank notes do not Granger-cause the price level. Using this evidence, Hendrickson argues that this suggests that Granger causality tests are potentially not useful in this context. The reason is that the strongest correlation between bank notes and the price level is the contemporaneous correlation. If the lagged effects are weaker than the contemporaneous effects then Granger causality tests might lead one to incorrectly conclude that bank notes are not important in explaining the price level. This high contemporaneous correlation is also consistent with Jonung's (1976) work on Sweden discussed below.

Finally, Antipa (2017) posits an altogether different explanation for the evolution of prices during the Restriction Period. Whereas the arguments between the Bullionists and Antibullionists were largely about whether the change in the price level and the exchange rate were due to monetary or nonmonetary factors, Antipa suggests that the evolution of prices can be explained by expectations regarding fiscal policy. In particular, she argues that the suspension of convertibility shifted sovereign default risk to a risk of inflation. To examine this hypothesis, Antipa tests for structural breaks in the agio, or the difference between the market price and the mint price of gold. She finds that breaks tend to coincide with significant political and military events. Given the high correlation between the agio and the price level, she argues that these fiscal events are of primary importance for explaining the behavior of the prices during this period. One difficulty is interpreting these results in light of the debate and other evidence. For example, her identification scheme seems to point to the distinct effects of political and military events independent of monetary expansion. However, a quantity theorist might argue that this does not rule out a monetary explanation. For example, if the government is paying for some fraction of spending with money creation (as Hendrickson's results suggest) and if individuals have rational expectations, then the agio as well as the prices of other goods might rise in anticipation of future monetization. Settling whether this is a quantity theoretic or fiscal explanation is clearly beyond the scope of this chapter, but this is an interesting area for future research.

Ireland

Lord King (1803) did not limit his analysis to Great Britain. He found that while the supply of bank notes issued by the Bank of England had increased by about 50% at the time of his writing, the notes of the Bank of Ireland had increased to "*four times the amount of notes in circulation at the time when the Act of Suspension was passed*" (Lord King 1803, p. 48) (the emphasis is in the original.). He argued that this difference explained why the Irish pound had depreciated more than the British pound. Parnell (1804) made much the same argument as Lord King, pointing out that the supply of bank notes issued by the Bank of Ireland had increased by a much greater percentage than the supply of bank notes supplied by the Bank of England.

Foster (1804) similarly argued that the depreciation in the Irish pound was caused by an excess supply of bank notes. He argued that the rising price of bullion and the depreciation of the Irish pound in foreign exchange markets coincided with a progressive increase in the supply of bank notes.

The depreciation of the Irish pound led to the creation of a parliamentary committee by the House of Commons to study its causes. The 1804 Report of the Select Committee on the Irish Currency (the "Irish Pound Report") was the resulting document produced by the committee. The report found that the depreciation of the Irish pound was due to an excess supply of notes produced by the Bank of Ireland relative to the supply of Bank of England notes. However, Fetter (1955) suggests that their evidence was more analytical than it was statistical.

Ó Gráda (1991) also expresses some doubt about this early evidence and the Irish Pound Report. In particular, he argues that the Report was too quick to dismiss the influence of real factors in the short term. Perhaps most importantly, however, Ó Gráda (1991, p. 9) suggests that the focus on the Irish supply of bank notes can be somewhat misleading since the Irish banking system was "almost nonexistent" before this period. As a result, following the suspension of convertibility, the supply of bank notes had to rise to substitute for the withdrawn specie. In fact, Ó Gráda (1993) presents the only evidence using modern time series analysis. Specifically, he tests the monetary explanation of the depreciation in the exchange rate by testing to see whether the exchange rate is cointegrated with the ratio of the supply of bank notes. He fails to find evidence of cointegration. While he notes that this does not necessarily rule out the monetary/bullionist explanation, it likely suggests that there were other important factors that played a role such that the simple monetary model insufficient. Indeed, if his earlier argument that the increase in bank notes simply substituted for the reduction in specie is true, then this would bias his empirical model against finding cointegration even if the quantity theory was strictly true.

Sweden

There is comparatively less that has been written on the Swedish controversy. The earliest evidence was presented by Pehr Nicholas Christiernin (1761). In his work, Christiernin presented evidence that the inflation in domestic prices and the

depreciation of the exchange rate were due to an excess supply of bank notes. However, contrary to his later British counterparts, he did not call for a subsequent deflation because he thought that it would lead to a significant reduction in economic activity (Eagly 1963). Despite being written prior to the British and Irish controversies, Christiernin's work was not referenced in those debates. This is likely because his work was not available in English at that time.

Empirical work on the Swedish controversy was not revived until Eagly (1969). He examines the data informally through the use of charts and tables and finds that prices and the depreciation of the exchange rate seem to follow similar patterns in the data as the supply of bank notes. Although charts and tables leave interpretations open to subjective interpretation, Eagly does provide evidence that when the Caps took power, they significantly reduced the supply of bank notes and the price level subsequently declined and the exchange rate appreciated. This deliberate reduction in the supply of bank notes gives some indication of the direction of causation.

Myhrman (1976) also examines the relationship between bank notes and the price level and exchange rate by examining charts. He finds that there is a positive relationship between the supply of bank notes and the price level. In doing so, he also notes that "the reverse causation problem is thereby avoided because the first period consisted of a deliberate policy of money supply increase . . . and the second period witnessed a decisive reduction in the money supply. It is almost like a controlled experiment" (Myhrman 1976, p. 187). Similarly, Jonung (1976) examines the relationship between money growth and prices and finds a strong contemporaneous correlation during this period.

The only empirical work on the Swedish bullionist controversy that uses modern multivariate time series techniques is Herger (2017). He uses a structural vector autogression model with three separate causal orderings the variables. He finds evidence that the supply of bank notes increased the price level and depreciated the exchange rate. Similarly, he finds little evidence to support the Hats' position that the nonmonetary factors explained inflation and the depreciation of the exchange rate.

Conclusion

The bullionist controversies in Great Britain, Ireland, and Sweden are of some significance because they represent some of the earliest debates about the determination of the price level and the exchange rate under a paper money standard. The debates therefore provide the potential to learn about the different implications of alternative monetary regimes. Macroeconomics seems to progress through critical thinking and debates about current events. Sound economic reasoning and modeling coupled with careful empirical evidence are the source of that progress.

The literature on the bullionist controversies remain somewhat limited given the importance of the debate. Nonetheless, some conclusions can be drawn from the existing research. It is hard to weight the evidence provided by contemporary scholars and subsequent researchers who relied on subjective interpretations of

tables and charts. These authors deserve credit for gathering and, in some case, creating time series data pertinent to empirical evaluation. However, in this early literature it is hard to distinguish analytical arguments and statistical evidence without the use of modern techniques. There is one exception. The works of Myhrman (1976) and Jonung (1976) on Sweden provide evidence of a strong correlation between the money supply and the price level. The Swedish example is somewhat of a natural experiment in the sense that the increase in the supply of bank notes under the Hats and the subsequent decrease under the Caps were both deliberate and exogenous. The strong correlation between the money supply and the price level is therefore evidence in favor of a bullionist position.

The analysis of the British and Irish controversies using modern techniques is somewhat mixed. The early literature seemed to either (a) not provide support for the Bullionist position, (b) provide support to the Antibullionist position, or (c) find mixed support. However, the most recent evidence presented by Hendrickson (2018) not only provides strong support for the Bullionist position, but also suggests that some of the earlier work might have inadvertently produced false negatives with regard to the Bullionist position. The work of Herger (2017) also provides support for what we would now call a quantity theoretic interpretation of the data and is consistent with earlier work on the Swedish controversy.

More evidence is needed about the Irish experience. It would seem useful to develop a model of currency substitution in order to determine how much of the increase in the supply of Irish pounds was substituting for specie and how much was due to excess issuance on the part of the Bank of Ireland. In addition, it would be useful to see if alternative model specifications can corroborate the recent work that provides strong support for a bullionist position.

References

Antipa PM (2017) How fiscal policy affects prices: Britain's first experience with paper money. J Econ Hist 76(4):1044–1076

Angell JW (1926) The theory of international prices. Harvard University Press, Cambridge

Christiernin PN (1761) Utdrag af Föreläsmingar angaende den i Svea Rike upstigne Wexel-Coursen (summary of lectures on the high price of foreign exchange in Sweden). Tryckt hos Directör Lars Salvius, Stockholm

Eagly RV (1963) Money, employment and prices: a Swedish view, 1761. Q J Econ 77(4):626–636

Eagly RV (1969) Monetary policy and politics in mid-eighteenth century Sweden. J Econ Hist 29(4):739–757

Edvinsson R, Ögren A (2014) Swedish money supply. In: Edvinsson R, Jacobson T, Waldenström D (eds) House prices, stock returns, national accounts, and the Riksbank balance sheet. Ekerlids Förlag, Stockholm, pp 1620–2012

Fetter FW (1955) The Irish pound, 1797–1826. A reprint of the report of the committee of 1804 of the British house of commons on the condition of the Irish currency. George Allen and Unwin Ltd, London

Foster JL (1804) An essay on the principles of commercial exchanges. J. Hatchard, London

Galton ST (1813) A chart, exhibiting the relation between the amount of Bank of England notes in circulation, the rate of foreign exchanges, and the price of gold and silver bullion and of wheat. J. Johnson & Co., London

Gayer AD, Rostow WW, Schwartz AJ (1953) The growth and fluctuation of the British economy, 1790–1850. Clarendon Press, Oxford
Glasner D (1985) A reinterpretation of classical monetary theory. South Econ J 52(1):46–67
Hendrickson JR (2018) The Bullionist controversy: theory and new evidence. J Money Credit Bank 50(1):203–241
Herger, Nils (2017) An empirical assessment of the Swedish bullionist controversy (working paper)
Jonung L (1976) Money and prices in Sweden 1732–1972. Scand J Econ 78(1):40–58
King (Lord) (1803) Thoughts on the effects of the Bank restrictions. Cadell and Davies, London
Laidler, David (2000) Highlights of the bullionist controversy (working papers 2000–2002), Department of Economics University of Western Ontario
Myhrman J (1976) Experiences of flexible exchange rates in earlier periods: theories, evidence, and a new view. Scand J Econ 78(2):169–196
Nachane DM, Hatekar NR (1995) The bullionist controversy: an empirical reappraisal. Manch Sch 63(4):412–425
Gráda CÓ (1991) Reassessing the Irish pound report of 1804. Bull Econ Res 43(1):5–19
Gráda CÓ (1993) The Irish paper pound of 1797 –1820: some cliometrics of the bullionist debate. Oxf Econ Pap 45(1):148–156
Officer LH (2000) The bullionist controversy: a time series analysis. Int J Financ Econ 5:197–209
Parnell H (1804) Observations upon the state of currency in Ireland, and upon the course of exchange between Dublin and London. Mahon, Dublin
Ricardo D (1810-11) The high price of bullion. In: Sraffa P (ed) Works and correspondence of *David Ricardo*. Cambridge University Press, Cambridge
Silberling NJ (1923) British prices and business cycles, 1779–1850. Rev Econ Stat 5(Suppl 2):223–247
Silberling NJ (1924a) Financial and monetary policy of great Britain during the Napoleonic wars. Q J Econ 38(2):214–233
Silberling NJ (1924b) Financial and monetary policy of great Britain during the Napoleonic wars. Q J Econ 38(3):397–439
Thornton H (1802 [1807]) An enquiry into the nature and effects of the paper credit of great Britain. James Humphreys, Philadelphia
Viner J (1937) Studies in the theory of international trade. Harper & Brothers, New York

Money in Wars

39

Kenneth Mouré

Contents

Abstract

Money, in Cicero's oft-cited phrase, provides "the sinews of war." But vital as money is to making war, it is a means, not an end, and wars produce great disruption to monetary order. Wars provoke monetary crises and create opportunity for monetary reforms. The increasing intensity and cost of "total war" in the twentieth century and the development of advanced, consumer economies in which monetary transactions are diffused throughout the economy have given monetary policy great importance for military power and for potential monetary disaster. This chapter surveys the relationship between the uses of money in wartime and monetary order, focusing on the two world wars as crises in which the power of money had very different effects in wartime and in postwar consequences.

Keywords

Money · Taxes · Fiscal state · Inflation · Fiat currency

K. Mouré (✉)
History and Classics, University of Alberta, Edmonton, AB, Canada
e-mail: moure@ualberta.ca

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_39

Introduction

Wars have produced greater disruption to monetary order than any other events in terms of the speed, violence, and extent of changes in domestic prices, international trade, and monetary stability. How states develop their resources for use in war and finance their expenditures influence not only the power and sustainability of their mobilization for war, but also their economic strength and stability when the war is over. Money, in Cicero's oft-quoted phrase, provides "the sinews of war." As such, money can be the determining factor in a country's ability to mobilize economic and military power. But once war is underway, the unpredictability of the war's duration and outcome means that monetary policies are subordinate to political and military goals. How to pay for the war is less important than winning, or at least avoiding defeat. The violence of wartime economic changes and how they are to be repaired or reversed are issues often left for when the war is over.

Paying war costs can inflict serious damage on currency values and on international payment systems. Financing military expenditure has been a major reason for the development of national banking and taxation systems. Undoing the damage of war finance has provided impetus for the restructuring of domestic monetary regimes and the development of new international arrangements to provide greater currency stability. The monetary dimensions of war experience are inseparably linked to war finance. This chapter will focus on the challenges posed by war and war finance with regard to domestic monetary stability, international monetary regimes, and the monetary needs of producers and consumers. After an overview of the relationship between war and monetary crises, the chapter will concentrate on monetary experience in the world wars in the twentieth century. In these wars, the importance of financial infrastructure and monetary policy is apparent in the development of national banks of issue, the evolution of modern central banking, and the increased civilian needs for currency in economies of mass consumption. The mobilization for "total war" – war in which the separation between military and civilian life was broken down by national economic mobilization and the targeting of civilians – made monetary policy and war finance matters that impacted all levels of society and disrupted a globalized world of international trade and finance.

Long-Term Overview

War requires money. In his *History of the Peloponnesian War*, Thucydides provides important reflections on money and power in time of war: the importance of money and finance in establishing Athenian naval power, and the subsequent "corrosion" of that power when increasing wealth was employed for private gain rather than public benefit (Kallet-Marx 1993; Kallet 2001). The continuing appeal of Cicero's phrase that "limitless money forms the sinews of war" (*nervos belli, pecuniam infinitam*: Philippics, V, 2, 5) lies in his evocative and timeless linking of military strength to fiscal capacity. Whether raising domestic armies, hiring mercenaries, manufacturing

or buying arms, or supplying forces in the field, financial power is essential to military success.

Machiavelli, writing in the early sixteenth century, an era of unreliable mercenary armies, disagreed: "it is not gold, as is vulgarly supposed, that is the sinews of war, but good soldiers; for while gold by itself will not gain you good soldiers, good soldiers may readily get you gold" (Machiavelli, *The Discourses*, 1517, Chap. 10). In his *Art of War* (1521), Machiavelli provides general rules of war in which he identifies four "sinews of war": "Men, iron, money, and bread are the sinews of war; but of these four, the first two are more necessary; for men and iron find money and bread, but bread and money do not find men and iron." (Machiavelli 2003, 159. His next rule: "The unarmed rich man is the reward of the poor soldier.") Machiavelli addressed the military problems of his day, a period in which the relationship between state financial capacity and military power was volatile, particularly in the Italian peninsula and in Central Europe. It was an era of "domain states" where warfare conducted by princes was a matter of personal sovereignty and vassals' obligations, drawing revenue from land, often through payments in kind.

Joseph Schumpeter contrasted this type of "domain state" finance, its irregularity and its breakdown in the face of recurrent fiscal crises linked to war costs, with the rise of the "tax state" in the early modern period (Schumpeter 1954). The "tax state" developed methods to collect revenue from a wider array of sources linked to economic transactions and the growth of money-based economic activities. This revenue was more reliable because it was more regular and the object of systematic state administration. Unpredictable and irregular war expenditures could then be met by a combination of loans and, critically, taxation to cover immediate costs and loan repayment. The development of nation states and state administration took place in conjunction with rising military costs from innovations in technologies of war, frequent wars, and the greater reliability of standing armies rather than mercenaries to maintain internal order and protect against external threats (Bonney 1995, 1999; Ormrod et al. 1999; Tilly 1992).

As "national" states developed from the sustained violence of interstate warfare in early modern Europe, military expenditure often constituted nearly half of state budgets in peacetime and up to 90% in wartime. The emergence of European "great powers" in the early modern period depended on the state's capacity to raise money through a combination of efficient taxation and stable, long-term borrowing (Kennedy 1988; Parker 1988). State capacity to manage a system of "raising and *simultaneously* spending vast sums of money acted like a bellows, fanning the development of western capitalism and of the nation-state itself" (Kennedy 1988, 100). Most states were "fiscally fragmented." Monarchs depended on local and regional authorities to whom they granted privileges in exchange for the collection and payment of tax revenues. Local authorities and aristocratic elites found their political autonomy threatened by fiscal reforms and new taxes, and could put up substantial resistance to the centralization of tax policy (the French Revolution provides a spectacular example). The centralized states able to impose uniform national tax policies collected higher per capita revenues, particularly when taxpayers with representation acted as a check on royal power. Their ability to restrict the

military spending that dominated state budgets also legitimized tax collection (Dincecco 2009, 2015).

British experience was precocious. As an island state, Britain faced lower threats of external attack and kept apart from Continental power struggles. Innovations in indirect taxation to avoid the opposition of wealthy elites and improved legal and administrative structures for collecting taxes provided more consistent and sustained revenue from a growing domestic economy (O'Brien 2011). Britain's centralized fiscal-military state after 1688 won great power status not on the basis of population size, geographic reach, or resource wealth, but rather by building a strong foundation for imperial power financed by efficient taxation, strong state credit to allow low-cost long-term borrowing and repayment, and a stable and internationally trusted currency. The creation of the Bank of England in 1694 assisted in establishing a system for reliable and low-cost domestic borrowing that "enabled England to spend on war out of all proportion to its tax revenue" in the eighteenth century (Dickson 1967, quote p. 9; Brewer 1989). Britain's ability to tap current and future revenue, rather than needing to debase its coinage or repudiate state debt, helped rationalize and stabilize government finance, which in turn facilitated economic and industrial growth.

In contrast to the stability of British financial and monetary institutions, in France the inefficiency and corruption in tax collection and excessive borrowing for war expenditures crippled state finance and finally generated revolution when Louis XVI convened the Estates General to reform the system of taxes and privileges in 1789. Poor financial management had burdened France with excessive debt, a tax system yielding low and unreliable revenues, debasement of the coinage, and the issue of paper currency in the last decades of Louis XIV's rule (Rowlands 2012). The problems were not insoluble. The conjunction of a political crisis over representations of interest and protection of privilege, a major subsistence crisis, and popular fears of violence and repression by the privileged orders sparked political revolution in combination with popular insurrection. Tax and tithe payments stopped in many regions awaiting major reform, and elected deputies formed factions to press for or forestall change. Customs duties and a direct land tax instituted in 1789 to rationalize the tax system failed to produce sufficient revenue to cover debt repayments, and the state seized Church lands in order to sell them at public auction.

Assignats, a paper currency guaranteed in value because they were "assigned" to specific nationalized Church properties, provided new funds to the Treasury beginning in December 1789. Initially issued as large-denomination bonds paying interest, in April 1790 they became a paper currency and soon demonstrated the dangerous potential for abuse in the quantity of notes being issued in ever-smaller denominations. Although the product of domestic revolution rather than war, the assignats were issued in a comparable atmosphere of crisis that included capital flight, hoarding, and rising state need for revenue. They were first intended to meet the need for money "to replace promptly that which has disappeared from commerce" (Crouzet 1993, 146). But war costs after April 1792 produced inflation and then, in 1795–1796, the first "modern" hyperinflation. In February 1796, the assignats, now worthless, were burned along with their printing plates in the Place

Vendôme. Their replacement, a new fiat currency, depreciated to 6% of its plate value by mid-summer. In 1797, as Britain took the pound off gold, France returned to a metal-backed currency (Harris 1930; Bordo and White 1991; Crouzet 1993; Brezis and Crouzet 1995).

Bordo and White compare British and French policies for war finance during the Napoleonic Wars and find that Britain's ability to repay borrowed money gave it much greater freedom of maneuver and financial strength. British war costs were covered by borrowing, while French war finance was restricted to increasing taxation and restricted borrowing at higher cost (Bordo and White 1991). A recent assessment of British policy in its period of fiat currency from February 1797 to May 1821, looking at how contemporaries evaluated the likelihood of debt repayment or reduction through currency devaluation, concludes that the changes in British prospects for military victory based on news from the Continent had a profound influence on interest rates and exchange rates (Antipa 2016). The British success in its war effort against Napoleon, which generated a higher debt-to-GDP ratio in Britain than in either the First or the Second World War, was based on money creation and borrowing, and on substantial tax increases that covered more than half of the wars' costs to 1809. The institutional support to manage and limit state borrowing for war purposes facilitated lower-cost borrowing and public confidence that currency value would be restored and debts repaid. But the outcome of the wars mattered too; default was far more likely for losers than winners (Antipa 2016; Sargent and Velde 1995).

The consequences of fiat currency issued in the US Civil War again demonstrated the dangers of inflation and the loss of confidence when governments issue paper currency to finance war expenditures (Mitchell 1903; Sharkey 1959). In the debate on the Legal Tender Act to authorize the Treasury issue of currency notes in 1862, Secretary of the Treasury Salmon Portland Chase declared that these notes would not be allowed to become "an irredeemable paper currency." He stated "no more certainly fatal expedient for impoverishing the masses and discrediting the government of any country can well be devised" (Sharkey 1959, 20). The "greenbacks" issued by the Union eventually totaled $431 million and contributed to a doubling of prices in the North during the war. They were a step, along with the creation of a National Banking System for consistency in the currency issue, towards the development of a single national currency and a national bank of issue (the Federal Reserve System would be established in 1913) (Sharkey 1959; Friedman and Schwartz 1963). The Confederacy covered more of its war costs by issuing no-interest Treasury notes that functioned as currency, although they were not made legal tender (it also issued interest-bearing notes). Their effect on prices and social welfare was ruinous. Prices in the Confederacy increased 27-fold, and Confederate states had to increase their relief payments to cope with desperate shortages and food riots (Hummel 2014). But currency issue alone did not decide the Confederate inflation. In both the North and South, several studies have argued that prices also moved in response to news from the battlefields, with battle outcomes and prospects for victory influencing the probability of eventual debt redemption in gold (Roll 1972; Burdekin and Wiedenmier 2001).

In the aftermath of the US Civil War and the Franco-Prussian War of 1870–1871, the desire for a return to "sound currency" encouraged development of the international gold standard, in which fixing the gold value of national currency became the norm for responsibly administered states. Britain had established a fixed rate to gold when it restored convertibility of the pound in 1821. Germany, which had been on a silver standard, based its new currency for the unified Reich on gold convertibility in 1872 and engaged in massive gold purchases to build its metallic reserve. These purchases were possible thanks to the large indemnity (5 billion francs) imposed on – and rapidly paid by – defeated France. The US Coinage Act of 1873 discontinued the free coinage of silver there. France and Belgium limited their coinage of 5-franc silver pieces in 1873 (both countries were inundated with silver and losing gold) in an effort to preserve bimetallism. But the unintended results were the further depreciation of silver in relation to gold, and the steady abandonment of silver as a metallic reserve by economic powers who adopted the gold standard. The issue caused political controversy in the United States in the 1870s, until convertibility to gold was restored in 1879, and again in the presidential election of 1896. Democratic candidate William Jennings Bryan told the Democratic National Convention in 1896 that, in reply to the arguments in favor of the gold standard, his party's response should be: "you shall not press down upon the brow of labor this crown of thorns. You shall not crucify mankind upon a cross of gold" (on France Flandreau 1996; on the US, Unger 1964; Friedman 1990a; Bensel 2000). The monetary disorders produced by wars between 1848 and 1871, the prestige and power associated with Britain's financial supremacy as a gold standard power, and the belief that gold convertibility would provide greater stability and limit discretionary monetary expansion promoted the spread of gold standard belief and government commitments to gold convertibility at fixed rates. The era became that of the "classical gold standard" from 1880 to 1914 (Gallarotti 1995).

The consolidation of the gold standard as an international monetary regime, the rapid growth of world trade and international finance, and the development of consumer markets and domestic transactions using currency notes provided a combination of financial and monetary forces that would create greater challenges and greater opportunities for the use and abuse of money in time of war when conflict erupted in the summer of 1914. The gold standard commitment to currency convertibility as a check on monetary expansion turned out to be a fair weather commitment, suspended to meet the urgent demands of war.

World War I

Expectations of imminent war set off an avalanche of sales on financial markets in late July 1914 as stock and bondholders rushed to obtain cash. Contemporaries described the events as a "breaking storm" or a "bolt from the blue." For Hartley Withers, a famous British financial journalist, "the fury of the tempest was such that no credit system could possibly have stood up against it" (Withers 1918, 3). The crisis seized all countries involved in international trade and finance, belligerents,

and neutrals alike. It centered in London as the world's preeminent financial market and the nerve center for international payments. Two changes in monetary demand impacted national and international markets: the demand for domestic currency to meet uncertain future needs (Withers commented "When war comes or is expected, the public is apt to lose its head and do things that are stupid and even wicked, under the influence of unreasoning fear, and because they think something dreadful is going to happen") (Withers 1918, 7) and the calling home of liquid assets and international credits. The sales of securities and the calling in of short-term loans created paralysis: the system froze. Apprehension struck markets in the last week of July, with panicked sales leading to the closing of stock markets across central Europe, then in London on 28 July, rapidly followed by the exchanges in New York, Tokyo, and Australia (Roberts 2013). There was a rush to sell securities and to convert bank notes to gold. The gold convertibility of major currencies was suspended, either formally (Russia on 27 July, Germany on 31 July, France on 5 August) or de facto (Britain) (Hardach 1977, 140). Foreign exchange markets ceased to function and most countries imposed debt moratoria, as the means of payment for many debtors were either frozen or unobtainable.

In London, banks paid out currency notes instead of gold, expecting a suspension of convertibility. Anxious Londoners queued in Threadneedle Street to obtain gold sovereigns from the Bank of England. After intense discussion, the government decided to maintain gold convertibility. John Maynard Keynes played a key role in convincing the Chancellor that the bankers' case for suspension of the gold standard was exaggerated and would damage London's future position as a gold market (Morgan 1952; Sayers 1976; de Cecco 1984; Roberts 2013). Although the gold content of the sovereign remained unchanged, the Bank of England purchased all gold produced in the Empire, and in 1916 gold importers were obliged to sell to the Bank at its statutory buying price, restricting gold imports (Brown 1940).

Youssef Cassis notes in his summary of the global financial crisis in July 1914 that the crisis was soon forgotten. It was the war itself that "changed the banking world and its broader economic and financial environment far more profoundly than any crisis could ever have done" (Cassis 2011, 22). It did so through the duration and cost of the war, and the methods improvised to cover war costs. No country had laid financial plans beyond meeting the costs of initial mobilization. A military "cult of the offensive" posited a short war to be won by superior offence, and left governments ill-prepared for the total mobilization, the massive consumption of resources, and the escalating financial costs the war would entail. The "short war illusion" also encouraged the concomitant belief that the financial measures needed for mobilization in 1914 and to sustain war costs thereafter would be temporary. Suspended obligations would be resumed after a brief interruption. Debts would be repaid, gold convertibility would be restored, currencies would return to their prewar parities. Immediate measures voted to cover mobilization costs in August 1914 were likewise short term. Planning to meet the extraordinary demand for war supplies, armaments, and manpower underwent necessary revision in 1915, but ad hoc financial planning prevailed in a situation where hopes for an end to the war were repeatedly disappointed (Horn 2002). The financial demands created ever-increasing

Table 1 Monetary impact of World War I (index 1913 = 100)

Country	Real GDP 1918	Currency in circulation 1920	Wholesale prices 1919
France	63.9	674	357
Germany	81.8	2678	415
Italy	111	800	821
Russia	67.7 (1917)	1061 (1917)	953 (1918)
United Kingdom	114.8	168	242
USA	113.2	201	206

Sources:
Real GDP: Broadberry and Harrison (2005), Table 1.4, 12; for Italy, Mitchell (1975)
Currency in circulation: for Europe, Kindleberger (1993), 287; for the USA, Friedman and Schwartz (1963), Table A-1, cash held by public plus demand deposits, indexed to July 1914. Russia is July 1914 = 100; note circulation on 1 October 1917; Gatrell (2005), 145
Prices: Hardach (1977), 172; Russia from Gatrell (2005), 146

pressure for greater state organization of the procurement of resources needed for war and required innovations in how the costs would be paid – through taxation, borrowing domestically and abroad, and by increasing the monetary circulation. It proved "fatally easy to create money out of nothing by borrowing from banks and issuing paper money" (Hawtrey 1931, 87). The resulting changes in government spending, monetary circulation and retail prices for major belligerents are shown in Table 1.

Belligerents covered the war credits for mobilization voted in August with short-term expedients: advances from central banks, increased currency issue without the backing of gold reserves, and the issue of currency notes by Treasuries and commercial banks. Some countries created special budgets for war expenditure, knowing they could not cover it with revenue receipts during the war. In Germany's case, war spending was placed in an "extraordinary budget" to be covered by short-term borrowing with no wartime plan for how debt repayment would be managed at the end of the war (Feldman 1993; Hardach 2015). In most countries, substantial efforts to develop longer-term financial strategies with less inflationary consequences began only in 1915 and relied heavily on borrowing. Britain, the most financially responsible of the European powers, allowed the Bank of England to exceed the legal limit on fiduciary issue, authorized the Treasury to issue low denomination currency notes and sold Treasury bills on the open market. In 1914, authorities feared a general economic collapse caused by the disruption to trade, and hoped for a return to "business as usual" (Daunton 2002). But the government also issued medium-term loans and raised taxes in November 1914 (Morgan 1952). In 1915, Treasury views allowed for some inflation to reduce civilian consumption, as well as tax measures that included increased income tax rates, duties on imported luxury goods, and an excess profits duty and increased Allied borrowing (Daunton 2002). France and Germany relied more heavily on short-term debt and increasing their note issue, adding domestic borrowing and the taxation of incomes and war profits in 1916 to restrain the growing pressure on prices. Germany had to rely on short-term domestic borrowing, as it had no wealthier allies. This debt was monetized, creating a monetary

Table 2 Postwar currency depreciation

Country	de facto stabilization	% of prewar parity
Germany	1923	0.0000000001
Austria	1922	0.00007
Hungary	1924	0.0069
Belgium	1926	14.5
France	1926	20.3
Italy	1926	27.3
Sweden	1922	100
Netherlands	1924	100
United Kingdom	1925	100
Denmark	1926	100

Source: The European Economy Between the Wars (1997) by Feinstein et al. Table 3.2 on pg. 46 [adapted]. By permission of Oxford University Press

overhang that would fuel postwar inflation. German taxation and borrowing were less successful than in Britain, the United States, or France, leading to greater monetization of debt and higher inflation (Balderston 1989; Holtfrerich 1986; Feldman 1993). Italy, too, faced limited domestic borrowing possibilities and relied on foreign loans from its allies and inflation to finance the war effort. In 1917–1918, Italian state revenue was barely 15% of state expenditure (Forsyth 1993; Fratianni and Spinelli 1997, 112). By November 1918, all belligerents had significantly increased their note circulation and government debt, making further inflation inescapable in the postwar period. Postwar inflation and currency depreciation varied with the extent of monetary overhang, as well as with failures in domestic financial management. These produced currency depreciations and in some cases hyperinflation and destruction of the currency, especially in Germany and Hungary (Strachan 2004; Ritschl 2005; Schulze 2005) (Table 2).

Foreign borrowing played a vital role in covering war costs, particularly European imports of food and war materiel from the United States. The United States benefited from the demand for goods and finance, turning from an international debtor to a major creditor, with a substantial surge in production and profit to industry and agriculture (Hardach 1977, 256–258). In recession when the war began, the United States experienced a sustained "peacetime" boom from December 1914 to its declaration of war in April 1917 and continued growth thereafter to August 1918. Real GNP rose 26% and non-farm employment increased by nearly 3.3 million (Rockoff 2005, 310–315). The British worked to coordinate Allied demands and channel purchasing and payments through London, with gold from France and Russia helping to maintain the critical sterling-dollar exchange rate. Britain's borrowing in the United States had, by the war's end, very nearly balanced its lending to European allies – mainly to Russia, France, and Italy. This balancing act reached a tipping point in July 1917, after the US entry into the war. British financial resources were nearly exhausted. British stress on the burden they carried and the danger of an imminent collapse of Allied purchasing were needed in order to persuade the US authorities to

finance Allied purchasing and support the "Allied exchange rate" for the dollar and the pound (at $4.76/£, slightly below prewar parity). Keynes, the Treasury's key advisor on external finance, had counseled keeping sterling on gold in August 1914 until suspension was "*absolutely* necessary." He now argued that without full US financial support to maintain the exchange rate and Allied purchasing, Britain should suspend convertibility rather than deplete Britain's remaining gold reserves (Burk 1979; 1985). The strength of exports and lending to belligerents brought a steady gold flow to the United States. Their gold reserves increased by 60%, and the US dollar became the standard for measuring currency depreciations and the distance to be closed in returning to "normal" monetary conditions after the war.

Monetary policy, subordinated to war finance, had sought means to cover war expenses without destroying the belief in a return to the "normal" conditions of the prewar gold standard. But increased currency in circulation and rising prices, a concentration of global gold reserves in the United States, and a massive increase in international debt so altered monetary conditions that this "return" was for many countries impossible, and for others unwise (Britain's return to gold in 1925). Interwar efforts to rebuild the gold standard as a gold exchange standard, patched together with a new effort at central bank cooperation, restored tenuous stability in the mid-1920s but were unable to maintain the flawed system after 1928 (Eichengreen 1992; League of Nations 1944; Clarke 1967).

The war disrupted domestic currency use in two ways. First, it increased the need for currency notes: for cash rather than credit transactions, the purchasing for war goods, the payment of troops, and hoarding of currency as a precautionary measure. Most European currency systems used coinage based on gold and silver, with part of the national metallic reserve circulating as coin (unlike the interwar "bullion" system with gold exchanged for notes only as ingots). Central banks withdrew gold and silver coins from circulation to augment their gold reserves. States issued paper currency to meet increased transaction needs. In Britain, the government authorized Treasury notes for this purpose. The Bank of France had printed and distributed 1.5 billion francs in 5 and 20 franc paper notes to replace coin in event of war and added 10 franc notes in 1916 (Blancheton 2001). New currency, printed on short notice and poorly designed, proved easy to counterfeit. The British Treasury printed the first "Bradburys" on ungummed postage stamp paper; the note looked like "a ticket for a cloakroom" (Clapham 1946, 38; Roberts 2013, 124). The need for low-denomination currency for daily purchases could be acute in regions invaded by the enemy, where the military demands for goods, services, and indemnities increased cash purchasing. In occupied northern France, a range of municipalities, savings banks, mining companies, and metal industries issued "monnaies de nécessité" – small denomination notes and tokens for daily transactions (Eck and Touchelay 2018). Throughout France, chambers of commerce met local needs by issuing their own currency bonds and tokens, a necessity in some regions until 1920 (Druell-Korn 2015). Germany and Austria-Hungary likewise authorized the issue of low-denomination notes to meet the need for currency for ordinary daily purchases (Strachan 2004, 30–32).

In most cases, these new low-denomination issues contributed only marginally to the inflationary increases in monetary circulation. But they did alter ordinary consumer practices. Queues for goods and state and local measures to ration food supplies became common practice by 1917. Consumers returned to earlier practices of direct purchase from farms, used family networks for food supplies, and bartered goods for food. The obvious inequalities in purchasing power and access to goods provoked deep resentment in a moral economy of shared civilian sacrifices necessary to win the war. "Equality of sacrifice" was a deceptive principle for rationing: needs varied within and across classes, and sacrifices were re-apportioned within communities, particularly at the family level (Pigou 1921, 136–149). The moral and political dimensions to the necessity of sacrifice prompted popular campaigns for state measures to tax wealth and war profits, prosecute "profiteers," and increase food allocations for the working and middle classes in Berlin and Vienna (Winter and Robert 1997, see essays by Winter, Robert, and Bonzon and Davis; on Germany, Davis 2000; on Austria-Hungary, Healy 2004).

During the war, monetary policy served the needs of total mobilization to win the war or at least to avoid defeat. Reliance on debt and the monetization of domestic debt created the conditions for postwar inflation and currency instability. Temporary measures adopted for the mobilization in 1914 endured. The belief that these measures could and would be reversed and prewar "normalcy" restored guided wartime monetary policy, and proved to be profoundly mistaken. In Britain, where financial policy had been managed relatively well, the Cunliffe Committee [the Committee on Currency and Foreign Exchanges after the War] foresaw a transition in 1918 that would take 10 years to restore the Bank of England's control of monetary policy and return the pound sterling to gold at prewar parity. Central bank cooperation, irregular and focused on domestic goals under the classical gold standard, took on new importance with the development of cooperation between governors of the Bank of England and the Federal Reserve Bank of New York (Clarke 1967; Flandreau 1997; Mouré 2002). Belief in the gold standard as a "knave-proof" regime to regulate monetary policy and prevent inflation took on the character of a religious faith. Britain reshackled the pound with "golden fetters" in April 1925, overvaluing the pound, a measure that would last until September 1931 (Moggridge 1972). Central bankers and policy makers welcomed the fetters in the 1920s, believing them essential to end the chaos of wartime inflation and restore global order. They then found themselves marching lockstep into monetary contraction and global depression after unbalanced growth in the 1920s (Eichengreen 1992; Ahamed 2009).

World War II

The turbulent financial aftermath of World War I, with international efforts to deal with war debts, currency depreciation, and hyperinflation, made obvious the need for better financial planning as a new war became ever more likely in the 1930s. In addition to the postwar financial problems, the Great Depression had made the 1930s a period of prolonged unemployment, budget deficits, and renewed currency

instability as the restored gold exchange standard collapsed. These economic challenges encouraged the turn to authoritarian governments in Europe, delayed rearmament spending in the Western democracies, and promoted belief in the need for a stronger government role to deal with market failures. The Great Depression had a profound impact on national monetary and fiscal policies and on the international regimes for trade and payments (For the US economy see Bordo et al. 1998). In 1939, governments and central banks were acutely aware of the problems generated by war finance in World War I and the unhappy aftermath of hyperinflation, currency instability, and prolonged depression. Abandonment of the gold standard in the 1930s gave governments greater freedom and greater responsibility to set monetary policy in the interests of their national economies. Policy makers struggled to escape the "gold standard ideology" that dominated interwar monetary policy (Eichengreen and Temin 2000). With British financial power weakened by the Great War and its aftermath, and without monetary leadership from the United States as the strongest financial power, there was no international consensus, neither for domestic monetary policy nor for the structure of the international monetary system (Kindleberger 1986). Even the tripartite cooperation between the United States, the United Kingdom, and France after the French devaluation in September 1936 was strained by poor communication and little serious effort to communicate between governments. Day-to-day management did improve thanks to communication between their exchange equalization accounts and the US Treasury's willingness to sell gold to the British and French accounts although neither currency had a fixed gold value.

Rearmament and war planning in the 1930s developed according to national agendas and resources. Without the external constraint of the gold standard, governments had greater domestic policy freedom, and Treasuries had greater influence than central banks over domestic policy. The impact of the Depression was evident in how each country set its priorities. Reserves of under-employed labor and capital could allow new rearmament spending within budget constraints without setting off renewed inflation, but this required increasing state intervention to redirect new earned income away from immediate consumption. In Germany, accelerated rearmament began in 1935 and benefited from a 1933 moratorium on foreign debt payments, exchange controls to prevent devaluation of the Reichsmark, and the willingness of the Reichsbank under Hjalmar Schacht to accommodate Nazi deficit spending (Ellis 1940; Tooze 2006). New rearmament spending by the Popular Front in France was the critical factor forcing France to devalue the franc Poincaré in September 1936 (Frankenstein 1982; Mouré 1991). Japan's military spending was supported by the Bank of Japan. In both Japan and Germany, restrictions on civilian consumption and the loss of labor power to press for higher wages limited the inflationary impact of increasing military expenditure. In Italy, deficit spending and the costs of wars in Ethiopia and Spain resulted in the devaluation of the lire, and the Fascist government removed the Bank of Italy's reserve requirement against paper currency, facilitating the monetization of budget deficits and accelerating inflation. Democratic governments struggled to balance national budgets and were more cautious in committing to military expenditure. Aggression by the Axis powers

from 1935 onward brought increasing need for defense spending, particularly after Munich in 1938, when Germany's drive for war became incontrovertible (Tooze and Martin 2015; Huff 2015; Fratianni and Spinelli 1997).

Europe went to war in 1939 without a shared international standard for currency convertibility and with no illusions that this war would be over by Christmas. Policy makers drew on national experience with financial management in World War I and in dealing with interwar inflation, currency instability, and world depression. But the course of the war was unknown: policy was improvised. The past provided examples of flawed policy rather than a blueprint for managing the new war. German policy since the "Four-Year Plan" for war preparation in 1936 pushed for rearmament and conflict with quick victories; otherwise the problems of finance, monetary stability, and inflation would prove intractable. "In the first year of the war Hitler was going for broke and to the amazement of the world, the gamble paid off" (Tooze and Martin 2015, 41). Italy launched its war in the Balkans in 1940 with poor military and financial preparation, after having delayed joining Germany's war against France until early June 1940 when its outcome was clear. Tight control of money creation to maintain the lira's gold parity had been abandoned in 1936. Price controls and rationing in wartime concealed significant inflationary pressure, with prices rocketing upward in 1943 after Mussolini was removed from power and the Allied invasion began the peninsula's slow liberation. Agricultural and industrial output had fallen steeply since 1939; tax revenue fell while government expenditure increased. The lire exchange rate, pegged at 19 to the dollar in 1939 and unquoted from 1940 to 1943, was reset by Allied authorities at 100 to the dollar in 1943, and stabilized at 575 to the dollar in December 1947 (Fratianni and Spinelli 1997).

Both France and Britain were democratic regimes less willing to run budget deficits, adopt exchange controls, and force labor to accept reduced consumption, and they planned for a defensive war that would require a prolonged mobilization focused on arms production. In France, the Popular Front began serious rearmament in 1936. Greater stability under a center-right coalition in 1938 allowed an increase in armaments output and in purchasing from the United States. But there was no clear strategy to cover war costs, which depended on borrowing and advances from the Bank of France in 1940 (Frankenstein 1982; Imlay 2003). British rearmament was financed by a combination of increased taxes and new borrowing in 1937–1938, with reluctance to impose higher taxes while the economy was in recession, counting instead on an increase in tax revenue as rearmament spending reduced unemployment (Peden 2000).

Once war had been declared, arms purchasing accelerated and British planning for war finance demonstrated the value of lessons learned. A new budget in September 1939 increased taxation to reduce state borrowing needs and civilian consumption (Peden 2000). Keynes published articles in *The Times* in November 1939, revised in book form in February 1940 with the title *How to Pay for the War*. He argued on the grounds of sound financial management and social justice for a system of deferred consumption by "forced savings" to reduce the level of civilian consumption. In peacetime, the amount of goods available depended on the amount of work done. In wartime, maximizing output needed constraints on civilian consumption to

devote increased effort to war goods. Paying for war production with inflation would limit consumption, but it would transfer purchasing power from ordinary consumers to the wealthy. Forced savings would curb wartime consumption and guard some of the earnings of "the wage and salary earner . . . for his future benefit and security, which would belong otherwise to the capitalist class" (Keynes 1940, 5). Although the British government did not adopt the forced savings plan, Keynes became an advisor in the Treasury. Britain's budget in 1941 adopted a national accounting framework in order to determine the amount of new tax revenue needed to prevent inflation. Britain increased income taxes and kept interest rates low, financing a "3% war" without significant inflation (Peden 2000; Daunton 2002).

In the United States, policies were even more successful at managing war finance. Distance from the regions of conflict meant there was little damage to domestic assets and infrastructure. The late US entry into the war after Pearl Harbor had allowed a ramping up of production to supply its democratic friends as "the arsenal of democracy," bringing the long Depression era of underproduction and unemployment to an end. When Japan and Germany declared war on the United States in December 1941, the US economy mobilized for full-scale war production. It drew on abundant natural resources, unequalled and underused productive capacity, and entrepreneurial energy that in combination could outperform all other belligerents in armament capacity. In 1944, the United States committed 45% of its economy to war production. Its production of military aircraft was more than double that of the next producer (the USSR, slightly ahead of Germany); its value of combat munitions output nearly doubled that of Germany and Japan combined (Rockoff 2012). This phenomenal growth, supplying its own armed forces and its allies with arms and food, and the full mobilization of its war economy was accomplished on a sounder basis than in the First World War. A higher proportion of war costs was paid by taxation (49% vs. 29%), in place of borrowing and printing money. Prices, rising in the period 1939–1942, were held in check by a system of price controls through the Office of Price Administration (OPA), and supplemented by powerful citizen support for controls that mobilized millions of women to monitor consumer prices (Galbraith 1952; Rockoff 1984; Jacobs 2005). This deferred some price inflation to the postwar period. Retail prices had been rising by about 10% per year in 1941 and 1942, and this fell below 2% from April 1943 to February 1946, but jumped 15% in the first 18 months after price controls were lifted in June 1946 (Rockoff 2012). But inflation did not cause price conflicts over the prioritization of war goods or reduce workers' consumption during the war.

British and US policy makers drew successfully on the lessons from WWI experience. Neither country suffered invasion and occupation. War finance and monetary policy in totalitarian states used less planning and greater coercion. In Germany, initial military successes helped meet war costs by exploiting the economic wealth of occupied countries (Explained further below). In the Pacific, both Japan and China relied heavily on money creation, with heavy inflationary consequences. The Japanese economy did not grow during the war years 1940–1945. Increased military expenditure came at the cost of reduced civilian consumption, which fell more than 43% (share of GNP) from 1940 to 1944 (Huff 2015, Table 2.1, 60).

Rising government deficits from 1937 onward were financed mainly through "voluntary savings" that were not voluntary and through the issue of bonds to financial institutions (Cohen 1949, 85–97). Consumer prices quadrupled from 1936 to 1944 (concealed in official statistics, according to which prices had only doubled). By mid-1945, black market goods cost from 10 to 100 times the official prices (Huff 2015, 63). Japan's war costs were covered mainly by money creation, boding ill for postwar stabilization. In its "Greater East Asia Co-Prosperity Sphere," Japan exploited specialized export industries by requiring governments to issue currency, particularly in Burma, Malaya, and the Philippines, which had massive increases in currency in circulation and prices but avoided hyperinflation because subject populations had no real alternatives to using the currency for trade and as a store of value (Huff and Majima 2015).

In China there was currency chaos. Japan issued four different occupation currencies through new banks of issue and printed military yen for its army's use. These "puppet currencies," issued in dollar denominations, were legal tender in the regions of Japanese control. Money creation and inflation financed the Japanese occupation (Huff 2015; Lester 1944). On the other side, Nationalist China used the printing press to finance its growing war expenditures, and Communist-controlled regions issued "about thirty" different local currencies between 1938 and 1945. Bank notes for the Nationalist forces had to be printed in the USA and were flown into China at the rate of over 150 t a month in the last years of the war (Huff 2015, quote p. 66; Young 1965). As in other regions of Japanese occupation in the Pacific, there was little economic development and virtually no state infrastructure to provide funding by direct taxation and tapping domestic savings. Inflation served as the means to tax an impoverished population. In Shanghai, prices when Japan surrendered in 1945 were 85,000 times their 1937 level.

Experience in occupied Europe provides some of the most unusual variations on the use of money as a weapon. Dramatic monetary adaptations were needed to deal with brutal German exploitation through looting and requisitions, and severe shortages of goods from the contraction of production and imports. Regions incorporated into the Reich – Austria and the Czech Sudetenland in 1938, parts of Poland in 1939, Alsace-Lorraine in 1940 – had to adopt the Reichsmark, which, in all cases but Austria, gave the Reich their domestic currency for use in purchasing abroad (Klemann and Kudryashov 2012, 190–193). German policy in Occupied regions settled into a consistent pattern after the initial use of German marks for military purposes in Poland in September 1939. The German authorities quickly found it more expedient to issue military scrip – the *Reichskreditkassenscheine* [RKK]. "Motorized banks of issue" followed the troops to supply RKK for military use in occupied regions only, not in Germany itself. Intended as a transition currency in areas of combat, the RKK proved an easy means to generate purchasing power for military authorities. RKK circulated (against regulations) between the different occupation zones and facilitated a broad range of private purchasing as well as massive black market transactions by official German agencies, especially in France (Scherner 2015). The RKK added to money in circulation without regard for the impact on prices in occupied regions. Agreeing to pay occupation costs in domestic

currency was one way national authorities hoped, mistakenly, to regain control of their domestic money supply (Bell 1942; Klemann and Kudryashov 2012, 194–201).

Once installed in their regions of conquest and occupation, German authorities imposed payments from the subject regions to cover "occupation costs" that were far in excess of their needs for occupation troops. These were levied in domestic currencies, in addition to requisitions of material supplies. "Under the cloak of occupation costs, the national treasuries of occupied Europe were plundered to pay for all the Wehrmacht wanted in any country" (Klemann and Kudryashov 2012, 204). They also set exchange rates in Germany's favor and employed clearing arrangements to limit the need for German payments for foreign goods. As the flow of traded goods went almost exclusively one-way into Germany, national treasuries were ordered to provide credits to cover clearing account payments to their own producers for German purchasing. In effect, "each country paid for its own exports to Germany" (Klemann and Kudryashov 2012, 208). Germany's clearing accounts deficit with countries in Occupied Western Europe at the end of 1944 totaled nearly 15 billion Reichsmarks (Klemann and Kudryashov 2012, Table 13.3). The money paid by German importers into the clearing account at the Reichsbank reduced inflationary pressure in Germany. Both the occupation cost levies and the clearing accounts imposed money creation on occupied territories to pay for German purchasing.

In France, the Germans set initial Occupation costs at 400 million francs per day, and the rate of exchange at 20 francs to the mark, which overvalued the mark by nearly 50%. By French calculation, this would pay for an occupation army of 18 million men. One estimate of actual occupation costs finds these could not have been more than 5% of the initial charge imposed on France, and German records show actual costs were 7% of the charges levied in the Netherlands (Liberman 1996, 47). The Germans reduced the daily cost to 300 million francs in May 1941 but increased it to 500 million in December 1942 after the Allied invasion of North Africa. These occupation payments equaled nearly 20% of French GDP in 1940, close to 37% in 1941 and 1942, and 55% in 1943 (Occhino et al. 2008). French authorities tried to limit the impact of this money creation on French prices by "closing the circuit" – issuing a variety of short-term bonds to bring the currency back to the treasury. This was partially successful through 1942, but the steady increase in Bank advances to the state to fund occupation costs became a point of contention. The Bank of France agreed to each advance reluctantly, seeing these as political decisions that threatened the independence of the bank and had potential to unleash a "dangerous inflation" (Margairaz 2002).

France made the greatest single contribution to German looting of occupied economies in terms of the total value delivered. Calculation of occupation payments to Germany from 1939 to 1945 that includes the credits Scandinavian countries were required to provide, and divided by population, shows France paid less on a per capita basis than the smaller countries, Denmark, the Netherlands, and Norway, and slightly more than Belgium and Greece (Klemann and Kudryashov 2012, Table 13.2). The German methods of extortion adopted diverse forms in the wealthier economies and forced money creation on their targets. In Eastern Europe, German exploitation was more brutal, with the Wehrmacht less concerned to pay for the goods it wanted

when dealing with peoples it considered inferior and creating banks of issue in some regions where there was no issuing authority (Poland, Serbia, Ukraine), to provide the currency they needed for military purchases (Klemann and Kudryashov 2012; Scherner 2015).

Massive requisitions, official procurement, and private purchasing by German armed forces and industry in Occupied Europe took scarce goods from shrinking economies. The impact in countries subject to German exploitation had a significant monetary dimension. Price controls and rationing were imposed to maximize the goods available for German purchasing. Output in occupied countries fell, owing to the shortages of raw materials, energy, transport, and manpower, as well as scarcity of fertilizer and pesticides in agriculture. The occupation authorities set food rations at levels that were deliberately insufficient. French authorities saw the ration levels set for France in September 1940 as intended to bring slow starvation and national decline. Across Occupied Europe, ration levels were insufficient and declined from 1940 to 1945 (Milward 1977). Rationing food was a way to "equalize sacrifice" in individual purchasing capacity in official markets, but everyone needed to find more food. Black markets flourished to meet consumer needs for essential goods and manufacturers' demand for raw materials. Prices for black market goods rose quickly and access to supplies was influenced not only by wealth but also geographic proximity, social networks, and connections with the Germans who held the ultimate authority for economic decisions, enforcement, and the transport of goods. Paul Sanders' research has shown how in Belgium and France, German black market purchasing played a major part in spreading black market practices and developing black market networks that operated with German permission beyond the control of national police (Sanders 2001). German agencies used black market purchasing to draw occupied nationals into their service and earn profit in seeking out goods that had been withheld from official markets for purposes of hoarding, to keep them out of German hands, or to sell for higher profit in parallel markets. This German purchasing was heavily concentrated in France. It included a "Christmas Action" (*Weihnachtsaktion*) in 1942 to buy consumer goods for Germans that included clothing, cosmetics, and toys to the value of nearly 57 million RM in Belgium and 244 million RM in France (Sanders 2001). Such purchases were made with currency from the payment of occupation costs.

For the residents in occupied countries, the shortages, price controls, and rationing required that they find alternate means of supply. Escaping regulations, especially in countries where authorities made a serious attempt at enforcement, meant increased consumer spending to acquire goods, and thus demand for currency, and the creation of alternative means of payment. Ration coupons determined eligibility to purchase essential goods, serving as a parallel currency. This made the coupons subject to fraud, theft, and large-scale counterfeiting. In France, black market prices for bread coupons distinguished between stolen coupons and counterfeits, the latter discounted for their greater risk of detection in use. The most frequent offences were overcharging for goods with fixed official prices. As this practice increased, state price indices failed to reflect the rising cost of survival in buying food and clothing and fuel for heating. Also common were practices that avoided price offences

by bartering goods or providing alternate means of exchange between friends and family members. The volume of food supplies transferred from rural producers to urban consumers via "family parcels" was massive and poorly monitored by the state. The exchange of goods without cash prices reduced demand for cash and inflationary pressure, as did agrarian producers' hoarding of currency. The monetary impact of widespread black market activity was thus limited for a time, masking the extent to which wartime money creation portended postwar inflation.

Black markets were not exclusive to occupied countries. They were a reflexive response to shortages, state controls, and unsatisfied consumer needs. In Germany, despite higher ration levels and harsh punishments (including more than 300 death sentences), black market activity spread during the war and became essential to survival in the postwar Allied occupation (Zierenberg 2015; Steege 2007). In Britain, consumers were able to maintain adequate caloric consumption levels for food, albeit with severely reduced choice and variety. The British rationing system was largely successful in providing "fair shares for all" (Zweiniger-Bargielowska 2000; Roodhouse 2013). In both Britain and the United States, limited access to critical war supplies (petrol, tires, sugar, and meat) fostered black markets for these goods (Rockoff 1984).

One extraordinary plan to use money as a weapon in the European theatre was the German scheme to weaken Britain's war effort by printing counterfeit sterling notes in order to sap confidence in British credit. Initially, the idea was to drop large quantities of low-denomination notes over British cities, expecting many citizens would try to use them. But British air superiority made that impossible as counterfeiting techniques were perfected in 1943, and the notes were used instead for Nazi espionage operations and some international purchasing. The counterfeiting scheme began as Operation Andrew, under Alfred Helmut Naujocks, to analyze and replicate the rag-based paper for British pound notes and design the engraved printing plates to mass produce £5 notes. It was continued as Operation Bernhard, led by SS Major Bernhard Krüger, who assembled a team of counterfeiters in the German concentration camp at Sachsenhausen in 1942. They refined the methods to counterfeit British official currency and began large-scale printing in Barracks 19, reaching as many as 100,000 notes per month, with denominations ranging from 1 to 50 pounds. The operation also developed plates and paper to print US $100 bills. In total, Bernhard produced more than £134 million in Sachsenhausen before the operation had to be moved in January 1945, first to Mauthausen, then to Redl-Zipf in the Austrian Alps, and finally to Ebensee in May 1945. The initial plan had been to produce £30 billion in £5 notes. As the war in Europe ended in May 1945, Operation Bernhard was eliminated, the machinery broken up, the paper burned, and they made frantic efforts to dispose of the crates of printed bills, many of which ended up dumped in the Traun River and Lake Toplitz in the Austrian Alps (Pirie 1961; Malkin 2006).

Liberating Europe from Nazi tyranny posed another challenge: what currency should Allied troops carry to pay for goods and services they needed as they advanced against Axis troops? They could not rely on local currency being readily available, and German use of RKK had set a powerful negative example for the use of invasion currency. Allied policy for Europe evolved from 1942 to the Normandy

invasion in 1944 to reconcile the practical needs of troops on the ground and the importance of monetary sovereignty to national governments, and how they would manage post-occupation monetary stabilization and wartime inflation. For Operation Torch, the Anglo-American invasion of North Africa in November 1942, Allied troops carried specially printed national currencies for military use. The US troops carried yellow seal dollars; British troops used British Military Authority [BMA] pounds. These invasion currencies, "spearhead currencies" in American parlance, allowed military purchasing and the payment of troops without need for domestic currency or concern that an enemy seizure of the currency could cause monetary disorder elsewhere (Petrov 1967).

In Italy, the Allies printed Allied Military lire, which raised questions about the eventual authority and responsibility for a note issue decided by US authorities. Alternate currencies issued in large quantity became popular for black market traffic. Italy had been an enemy country; an Allied Control Commission was established to administer the liberated regions. Elsewhere in Occupied Europe, national governments in exile did not want their monetary sovereignty compromised by Allied currency issue in their countries. They argued against the use of yellow seal dollars in Continental Europe and persuaded Britain and the United States to print national currencies for troop use, under the authority of the governments in exile. France posed a problem in this regard: Roosevelt refused to acknowledge Charles de Gaulle's provisional government. Allied troops landed in Normandy with French franc notes, printed in the United States, bearing the phrase "émis en France," with no issuing authority. Although declared counterfeit by de Gaulle and by the Germans in early June, French citizens in liberated areas accepted these invasion francs. Provisional government accords on their use were reached with the British in June 1944 and with the Americans in August.

In the Pacific theatre, invasion currency was less complicated. The United States relied on dollars overstamped "HAWAII" in the Central Pacific. (The US dollars in Hawaii had been overstamped in 1942; in case Hawaii was invaded by the Japanese, the currency could then be declared void as legal tender.) For the Dutch East Indies, the Netherlands government in exile authorized guilders; for the Philippines, a Victory peso was printed in the United States. For the Allied occupations of Germany and Japan, the United States printed Allied military marks and yen (Petrov 1967; Rundell 1964, 1980) (Table 3).

The prospects for an end to the war and an Allied victory encouraged planning to restore monetary order at the domestic and international levels. Several countries, most notably Belgium, planned monetary reforms to reduce the excess issue of wartime currency. Belgium's government in exile returned in September 1944 with a plan to exchange wartime currency for new Belgian francs printed in Britain, a plan they had to delay until 6 October to wait for arrival of the new francs from London. The Belgian reform required the surrender of all currency notes of 100 francs and above to bank authorities, which ceased to be legal tender. Individuals received an equal quantity of new notes up to 2000 francs. Amounts over 2000 francs were deposited in blocked accounts, release of which was allowed gradually as economic activity recovered in 1944–1945, with frozen balances in October 1945 converted to

Table 3 Monetary impact of World War II (index 1938 = 100)

Country	Real GDP 1944	Military outlay 1943 as % national income/GDP	Money supply 1945	Wholesale prices 1946
France	54.5	–	523	840
Germany	124.5	70	–	110 (1944)
Italy	97.2 (1943)	21	–	2881
Japan	111.8	43	951.9	167.4 (1944)
USSR	137.9	64	–	–
United Kingdom	121.8	55	(M3) 208.5	147.7 (r)
USA	187.4	42	(M1) 330	153.8

Sources:
Real GDP 1944: Harrison (1998), Table 1.3, 10; for France, Maddison (2003), Table 1b, 50
Military outlay: Harrison (1998), Table 1.8, 21; Soviet figure from Davies et al. (1994), 322 (as % of GDP)
Money supply 1945: Harrison (1998); French figure calculated from Banque de France annual reports
Wholesale prices 1946: Harrison (1998); Mitchell (1975); French wholesale price index for December 1946 as reported in the Bank of France annual report; the INSEE WPI average for 1946 was 618

a compulsory long-term loan to the state (Van der Wee and Verbreyt 2010). The twofold purpose of the reform was to reduce the currency in circulation (which it did, from 300 billion to 57.4 billion Belgian francs) and to seize funds of questionable wartime origin. The latter objective, aimed at war profiteering and collaboration, was weakened by the delay, which allowed the conversion of cash into real goods. Monetary reforms in other countries – France, the Netherlands, Norway, Denmark, Finland, Czechoslovakia, and Austria in 1945 – were less ambitious and too late to confiscate illicit earnings.

At the international level, the disruption of war allowed a redesign of the international system. Trade disputes, competitive devaluations, and speculative capital movements had fostered protectionism, instability, economic nationalism, and military aggression between the wars. In the Atlantic Charter in August 1941, the United States and Britain agreed to a statement of principles for the peace that included equal access to trade and raw materials and "the fullest collaboration between all nations in the economic field" to produce greater security and economic advancement. In the discussions that followed, rival British and American plans for a new international order were merged, to establish an international system of pegged exchange rates with an adjustment process to provide international supervision and aid for countries experiencing trade and currency difficulties (the International Monetary Fund and the International Bank for Reconstruction and Development). More than 700 delegates from 45 countries met at Bretton Woods, NH, in July 1944 to finalize details for the postwar monetary regime, which would provide for a gradual return to convertibility and the full operation of the "Bretton Woods system" from 1959 to 1971 (Bordo and Eichengreen 1993; James 1996; Steil 2013). The

cooperative will required, and the critical change in the willingness of the United States to commit resources to international stabilization, were products of the international alliance formed under the pressure of war to defeat Germany, Italy, and Japan.

Conclusion

Money's use in wartime has changed significantly as economies, institutions, and markets have developed over time. This chapter has focused on the two world wars in the twentieth century, on money in advanced economies, and on how the costs of "total war" have been paid. The overwhelming, destructive power of total war has required a greater mobilization of financial power, and greater attention to monetary planning to keep monetary disasters from following as inexorable consequences of war. The differences in the experiences of individual nations in the two world wars highlight the importance of war in altering national monetary policies and international monetary regimes.

The three basic ways to cover war costs did not change: taxation, borrowing (domestic and foreign), and printing money. The era of total war increased financial costs dramatically. Money creation still provided the easiest short-term solution to pay the escalating costs of war. But the demand for currency in economies that had moved from subsistence to industrialization and then to mass consumption increased the potential to raise resources for war finance from all three sources, as well as the damage to be done by excess money creation. The establishment of national banks of issue and their evolution towards modern central banking between the wars facilitated monetary manipulation to pay war costs. But it also developed the financial infrastructure for lower borrowing cost with stronger popular support. The consequences of relying on the printing press became more immediate and visible as inflation took hold. The collapse of currencies in postwar hyperinflations gave clear warning of the need for monetary discipline in managing war finance. Printing money was easy, and its consequences could be deferred. But the strength of financial organization for effective taxation, currency stability, and low-cost borrowing now provided the "sinews of war."

The changes in financial planning and economic mobilization in the two world wars show the importance of financial infrastructure for a sustained effort in total war, and that governments did learn from experience. World War I was fought with little initial financial planning, expecting a short war and leaving the financial costs to be dealt with afterwards. Few could foresee the national and international consequences of inflationary war finance. In World War II, Britain and the United States covered their war costs without significant wartime inflation. They became the main architects for a new international monetary system designed to provide greater stability after the war. Totalitarian states could suppress inflationary pressures and impose costs on others; Germany and Japan used monetary policy and financial exactions in the countries they occupied to pay occupation costs and extract resources for their battles elsewhere. Occupied countries and those with

under-developed banking and financial services had less freedom and less protection from monetary disaster. They experienced higher inflation, in some cases catastrophic hyperinflation. The greatest monetary problems occurred in the countries occupied and exploited by the Axis powers and in those lacking administrations capable of resisting the demands for more money. Money served as a weapon of war for aggressors, an instrument for exploitation, but it remained vital to provide the means to counter aggression and to restore economic stability and security in peacetime.

Cross-References

- ▶ International Monetary Regimes: The Gold Standard
- ▶ International Monetary Regimes: The Interwar Gold Exchange Standard
- ▶ The Historical Evolution of Central Banking

References

Ahamed L (2009) Lords of finance: the bankers who broke the world. Penguin Books, New York

Antipa PM (2016) How fiscal policy affects prices: Britain's first experience with paper money. J Econ Hist 76(4): 1044–1077

Balderston T (1989) War finance and inflation in Britain and Germany, 1914–1918. Econ Hist Rev 42(2):222–244

Bell HH (1942) Monetary problems of military occupation. Mil Aff 6(3):77–88

Bensel RF (2000) The political economy of American industrialization, 1877–1900. Cambridge University Press, Cambridge

Blancheton B (2001) Le pape et l'empereur: La Banque de France, la direction du Trésor et la politique monétaire de la France (1914–1928). Albin Michel, Paris

Bonney R (ed) (1995) Economic systems and state finance. Clarendon Press, Oxford

Bonney R (ed) (1999) The rise of the fiscal state in Europe, c. 1200–1815. Oxford University Press, Oxford

Bordo MD, Eichengreen B (eds) (1993) A retrospective on the Bretton Woods system: lessons for international monetary reform. University of Chicago Press, Chicago

Bordo MD, White EN (1991) A tale of two currencies: British and French finance during the Napoleonic Wars. J Econ Hist 51(2):303–316

Bordo MD, Goldin C, White EN (eds) (1998) The defining moment: the Great Depression and the American economy in the twentieth century. University of Chicago Press, Chicago

Brewer J (1989) The sinews of power: War, money and the English State, 1688–1783. Alfred A. Knopf, New York

Brezis ES, Crouzet FH (1995) The role of assignats during the French Revolution: an evil or a rescuer? J Eur Econ Hist 24(1):7–40

Broadberry S, Harrison M (eds) (2005) The economics of World War I. Cambridge University Press, Cambridge

Brown WA Jr (1940) The international gold standard reinterpreted 1914–1934. National Bureau of Economic Research, New York

Burdekin RCK, Weidenmier MD (2001) Inflation is always and everywhere a monetary phenomenon: Richmond vs. Houston in 1864. Am Econ Rev 91(5):1621–1630

Burk K (1979) J.M. Keynes and the exchange rate crisis of July 1917. Econ Hist Rev 32(3):405–416

Burk K (1985) Britain, America and the sinews of war, 1914–1918. George Allen & Unwin, London
Cassis Y (2011) Crises and opportunities: the shaping of modern finance. Oxford University Press, Oxford
Clapham J (1946/1976) Sir John Clapham's account of the financial crisis in August 1914. In: Sayers RS (ed) The Bank of England 1891–1944, appendixes. Cambridge University Press, Cambridge, pp 31–45
Clarke SVO (1967) Central bank cooperation, 1924–31. Federal Reserve Bank of New York, New York
Cohen JB (1949) Japan's economy in war and reconstruction. University of Minnesota Press, Minneapolis
Crouzet F (1993) La grande inflation: La monnaie en France de Louis XVI à Napoléon. Fayard, Paris
Daunton M (2002) Just taxes: the politics of taxation in Britain, 1914–1979. Cambridge University Press, Cambridge
Davies RW, Harrison M, Wheatcroft SG (eds) (1994) The economic transformation of the Soviet Union, 1913–1945. Cambridge University Press, Cambridge
Davis BJ (2000) Home fires burning: food, politics, and everyday life in World War I Berlin. University of North Carolina Press, Chapel Hill
De Cecco M (1984) The international gold standard: money and empire. Frances Pinter, London
Dickson PGM (1967) The financial revolution in England: a study in the development of public credit 1688–1756. Macmillan, London
Dincecco M (2009) Fiscal centralization, limited government, and public revenues in Europe, 1650–1913. J Econ Hist 69(1):48–103
Dincecco M (2015) The rise of effective states in Europe. J Econ Hist 75(3):901–918
Druell-Korn C (2015) Du papier, de grâce du papier pour l'amour de Dieu. Le ministère du Commerce au secours de la circulation monétaire pendant la Grande Guerre: monnaies de nécessité consulaires et chèques postaux. In: Descamps F, Quennouëlle-Corre L (eds) La mobilisation financière, le troisième front. CHEFF, Paris, pp 75–96
Eck J-F, Touchelay B (2018) La Banque de France face aux monnaies de nécessité émises en 1914–1918: le cas du Nord de la France occupée. In: Banques Centrales dans la Grande Guerre. Presses de Sciences Po, Paris
Eichengreen B (1992) Golden fetters: the gold standard and the Great Depression 1919–1939. Oxford University Press, New York
Eichengreen B, Temin P (2000) The gold standard and the Great Depression. Contemp Eur Hist 9(2):183–207
Ellis HW (1940) German exchange control, 1931–1939: from an emergency measure to a totalitarian institution. Q J Econ 54(4/2):1–158
Feinstein CH, Temin P, Toniolo G (1997) The European economy between the wars. Oxford University Press, Oxford, UK
Feldman GD (1993) The great disorder: politics, economics, and society in the German inflation 1914–1924. Oxford University Press, New York
Flandreau M (1996) The French crime of 1873: an essay on the emergence of the international gold standard, 1870–1880. J Econ Hist 56(4):862–897
Flandreau M (1997) Central bank cooperation in historical perspective: a sceptical view. Econ Hist Rev 50(4):735–763
Forsyth DJ (1993) The crisis of liberal Italy: monetary and financial policy, 1914–1922. Cambridge University Press, Cambridge
Frankenstein R (1982) Le prix du réarmement français (1935–1939). Publications de la Sorbonne, Paris
Fratianni M, Spinelli F (1997) A monetary history of Italy. Cambridge University Press, Cambridge
Friedman M (1990a) The crime of 1873. J Polit Econ 98(6):1154–1194
Friedman M (1990b) Bimetallism revisited. J Econ Perspect 4(4):85–104
Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1867–1960. Princeton University Press, Princeton

Galbraith JK (1952) A theory of price control: the classic account. Harvard University Press, Cambridge, MA; 1980 reprint
Gallarotti GM (1995) The anatomy of an international monetary regime: the classical gold standard, 1880–1914. Oxford University Press, New York
Gatrell P (2005) Russia's First World War: a social and economic history. Routledge, New York
Hardach G (1977) The First World War 1914–1918. University of California Press, Berkeley
Hardach G (2015) De l'argent pour Armageddon. La mobilization financière de l'Allemagne entre 1914 et 1918. In: Descamps F, Quennouëlle-Corre L (eds) La mobilisation financière, le troisième front. CHEFF, Paris, pp 75–103
Harris SE (1930) The assignats. Harvard University Press, Boston
Harrison M (ed) (1998) The economics of World War II: six great powers in international comparison. Cambridge University Press, Cambridge
Hawtrey RG (1931) The gold standard in theory and practice, 2nd edn. Green and Co., London
Healy M (2004) Vienna and the fall of the Habsburg Empire: total war and everyday life in World War I. Cambridge University Press, Cambridge
Holtfrerich C (1986) The German inflation 1914–1923: causes and effects in international perspective (trans: Balderston T). Walter de Gruyter, Berlin/New York
Horn M (2002) Britain, France and the financing of the First World War. McGill-Queen's University Press, Kingston/Montreal
Huff G (2015) Finance for war in Asia and its aftermath. In: Geyer M, Tooze A (eds) The Cambridge history of the Second World War, volume III, Total war: economy, society and culture. Cambridge University Press, Cambridge, pp 56–93
Huff G, Majima S (2015) Paying for war, 1941–1945: how Japan financed Southeast Asia's occupation. In: Boldorf M, Okazaki T (eds) Economies under occupation: the Hegemony of Nazi Germany and Imperial Japan in World War II. Routledge, London, pp 55–70
Hummel JR (2014) Emancipating slaves, enslaving free men: a history of the American Civil War. Open Court, Chicago
Imlay T (2003) Facing the Second World War: strategy, politics, and economics in Britain and France, 1938–1940. Oxford University Press, Oxford
Jacobs M (2005) Pocketbook politics: economic citizenship in twentieth-century America. Princeton University Press, Princeton
James H (1996) International monetary cooperation since Bretton Woods. Oxford University Press, New York
Kallet L (2001) Money and the corrosion of power in Thucydides: the Sicilian expedition and its aftermath. University of California Press, Berkeley
Kallet-Marx L (1993) Money, expense and naval power in Thucydides' History, 1.-5.24. University of California Press, Berkeley
Kennedy P (1988) The rise and fall of the great powers: economic change and military conflict from 1500 to 2000. Random House, New York
Keynes JM (1940) How to pay for the war. Macmillan and Co., London
Kindleberger CP (1986) The world in depression 1929–1939, rev edn. University of California Press, Berkeley
Kindleberger CP (1993) A financial history of Western Europe, 2nd edn. Oxford University Press, New York
Klemann H, Kudryashov S (2012) Occupied economies: an economic history of Nazi-Occupied Europe, 1931–1945. Berg, London
League of Nations (1944) International currency experience: lessons of the inter-war period. League of Nations, Geneva
Lester RA (1944) International aspects of wartime monetary experience. Princeton essays in international finance, no. 3. International Finance Section, Princeton
Liberman P (1996) Does conquest pay? The exploitation of occupied industrial societies. Princeton University Press, Princeton

Machiavelli N (2003) Art of war, translated, edited and with a commentary by Christopher Lynch. University of Chicago Press, Chicago
Maddison A (2003) The world economy: historical statistics. OECD, Paris
Malkin L (2006) Krueger's men: the secret Nazi counterfeit plot and the prisoners of Block 19. Little, Brown and Co., New York
Margairaz M (2002) La Banque de France et l'Occupation. In: Margairaz M (ed) Banques, Banque de France et Seconde Guerre mondiale. Albin Michel, Paris, pp 37–84
Milward AS (1977) War, economy and society, 1939–1945. University of California Press, Berkeley
Mitchell WC (1903) A history of the greenbacks, with special reference to the economic consequences of their issue: 1862–65. University of Chicago Press, Chicago; reprint 1960
Mitchell BR (1975) European historical statistics 1750–1970, abridged edn. Columbia University Press, New York
Moggridge DE (1972) British monetary policy 1924–1931: the Norman Conquest of $4.86. Cambridge University Press, Cambridge
Morgan EV (1952) Studies in British financial policy, 1914–1925. Macmillan & Co., London
Mouré K (1991) Managing the franc Poincaré: economic understanding and political constraint in French monetary policy, 1928–1936. Cambridge University Press, Cambridge
Mouré K (2002) The gold standard illusion: France, the Bank of France and the international gold standard 1914–1939. Oxford University Press, Oxford
O'Brien P (2011) The nature and historical evolution of an exceptional fiscal state and its possible significance for the precocious commercialization and industrialization of the British economy from Cromwell to Nelson. Econ Hist Rev 64(2):408–446
Occhino F, Oosterlinck K, White EN (2008) How much can a victor force the vanquished to pay? France under the Nazi boot. J Econ Hist 68(1):1–45
Ormrod WM, Bonney M, Bonney R (1999) Crises, revolutions and self-sustained growth: essays in European fiscal history, 1130–1830. Shaun Tyas, Stamford
Parker G (1988) The military revolution: military innovation and the rise of the west, 1500–1800. Cambridge University Press, Cambridge
Peden GC (2000) The Treasury and British public policy, 1906–1959. Oxford University Press, Oxford
Petrov V (1967) Money and conquest: allied occupation currencies in World War II. Johns Hopkins Press, Baltimore
Pigou AC (1921) The political economy of war. Macmillan and Co., London
Pirie A (1961) Operation Bernhard. William Morrow & Company, New York
Ritschl A (2005) The pity of peace: Germany's economy at war, 1914–1918 and beyond. In: Broadberry S, Harrison M (eds) The economics of World War I. Cambridge University Press, Cambridge, pp 41–76
Roberts R (2013) Saving the city: the great financial crisis of 1914. Oxford University Press, Oxford
Rockoff H (1984) Drastic measures: a history of wage and price controls in the United States. Cambridge University Press, Cambridge
Rockoff H (2005) Until it's over, over there: the US economy in World War I. In Broadberry and Harrison, The economics of World War I. Cambridge University Press, New York, pp 310–343
Rockoff H (2012) America's economic way of war: war and the US economy from the Spanish-American War to the Persian Gulf War. Cambridge University Press, Cambridge
Roll R (1972) Interest rates and price expectations during the Civil War. J Econ Hist 32(2):476–498. s
Roodhouse M (2013) Black market Britain 1939–1955. Oxford University Press, Oxford
Rowlands G (2012) The financial decline of a great power: war, influence and money in Louis XIV's France. Oxford University Press, Oxford
Rundell W Jr (1964) Black market money: the collapse of U.S. military control in World War II. Louisiana State University Press, Baton Rouge

Rundell W Jr (1980) Military money: a fiscal history of the U.S. Army overseas in World War II. Texas A&M University Press, College Station
Sanders P (2001) Histoire du marché noir 1940–1946. Perrin, Paris
Sargent TJ, Velde FR (1995) Macroeconomic features of the French Revolution. J Pol Econ 103(3):474–518
Sayers RS (1976) The Bank of England 1891–1944. Cambridge University Press, New York
Scherner J (2015) The German system of financing occupation. In: Boldorf M, Okazaki T (eds) Economies under occupation: the Hegemony of Nazi Germany and Imperial Japan in World War II. Routledge, London, pp 39–54
Schulze M-S (2005) Austria-Hungary's economy in World War I. In: Broadberry S, Harrison M (eds) The economics of World War I. Cambridge University Press, Cambridge, pp 77–111
Schumpeter JA (1954) The crisis of the tax state. Int Econ Pap 4:5–38
Sharkey RP (1959) Money, class and party: an economic study of Civil War and reconstruction. The Johns Hopkins Press, Baltimore
Steege P (2007) Black market, cold war: everyday life in Berlin, 1946–1949. Cambridge University Press, Cambridge
Steil B (2013) The Battle of Bretton Woods: John Maynard Keynes, Harry Dexter White, and the making of a new world order. Princeton University Press, Princeton
Strachan H (2004) Financing the First World War. Oxford University Press, Oxford
Tilly C (1992) Coercion, capital, and European states, AD 990–1992. Blackwell, Oxford
Tooze A (2006) The wages of destruction: the making and breaking of the Nazi economy. Allen Lane, London
Tooze A, Martin J (2015) The economics of the war with Nazi Germany. In: Geyer M, Tooze A (eds) The Cambridge history of the Second World War, volume III, Total war: economy, society and culture. Cambridge University Press, Cambridge, pp 27–55
Unger I (1964) The greenback era: a social and political history of American Finance, 1865–1879. Princeton University Press, Princeton
Van der Wee H, Verbreyt M (2010) A small nation in the turmoil of the Second World War: money, finance and occupation (Belgium, its enemies, its friends, 1939–1945) (trans: Parker F). Leuven University Press, Leuven
Winter J, Robert J-L (eds) (1997) Capital cities at war: Paris, London, Berlin 1914–1919. Cambridge University Press, Cambridge
Withers H (1918) War and Lombard street. John Murray, London
Young AN (1965) China's wartime finance and inflation, 1937–1945. Harvard University Press, Cambridge, MA
Zierenberg M (2015) Berlin's Black market: 1939–1950. Palgrave Macmillan, New York
Zweiniger-Bargielowska I (2000) Austerity in Britain: rationing, controls, and consumption, 1939–1955. Oxford University Press, Oxford

The Anatomy of Inflation: An Economic History Perspective

40

Pierre L. Siklos and Martin T. Bohl

Contents

Abstract

Understanding the dynamics of inflation continues to remain at the center of work in macroeconomics. This chapter provides an up-to-date survey of inflation and its principal driving forces. The chapter starts by reviewing stylized fact about inflation and summarizing the behavior of inflation in some commodity money systems. It then summarizes the main theories that are used to explain movements in inflation. We also discuss the role played by inflation volatility and episodes of hyperinflation.

P. L. Siklos (✉)
Wilfrid Laurier University and Balsillie School of International Affairs, Waterloo, ON, Canada
e-mail: psiklos@wlu.ca

M. T. Bohl
School of Business and Economics, University of Münster, Münster, Germany
e-mail: martin.bohl@wiwi.uni-muenster.de

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_35

Keywords
Inflation · Inflation expectations · Hyperinflation · Central banks · Monetary policy

Introduction

Over 40 years ago Laidler and Parkin (1975) published a survey of inflation. Readers were reminded that inflation has social consequences including effects on the distribution of income, wealth effects, and the purchasing power of money. At the time economies around the world were beginning to grapple with the fallout from the first oil price shock. A second one would be inflicted on the world economy a few years later leading to a combination of high inflation and unsatisfactory economic growth that came to be called stagflation. Rather strikingly their survey concluded as follows: "Until we have a much more fully articulated analysis of the formation of expectations and of the interaction of expectations formation and the behaviour of economic agents it is hard to see how we are going to make any significant further progress in understanding inflation." Beyond understanding the formation of inflation expectations, we continue to struggle to disentangle the demand and supply side influences of oil price shocks on inflation (e.g., Parkin 1980; Bordo and Orphanides 2013).

Fast forward over 30 years later as the gathering storm of what would eventually be called the global financial crisis (GFC) was about to change the course of history, then Fed Chairman Ben Bernanke (2008) would utter the following words: "But there is much we do not understand about inflation expectations, their determination, and their implications." As this is written, we no longer worry about the consequences of high inflation. Instead it is low inflation and low economic growth, that is, lowflation and secular stagnation, that preoccupy central banks. In any case, inflation and understanding its dynamics continue to remain at the center of the work of macroeconomics.

In spite of considerable progress in both the theoretical and empirical realms, there continues to be much that can be learned from how inflation has evolved. One area where both theory and empirics provide clear answers about the dynamics of inflation but has been neglected in recent years concerns the economics of hyperinflation. Other than the relatively recent experience of hyperinflation in Zimbabwe and Venezuela, episodes of this kind have become a curiosity. Curiously, one does not have to wait very long once the GFC erupted for several prominent observers to worry that the response of the monetary authorities, via massive injections of liquidity now referred to as quantitative easing (QE), would lead to a return to high inflation (e.g., Meltzer 2009), if not hyperinflation. Hence, a more up-to-date survey of inflation ought to include a reminder of what episodes of hyperinflation teach us about the dynamics of inflation.

The remainder of this survey is organized as follows. The next section provides a short list of some stylized facts about the evolution of inflation across countries and time. We next provide a brief summary of the behavior of inflation in some

commodity money systems. This provides a contrast for the challenges facing a fiat money system where central banks are expected to control inflation and the connection with the policy strategy in place. We then turn to an overview of the extreme case of hyperinflation which illuminates some of the key insights in modern monetary and macroeconomics. The chapter concludes with a summary and future questions that remain to be addressed.

Some Stylized Facts About Inflation

Almost half a century ago, Milton Friedman (1970) made famous the following proposition: "Inflation is always and everywhere a monetary phenomenon in the sense that it can be produced only by a more rapid increase in the quantity of money than in output . . .." Figures 1 and 2 show the annual history of price and money supply movements in the UK and Norway. For prices the data go back to at least 1600, while for the two money supply indicators, the data are shown only since the late 1800s. While price-level rises and money supply-level changes do appear to parallel each other, the relationship between the various time series also suggests that there may be other factors that contribute to changes in inflation over the time.

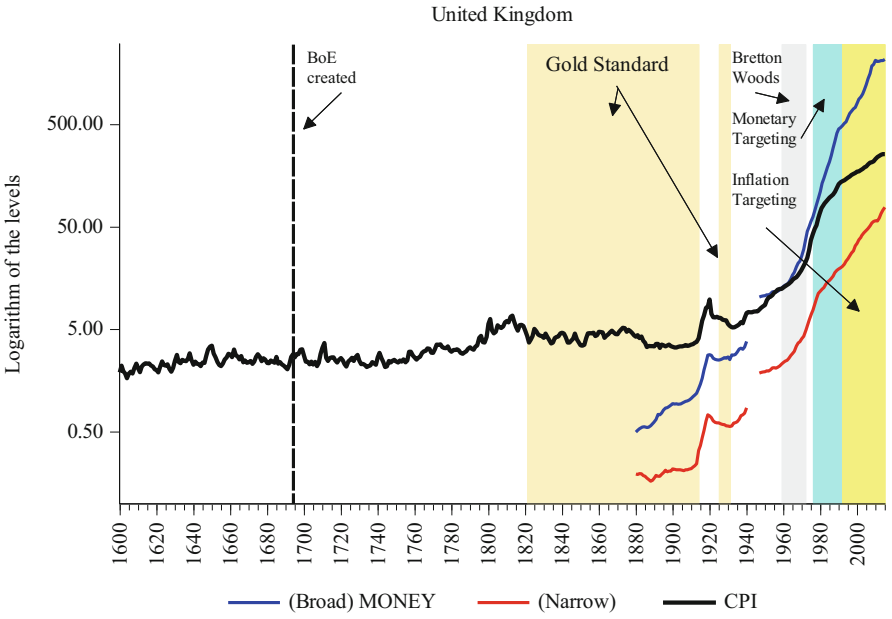


Fig. 1 Consumer prices and the money supply in the UK Since 1600. Data and sources are from Bordo and Siklos (2016, 2018). Narrow money is essentially currency and bank reserves; broad money includes deposits. Definitions of reserves and deposits change over time. CPI is a measure of consumer prices. Precise dating of policy regimes is given in Bordo and Siklos (2018)

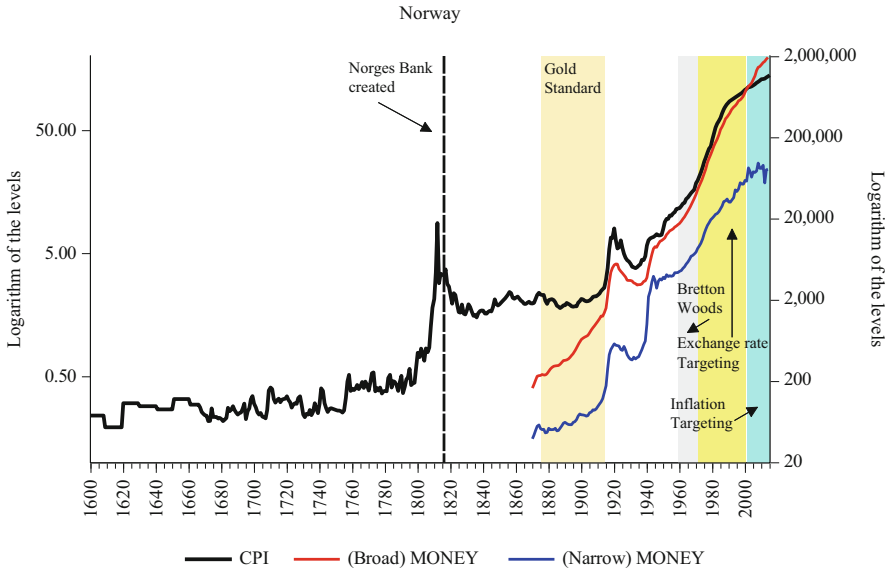


Fig. 2 Consumer prices and the money supply in Norway Since 1600. See note to Fig. 1

One obvious element that explains the relationship between money and prices is the monetary policy strategy in place. Figures 1 and 2 also highlight the duration of four monetary policy regimes that have so far been established. Well after each central bank was created (also shown in the figures), the gold standard was introduced. In both the UK and Norway, this regime is identified not only by stable prices, at least relative to the regimes that follow, but also by its duration. This is especially visible in the UK's case. Three other regimes mark monetary history over the past four centuries. They are Bretton Woods, where a form of exchange rate pegging was in place; monetary or exchange rate targeting introduced to stem the rise in inflation once the Bretton Woods era essentially ended and, together with the two oil price shocks of the 1970s, led to greater emphasis on inflation control; finally, inflation targeting has become the monetary regime of choice in a large number of countries (Siklos 2017) over the past two to three decades once it became clear that other alternatives had failed to produce low and stable inflation.

As economies gravitated toward inflation targeting, there was also a movement away from reliance on monetary aggregates to determine the likely inflationary outcome of changes in the stance of monetary policy. Instead, the focus shifted to using an interest rate instrument that would be adjusted according to the degree of economic slack and whether inflation exceeded some mandated or implicit level deemed acceptable. This modus operandi was captured by the now eponymous Taylor rule (Taylor 1993). Until the GFC of 2007–2009, the policy recommendations based on such a rule dominated the conduct of monetary policy.

Together with questions stemming from the inability to achieve a consensus on how best to define "money" in light of the technological changes that influenced

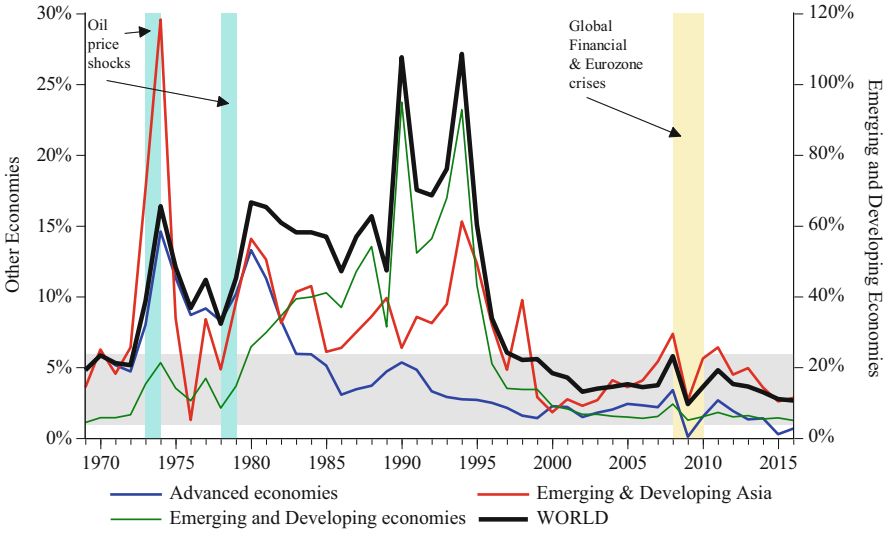


Fig. 3 Inflation rates in various parts of the globe since 1969. Inflation is the percent change in the CPI. Country groupings are as defined by the International Monetary Fund. Data are annual from International Financial Statistics, July 2017 CD-ROM. Oil shocks are 1973–1974, and 1979–1980. Global financial crisis is 2008–2009

developments in the financial sector, money as a determinant of inflation disappeared into the background. The theoretical work of Woodford and others also helped bury the role of money as a relic of past ideas (e.g., Woodford 2003). However, reports of money's demise are perhaps greatly exaggerated. Part of the reason for relegating money to the dustbin of history were the growing number of studies worldwide that suggested that the long-run relationship between a monetary aggregate, almost any that one cared to used, interest rates, and GDP, that is, the constituents of a stable money demand function, broke down (e.g., Friedman and Kuttner 1992; Siklos and Barton 2001). However, Lucas and Nicolini (2015) and Benati et al. (2016) have produced evidence that revives the notion that a stable money demand exists in a large number of countries.

Figure 3 plots inflation in major parts of the globe. There are at least three notable features in the data. First, the record of inflation over the past 50 years or so indicates considerable diversity in inflation performance around the world. Historically, inflation has tended to be lowest in advanced economies and highest in the emerging and developing world. Second, inflation rates are volatile and exhibit sharp movements over time. Clearly, the two oil price shocks and the GFC, both highlighted in the figure, explain some of these movements. The figure also hints at some possible global co-movement in inflation around the globe. Indeed, this recent phenomenon gave rise to continuing interest in estimating the extent to which movements in inflation are influenced by global factors (e.g., Ciccarelli and Mojon 2010) or the rise of China (e.g., Pang and Siklos 2016).

Next, even if we take a liberal view of how low and stable inflation might be defined, namely, a range of 1–6%, achieving low and stable inflation is a fairly recent phenomenon (also see Bordo and Siklos 2016). In advanced economies, low and stable inflation is a range that encompasses 1–3% in CPI inflation with a mean target of usually 2%. The same range tends to be more variable, wider, and associated with a higher mean inflation target in less advanced economies. Finally, a remarkable feature of the global inflation record is the apparent tendency for inflation to converge since the early 2000s. Not since the 1950s to early 1960s have inflation rates remained so low or stable. Indeed, as this is written, policymakers are asking whether the emphasis on inflation control has been too strict and whether inflation worldwide is too low. Alternatively, the suggestion has also been made that the growth of e-commerce and the emerge of large firms with a global reach and market dominance (e.g., Amazon) may also partially explain the persistence of low inflation in spite of the economic recovery since the crisis of 2007–2008 and the introduction of QE. However, the evidence to date is limited and does not yet point to these forces as the principal explanation of the low inflation environment as this is written (e.g., Charbonneau et al. 2017; Gorodnichenko and Talavera 2017). All of the foregoing developments mark the most serious threat to the inflation-targeting regime in particular since this type of monetary policy strategy came on the scene in the late 1980s and early 1990s, pioneered by New Zealand and Canada. Nevertheless, low inflation has become a global phenomenon, and all monetary policy regimes are being called into question.

So far we have assumed that there is a widely accepted consensus about the precision with which inflation is measured. Nothing is further from the truth. In particular, it is easy to confuse a general increase in the price level, the most general definition of inflation, with a change in relative prices. The former is what monetary policy is tasked to control, while the latter reflects the forces of aggregate demand and supply. Even if the foregoing distinction is clearly understood, there is then the problem of choosing the right price level against which policy decisions are evaluated.

To illustrate, consider the case of Canada. Figure 4 plots seven different indicators of inflation in Canada since 2000. The data were obtained from http://www.bankofcanada.ca/rates/price-indexes/bcpi/. Two of the seven are highlighted, namely, inflation in the CPI, also often referred to as headline inflation, and CPI that excludes food, energy prices, and indirect taxes (CPI XFET), also known as core inflation. The former is the one that the public has come to expect and is the focus of monetary policy performance; the latter is considerably less volatile and, by excluding price variations likely due to mainly changes in aggregate supply conditions, provides clearer signals to the central bank about whether a correction in the stance of monetary policy is needed. Nevertheless, there are other ways of "smoothing" out volatile components in headline CPI. Hence, CPI common, CPI median, and CPI trim represent three different methodologies used to filter out the "unwanted" variation in aggregate price changes. Methodological details are outside the scope of this chapter, but the point is that there are many ways of defining inflation.

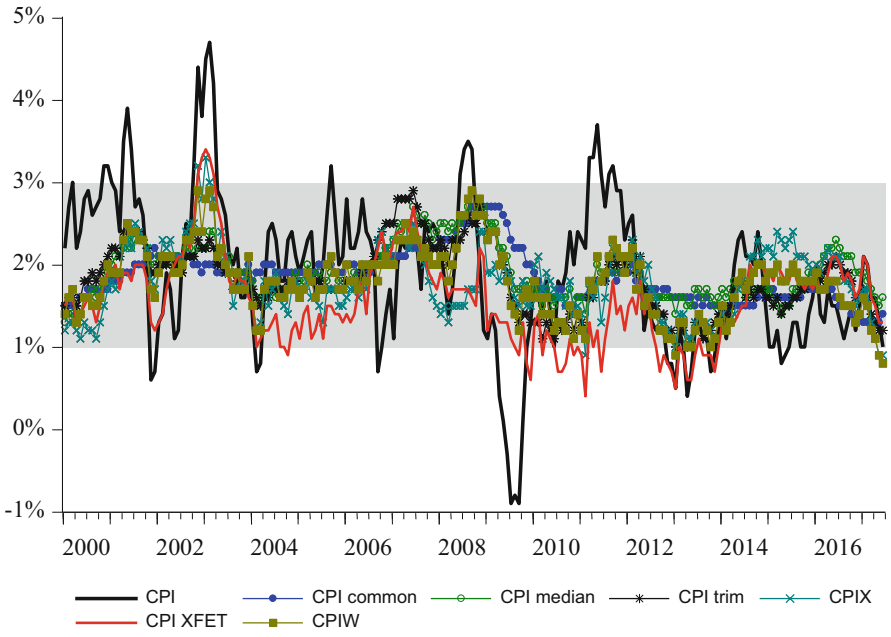


Fig. 4 Varieties of inflation rates: the case of Canada, 2000–2017. Data source is given in the text. The data source also provides detailed definitions for each measure of inflation. The shaded area represents the 1–3% inflation target range the Bank of Canada is expected to meet. Inflation is 100 times the first log difference of consumer prices. Data are monthly

Although a consumer price index, or CPI, is a term used around the globe to express changes in the purchasing power of money over time, there are subtle but important differences depending, for example, on how housing and interest rates costs are factored in, among other differences that impact the measurement of prices. It would take us far afield to devote any more space to these issues. Instead, the point is that since the CPI captures a representative basket of goods and services, the monetary authorities must make allowances for some of the components of CPI that are likely to be highly volatile, such as food and energy prices. Of course, food and energy prices need not be the only component one may wish to focus on. Instead, policymakers, and others interested in the evolution of price movements, may also be interested in particular types of prices, such as changes in transportation, computing, and oil prices, to give three examples, because some of these may be harbingers of broader changes in future CPI movements if changes in these prices feed into aggregate prices.

Why should we care? The degree to which changes in individual prices of goods and services feed into aggregate prices translates into inflation. Perhaps more importantly, changes in inflation may well lead to changes in inflation expectations. Economic theory and the tenets of modern central banking give pride of place to the role of inflation expectations. Indeed, the behavior of both observed and expected

inflation has taken on an even more critical role since inflation rates, especially in the advanced economies, have undershot announced objectives for several consecutive years.

Persistently low inflation is potentially as much of a worry as persistently high inflation because it raises the possibility that current inflation will be strongly influenced by its past history. In slightly more technical terms, this means that inflation can be well modeled as a first-order autoregressive process, that is:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \varepsilon_t. \tag{1}$$

If the parameter β_1 is high, this implies a high degree of persistence. Now write Eq. (1) as follows:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \beta_2 I_t + \varepsilon_t. \tag{2}$$

If observed inflation is below the inflation target, then β_2 has the effect in Eq. (2) of actually raising the degree of inflation persistence. Just as the introduction of inflation targeting reduced inflation persistence (e.g., Siklos 1999; Benati 2008), there is the possibility that the GFC reversed the direction of persistence.

To illustrate, we use the data for inflation in advanced economies referred to above. For the sample of annual data from 1969 to 2005, the estimate of β_1 in Eq. (1) is 0.80. The same coefficient drops to 0.46 for the period after 1995 when most advanced economies, including ones that adopted an explicit numerical target for inflation, placed inflation control as the heart of their monetary policy strategy. Finally, for the period from 2007 to 2016, the persistence parameter falls still further to 0.07 and is statistically insignificant, an indication of the disappearance of persistence in inflation in the conventional sense since the GFC. When emerging and developing economies are considered, persistence falls dramatically from 0.77 (1970–2016 period) to 0.11 (2001–2016). The year 2001 is chosen as this is approximately when several emerging markets began to adopt inflation targeting (e.g., Siklos 2017).

Estimates of Eq. (2) confirm the significant drop in persistence (not shown). Assuming that the GFC is defined as a period that begins in 2007 and ends in 2009, we observe a rise in inflation persistence although the coefficient is only marginally statistically significant at conventional levels (i.e., 10% level). Similar results for both β_2 I_t and the GFC are obtained for the data from emerging market economies. Clearly, the experience of individual economies within each group will differ significantly.

What helps anchor inflation expectations? If inflation is high and the public does not believe the promises made by the central bank to reduce inflation, then their expectations will not adjust until it is convinced that inflation is under control and remains consistently lower than before.

We next take a step back from differences in how inflation is measured or the quality of such measurements and, in what follows, assume that the focus of our analysis is some indicator of headline inflation.

Inflation in the Commodity Currency Era

Throughout history, there have been several commodity-based currencies. These include regimes based on commodities ranging from playing cards to the bimetallic systems used in parts of Europe and the USA. However, the most important commodity money system has undoubtedly been the gold standard (e.g., Bordo 1999).

It is important to remind readers that there is more than one type of gold standard regime depending on the extent to which the monetary system is tethered to the price of gold. In theory, the price of gold dictated the quantity of money in circulation. While gold coins could circulate, it quickly became apparent that it was more efficient, and safer, to print notes that were fully backed by the available gold reserves. Therefore, unless there were gold discoveries that allowed growth in the available quantity of money, the price of gold would have to change to accommodate a rise in the demand for currency in response to economic growth. Strictly speaking then, the gold standard constrains price-level movements by fixing the quantity of money. This explains a good portion of the changes in prices in the UK and Norway in the nineteenth century.

In practical terms, all countries that adopt such a gold standard also ought to lead to common price-level movements. The reason is that, with unfettered capital flows, the gold standard induces an automatic balance of payments adjustment. For example, a positive economic shock in one country would, since the quantity of money is more or less constant in the short run, produce a fall in prices. This makes exports more attractive, thereby producing, other things equal, a balance of payments surplus which also translates into an inflow of gold. The latter creates forces that offset the initial fall in prices. Inflation does exist under a gold standard, but the price-specie-flow mechanism returns inflation to zero once the shocks end. Note that if inflation is possible, deflation is also certain under the classical gold standard. As shown in Fig. 5 for six economies, movements in prices broadly parallel each other.

Indeed, in the absence of some of the uncontrollable factors and assuming that governments and their central banks adhere to the "rules of the game," inflation should behave like a stationary process, that is, a time series with a roughly zero mean and a constant variance. Consider the period referred to as the "classical" gold standard, that is, when most countries that adopted the regime followed its precepts yielding little inflation. The mean inflation rate, for example, in the UK between 1880 and 1914 was −0.19% with a standard deviation of 4.07%; the same figures for Norway are, respectively, 0.23% and 3.55%. As others have noted (e.g., Eichengreen 1992; Bordo 1981; Bordo and Schwartz 1984), this was also a period of strong economic growth. Unfortunately, the regime could not survive the onset of the World War I when it broke down. Instead, an era of inflation emerged as is apparent from Fig. 6 which highlights inflation performance in the UK and Norway during and after the classical gold standard era.

The end of that conflict brought about a nostalgia for the return to rules to guide the conduct of in monetary policy. However, for reasons alluded to above, the price at which gold is fixed is critical, and by the early 1930s, it was massive outflows of

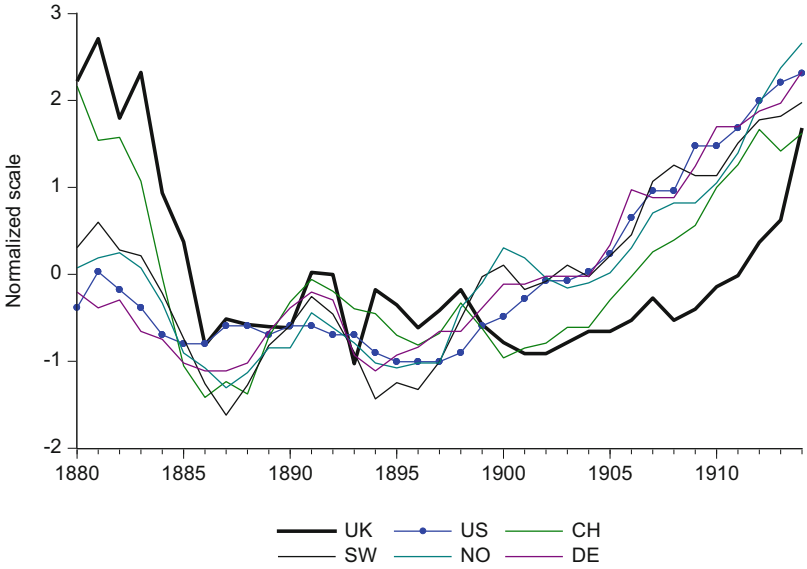


Fig. 5 Price levels during the gold standard, 1880–1914. The price levels (different base years) are normalized (i.e., divided by their respective standard deviations) to ensure comparability. UK is United Kingdom, US is United States, CH is Switzerland, SW is Sweden, NO is Norway, and DE is Germany. Data are annual from Bordo and Siklos (2018)

gold, following an inappropriate fixing of the price of gold, that forced the UK to abandon the standard. The USA also hastened the demise of the gold exchange standard, as it was now called, when President Franklin Roosevelt suspended the regime. By this time, policymakers were no longer as keen to adhere to the version of the gold standard that prevailed during the classical period so that the "Golden Fetters" no longer bound the quantity of money to the price of gold to the same degree. Many decades later, after several bouts of high inflation (see below), the attraction to gold among some observers remains. Nevertheless, to avoid inflation running out of control, as well as for practical reasons, policymakers prefer today to tie their hands by following other kinds of rules (e.g., by following a Taylor-type rule previously mentioned) rather than return to what Keynes (1924) once called "a barbarous relic."

Central Banks and Inflation

The Road to Inflation Targeting

The end of the World War II may have ended the appeal of the gold standard, but it did not end preference for stable exchange rates since to do so otherwise would allow the potential for beggar-thy-neighbor policies that contributed to the turmoil in

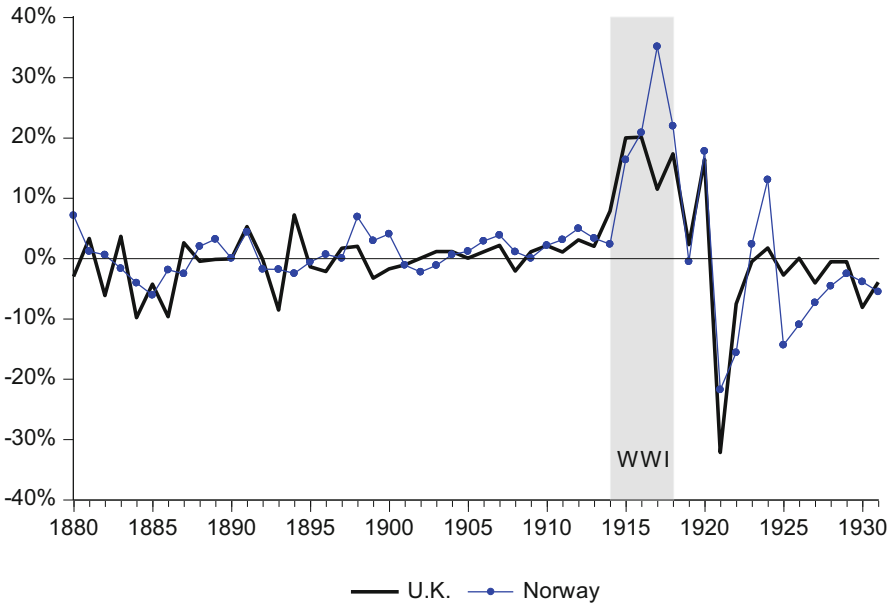


Fig. 6 Inflation in the UK and Norway during the gold standard, 1880–1931. Source is the same as for Fig. 5. Inflation is 100 times the log first difference of the price level

international economic relations during the 1920s. Hence, the Bretton Woods system ushered in an era of stable exchange rates around a narrow fluctuation band. Exchange rates could be adjusted but only if sanctioned by a newly formed international organization, the International Monetary Fund, and conditional of reforms undertaken to mitigate the likelihood of a future devaluation. In the immediate postwar era, the focus was on the performance of the current account balance. It would take almost two decades after the end of the 1939–1945 conflict for movements in financial flows to begin and eventually to take center stage in international policy-making.

Despite best-laid plans, the agreement at Bretton Woods put the USA, the dominant post-1945 global economic power, at the heart of an asymmetric system that perpetuated an imbalance exacerbated by two oil price shocks during the 1970s as well as the rising costs of the Vietnam war. Since the global currency of trade was in US dollars, a US current account deficit would be required to finance a surplus elsewhere. Accordingly, any additional shocks to the US economy that would produce continued rises in trade deficits would simply create the conditions that would become unsustainable. The problem was, of course, well known, in the form of the Triffin paradox, even before the eventual collapse of the Bretton Woods system.

Eventually, the pressure on large surplus economies with a desire for low and stable inflation, such as Germany, produced a divergence in inflationary pressure that could no longer be contained. Hence, the exchange rate anchor that was supposed to

hold inflation differentials within some limits was no longer sustainable. Increasingly, advanced economies began to adopt floating exchange rates which, in principle, permitted inflation to be determined entirely based on domestic considerations by acting as a shock absorber against external shocks.

The only exceptions were several countries in the European continent which began preparations for what would eventually become the single currency area, namely, the eurozone. Potential member economies experimented with several variants of exchange rate regimes with constraints on exchange rate variability again as a means of limiting inflation differentials. Success was, at best, also limited. In spite of setbacks, including the 1992 exchange rate crisis which was an almost existential one, politics trumped economics, and the eurozone came into being in 1999 with the new single currency circulating by 2001. From the standpoint of the history of inflation, the eurozone is an interesting example for the treaty that established the phases leading to the introduction of the single currency explicitly required prospective, and future, members to meet an inflation convergence requirement. Indeed, a reading of the history of the creation of the European Monetary System (e.g., James 2016) highlights that the focus of concern on the part of the originators of the idea of a single currency was attention to inflation in the prospective member states.

Once the anchor implicit in the Bretton Woods system could no longer be sustained, the search began for a new anchor that would both placate those worried that some countries might take advantage over others via exchange rates, thereby impacting inflation, and generate a domestic inflation rate that was economically and politically acceptable. The next candidate, a reflection of the ascendancy of monetarism, was to target some indicator of money growth. Unfortunately, the proposed system was more suitable in theory than in practice. In a world where the central bank can control the amount of currency in circulation and bank reserves and has good knowledge of the multipliers linking the monetary base, namely, the sum of currency and reserves, combined with Friedman's dictum cited at the beginning of the chapter, there is a theoretical and empirical link between money growth and inflation. There are other factors as well such as the velocity of circulation and the interest sensitivity of money demand. If the money demand function is relatively stable, then monetary targeting does offer the possibility of inflation control. However, when the money demand function becomes unstable, then the money growth-inflation link cannot be severed.

Beyond the potential instability of money demand, a central bank that implicitly announces an inflation objective based on a projected level of money growth, usually with a tolerance range to account for uncontrollable factors and permit some flexibility, must revise the money growth objective from the point at which a new money growth target must be announced. As a result, because bygones are bygones, the money growth targeting regime had a built-in base drift problem. Moreover, because it can take some time for any tightening of loosening of policy via changes in money growth to eventually influence inflation, this posed additional challenges for central banks attempting to tame inflation.

After only a few years of trying, the monetary targeting regime collapsed. As former Governor of the Bank of Canada, Gerald Bouey famously said: "We did not abandon M1, M1 abandoned us." The unhappy history of monetary targeting in six advanced economies is detailed in Bernanke and Mishkin (1992). Several emerging market economies also claimed to target money growth, but as with the experience of advanced economies, there was little success in adopting this strategy to control inflation (e.g., Mahadeva and Sterne 2000). Indeed, as the monetary targeting experiment showed signs of failure, the number of aggregates that central banks monitored began to mushroom. This did not mean that monetary aggregates could not be useful for potentially controlling inflation (e.g., Friedman and Kuttner 1992; Siklos and Barton 2001), but the link became too unreliable for the monetary authorities to retain the strategy.

With exchange rates and money growth falling out of favor as anchors, the search for a new anchor continued. Meanwhile, central banks were increasingly operating in a world where reserve requirements were no longer the primary instrument used to control the stance of monetary policy. Instead, an interest rate, typically on overnight or short-term funds, was the instrument of choice to influence liquidity in the financial system. The resulting policy rate also had the advantage of being far more transparent than any monetary aggregate. Increases in a policy rate, ceteris paribus, implied a tightening of monetary policy, while a loosening would be signaled via a fall in the policy rate. The policy rate became so ubiquitous as the monetary policy instrument par excellence that it served to launch one of the most successful tools to assess the stance of policy and the performance of central banks, namely, the so-called Taylor rule (Taylor 1993).

Once a link between an interest rate and monetary policy was established, the theoretical device of the Fisher equation provided the necessary link between expected inflation and interest rates. All that was then required was a strategy that guided the public's expectations concerning monetary policy. In essence this is how inflation targeting was born. By becoming accountable for meeting an inflation objective in the medium term, the central bank communicated via interest rate changes its determination to meet this objective. Moreover, if the mandate to achieve a stable inflation rate, possibly with a tolerance band, was agreed to or determined by political officials, then not only was there some added transparency but, perhaps more importantly, there was the additional requirement of accountability for public officials that would be met. Inflation targeting spread quickly as some of the early adopters (e.g., New Zealand and Canada) experience successes aided no doubt by the benign environment that characterized the Great Moderation (e.g., Bernanke 2004; Blanchard and Simon 2001; Siklos 2017).

The literature on inflation targeting is vast. It is worth noting that no central bank or government has recanted its desire to aim for low and stable inflation. Moreover, in spite of the fact that a sizeable portion of the world economy has adopted the inflation-targeting policy strategy, the GFC did mark a shift in emphasis away from a concern over inflation being too high to inflation becoming too low, a point noted earlier. As a result, new adoptions of the inflation-targeting regime fell to a trickle after the GFC as central banks and governments became

focused on financial system stability. If we date the adoption of inflation targeting to the early 1990s in advanced economies and the early 2000s among emerging markets, the inflation-targeting strategy has become one of the most durable, if not successful, monetary policy strategies ever implemented. The fact that the original idea for controlling inflation via direct emphasis on the performance of the price level goes back to at least the 1930s (e.g., Berg and Jonung 1999) demonstrates that some old ideas prove their mettle at the right time and not necessarily when they are first proposed.

This is not to deny that there are serious criticisms of the inflation-targeting regime. Some have complained that the risks of deflation are magnified by a preference for low inflation, whether or not there is a tolerance factor. Others point to what they perceive as central banks becoming obsessed with inflation to the exclusion of a concern for real economic factors. As Mervyn King, former Governor of the Bank of England, once stated (King 1997) there is little evidence that central banks with an inflation-targeting mandate are "inflation nutters." Instead and in line with the Taylor rule formulation, inflation-targeting central banks practice a flexible form of inflation targeting (Svensson 2009).

While the GFC has brought the question of low inflation to the forefront, there is another momentous development that may also have contributed to the low inflation environment over the past decade or so, namely, the rise of China. Beginning in the 1980s and accelerating during the 1990s, China has become the second largest economy in the world. The inflationary experience of China has ranged from high inflation to low inflation with a bout of deflation in between (e.g., Burdekin 2008). Some saw deflation in China which took place during the 1998–2003 period, in contrast with the protracted one in Japan, as supply side driven, or of the "good variety" (e.g., Bernanke 2002), while others have argued that demand factors were also at play (e.g., Burdekin and Siklos 2004; Siklos and Zhang 2010; Borio et al. 2015).

In any event, the combination of a fixed exchange rate and low or falling prices, together with massive increases in output, led the argument that low inflation was being exported to the rest of the world (e.g., Côté and De Resende 2008; Kamin et al. 2004). Even if there is some truth to the argument, its empirical validity must be assessed against a few of other developments especially in advanced economies. First, real commodity prices fell rapidly during the 1980s and then remained fairly stable until the early 2000s. Second, monetary policy was becoming tighter in most advanced economies from the mid-1990s again until the early 2000s, that is, until some economies, notably the USA, were experiencing their first bout of low inflation that threatened to become a deflation no doubt exacerbated by the brief recession of 2001 following the bursting of the tech bubble. Finally, the era in question also coincided with the spread of globalization which also put downward pressure on the growth in consumer prices. We may never know how much of the low inflation was due to the growth of China alone, but the fact that inflation continues to be low, many years after China relaxed its fixed exchange rate regime and saw inflation rise, suggests that its role was at best a passing phase in the history of inflation.

It's a Question of Credibility

In the highly unlikely event that the world returns to a commodity money standard, the current fiat money standard will persist and, with it, the risk of potentially significant movements in inflation. While institutional efforts to limit the range of inflation rates experienced in any given economy have spread around the world, ultimately it is the market and the public's trust in central banks and governments to deliver low and stable inflation rates that remains essential. Put differently, a successful monetary policy strategy must ensure that the monetary authority remains credible.

Unfortunately, there is no universal agreement on what constitutes credibility. Matters are made more difficult because even if the central bank can control inflation in the medium term, there are other sources of shocks that can also influence price developments, including fiscal policy and external shocks, to name just two. Nevertheless, at the core of any definition of credibility, there must be the belief that the central bank's views about future inflationary developments are reasonably accurate. Put differently, forecast errors about future inflation ideally should be small and not persist over time. This is the easiest way for market and the public to anchor their expectations, and the added constraint of an inflation objective for which the central bank is held to account also helps. Bordo and Siklos (2016) review the issues surrounding the measurement of credibility and provide long-run historical evidence.

With one exception (i.e., the US Federal Reserve), it is only fairly recently that central banks have published forecasts of members of their policy committee or their staff. Previously, one had to rely on several competing inflation forecasts and surveys of the public's views about future inflation (e.g., Siklos 2018). In any case, one might have expected that, with the arrival of lower and more stable inflation, forecast errors would at least have decreased over time. Sadly, this is not necessarily the case. Figure 7 plots the forecast errors for the G7 economies since the late 1980s. The forecasts are the ones published by *The Economist* on a monthly basis, and they are representative of private sector forecasts (e.g., Survey of Professional Forecasters, Consensus Economics forecasts). Large forecast errors persist and can be large and not only because of the shock of the GFC. Even the introduction of forecasts by central banks, a development of the 2000s, with the exception of the USA, has not contributed to a reduction in the size of forecast errors. If the errors shown give an indication of the scope for unexpected inflation, then there is still, as 2017 comes to an end, considerable room for improvement. After all, most of the negative effects from inflation originate from the portion that is unexpected. It is also notable that, across the G7, forecasters have been routinely overestimated inflation in the years since the GFC.

The Link Between Inflation and Relative Price Variability

The recent literature on the welfare cost of inflation emphasizes the impact of inflation on the variability of relative prices. Expected and unexpected inflations

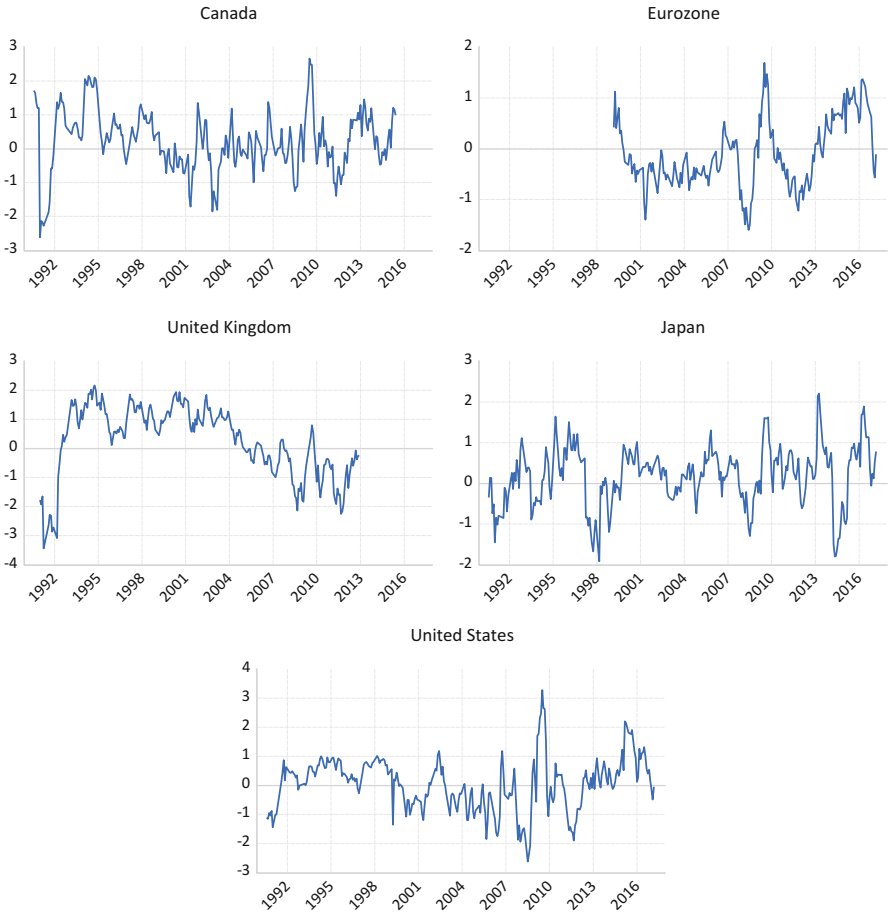


Fig. 7 Inflation forecast errors, G7. Inflation in CPI from June 2017 International Financial Statistics CD-ROM. Current year forecasts converted to fixed horizon (i.e., 1 year ahead) forecasts using the formula in Siklos (2013). Data are monthly from *The Economist*. Shown is the difference between the forecasted and observed inflation rates. A positive number signifies a pessimistic forecast relative to outturns and vice versa for an optimistic forecast

have both been proposed to increase relative price variability and to distort the information content of prices. Theories explaining the welfare cost of inflation through its impact on relative price variability are typically based on menu cost (Sheshinski and Weiss 1977; Rotemberg 1983) and signal extraction models (Lucas 1973; Barro 1976; Hercowitz 1981). Both approaches predict that inflation increases relative price variability and, thus, decreases the informativeness of prices resulting in a decrease of welfare due to the misallocation of resources. While menu cost models emphasize the effects of expected inflation, the focus of signal extraction models is on the cost of unexpected inflation.

In line with the theoretical predictions, several studies have provided evidence in favor of a positive impact of inflation on relative price variability for various countries (Debelle and Lamont 1997; Aarstol 1999; Jaramillo 1999; Konieczny and Skrzypacz 2005). Nevertheless, following Lastrapes (2006), the relationship between US inflation and relative price variability breaks down in the mid-1980s. Fielding and Mizen (2000) and Silver and Ioannidis (2001) find evidence of a negative relationship between relative price variability and inflation for several European countries. A possible explanation for the mixed empirical evidence is that the investigations rely on the linearity of the relative price variability relationship and that the relationship between inflation and relative price variability is more complex than the linear one (among others, Baglan et al. 2016; Nautz and Scharff 2012).

Extraordinary Spells of Inflation: Hyperinflations

While the discussion above dealt with low and moderate inflation rates, we now jump to the extraordinary case of hyperinflation episodes. According to the definition in Cagan's (1956) seminal study, hyperinflations start whenever the monthly inflation rate exceeds 50% and end in the month the inflation rate drops below that amount and remains below for at least 1 year. The definition of hyperinflation periods is arbitrary in nature but widely accepted in the literature. Applying this definition, the updated version of Hanke-Krus' World Hyperinflation Table contains 57 hyperinflation episodes as of December 2016 (Hanke and Krus 2012; Hanke and Bushnell 2016). Modern examples of hyperinflations include Zimbabwe and, since November 2016, Venezuela.

Hyperinflations are destructive events for societies. Hence, an important reason for economists to study hyperinflations is to learn how to prevent their outbreak and their damaging effects. Moreover, Cagan (1956) argues that hyperinflation periods provide the unique opportunity to investigate monetary phenomena largely in isolation from changes in real macroeconomic variables because the dynamics of real factors are small compared to the enormous changes in monetary measures. Hyperinflations can be viewed as the laboratory experiments in the field of macroeconomics. Like earthquakes and hurricanes, hyperinflationary episodes provide to researchers an unusual type of data. These data can be exploited to learn important policy lessons which cannot be drawn relying on data produced under normal economic circumstances.

As a considerable literature exists about hyperinflations, this chapter concentrates on outlining the principal arguments, while the empirical findings are discussed to a lesser extent. The survey starts with the reasons for the origin of hyperinflations in section "Origins." Issues on the money demand function and the discussion on self-generating inflation processes can be found in section "Development." Section "Ending" is devoted to the termination of hyperinflation episodes.

Origins

The most obvious reason for the origin of hyperinflations is the excessive growth in the supply of money. The increases in the supply of money necessary to produce such high inflation rates can only occur in systems with fiat money ruling out hyperinflation periods under metallic currency systems or a gold standard. To address the deeper question why central banks excessively print money, the government budget constraint needs to be considered. Governments being unable to finance their expenditures on goods and services through regular means of borrowing and taxation resort to monetary financing as the remaining source. The central bank's money printing produces seignorage revenues for the government and increases the supply of money which is, if maintained, ultimately responsible for the hyperinflation. While these arguments are fairly straightforward, it is more difficult to answer the question under what circumstances huge budgetary imbalances are possible to trigger hyperinflations.

Among others, Hamilton (1977) connects wars and hyperinflations. Increases of military expenditures lead to high public deficits which are eventually financed by printing paper currency. However, Capie (1986) argues that it is not war itself that leads to hyperinflation because a war between countries raises patriotism and, hence, enables the government to borrow and collect taxes from the public. Rather civil wars, revolutions, or social unrest of different groups within a country triggers hyperinflations. In these internal unstable conditions, the collection of taxes and borrowing from the public become difficult which ultimately leads to monetary financing.

Weak or inexperienced governments are another potential source triggering hyperinflation episodes. In this case, monetary financing results due to the government's inability to collect taxes and their temptation to provide financial transfers and subsidies to different groups of the society in order to strengthen the political base. Lastly, reparation demands and occupation payments from the allied powers are the potential triggers of hyperinflation (Bresciani-Turroni 1937; Kindleberger 1987).

In general, a necessary condition for the emergence of hyperinflation is the lack of independence of the central bank from the government. If the dependence is strong enough, the government has an access to the printing press to finance a huge budget. The drastic increase in money supply is the common feature of all hyperinflations.

Development

Cagan's (1956) money demand function is a special case of Friedman's (1956) demand for money function relating real money balances solely to the expected inflation rate:

$$(m_t - p_t) = \alpha \pi^e_{t+1} + \varepsilon_t. \tag{3}$$

The parameter α is expected to be negative, because a rise in inflation expectations π^e_{t+1} is hypothesized to produce a reduction in the demand for real balances $(m_t - p_t)$. Under the circumstances of hyperinflation, relative changes in real income and interest rates are small compared to changes in nominal money and aggregate prices so that other principal determinants of money demand can be excluded in Eq. (3). One objective of Cagan's study was to demonstrate the existence of a stable money demand function even under severe monetary stress of hyperinflationary periods.

The estimation of the money demand function (3) requires data on the money stock and the price level which are sourced from the countries' official statistics. While data on real money stock are directly observable, the expected inflation rate is per se an unobservable variable. The empirical literature solved this data issue in different ways. Cagan (1956) relied on the adaptive model of expectations formation in which the expected inflation rate is a function of both past inflation and inflation expectation. Solving recursively, the real money stock is a function of a distributed lag of past actual inflation rates.

The conceptual difficulties and the literature on rational expectations have led to an examination of the conditions under which the adaptive expectations process is rational (Muth 1961; Sargent and Wallace 1973; Sargent 1977). Alternatively, Frenkel (1977) proposes a direct measure of inflation expectations that is based on observed data from the forward market for foreign exchange. The premium (discount) on a forward contract for foreign exchange measures the anticipated depreciation (appreciation) of the domestic currency in terms of foreign exchange. Hence, the specification of the demand for money includes the anticipated change in the exchange rate as measured by the forward premium (or discount).

While Frenkel's approach is appealing, the role the forward premium plays in the demand function is not straightforward. In Frenkel's approach, the forward premium serves as a measure of inflation expectations which is the true argument in the money demand function. An alternative interpretation emphasizes the substitution possibilities between domestic and foreign currencies. Under condition of hyperinflation, agents have strong incentives to substitute foreign for domestic currency. With this interpretation, the expected rate of currency depreciation is the true variable in the money demand function. Moreover, both domestic goods and services as well as foreign currencies may be alternatives to holding domestic money. With this interpretation, expected inflation and expected depreciation are both arguments in the money demand function (Abel et al. 1979).

Cagan's framework suffers not only from using the adaptive expectations approach due to its well-known conceptual difficulties and econometric drawbacks, but in addition, – understandable given the econometric knowledge of the time – from reliance on non-stationary time series. Both aspects are taken into account simultaneously by Taylor (1991). Relying on the cointegration technique (Engle and Granger 1987; Johansen 1988), Taylor puts forward a test of the Cagan model which is subject only to very weak assumptions concerning the inflation expectations, i.e., agents' inflation forecasting errors are stationary. Consequently, the actual inflation

rate rather than a measure of the expected inflation rate appears in Eq. (3). If the time series of real money balances and the inflation rate are integrated of order one and are cointegrated, estimates of the parameter of interest are then super-consistent and robust to simultaneity and omitted variables. Moreover, Cagan's conjectures concerning the relative magnitude of real and financial variables during hyperinflation are supported subject to the very weak assumption on agents' expectations that forecasting errors are stationary. The application of cointegration techniques to estimate Cagan's money demand function has the additional advantage that a valid cointegrating relation between real money balances and inflation rate implies the stability of the money demand function.

Relying on Taylor's argument, the actual inflation rate instead a measure of the expected inflation rate is used in Eq. (3). As hyperinflations arise under extreme conditions, like wars, revolutions, and political mismanagement, barriers to the recording and publication of reliable official inflation statistics exist. The lack of reliable price index data can be overcome relying on the principle of purchasing power parity. In its relative form, the inflation rate differential between two countries is equal to the change of the exchange rate. In case a free market for currencies exists and market data are available, changes in the exchange rate are reliable estimates of the country's inflation rate (Frenkel 1976; Hanke and Kwok 2009).

With a few exceptions only, the empirical findings show a negative estimate of the money demand parameter α. Moreover, the findings mostly indicate a cointegrating relation between real money balances and the inflation rate. The empirical evidence varies to some extent depending on the individual country, sample period, data set, treatment of extreme values, and econometric specification. Hence, Cagan's original findings concerning the parameter of interest and the stability of the money demand function are largely confirmed by the application of up-to-date econometric techniques.

Cagan's work was not only the origin of research on the demand for money during hyperinflation, but it was also the beginning of studies on the possibility of self-generating hyperinflations. Following Cagan's argumentation, in a self-generating inflation, a small rise in money supply causes a disproportional increase in the price level. More current treatments utilize the argumentation of a rational bubble in prices. Rational bubbles reflect price-level dynamics which are driven by self-fulfilling expectation independently of fundamental determinants. The appropriate policy to deal with hyperinflations is different depending on the nature of the underlying inflation-generating process. If bubbles are not present, the cause of hyperinflation is solely fundamental. A necessary and sufficient condition to stop hyperinflation is to take firm control of the market fundamentals. If, however, hyperinflation is driven by rational bubbles, restrictive policy actions will not necessarily bring an end to the hyperinflation. Rather, inflation stabilization program has to consist of actions needed to prick the bubble process.

Flood and Garber (1980) propose direct tests for price bubbles. However, as direct tests place strong restrictions both on the relation between inflation and market fundamentals and on the nature of bubbles, the correct specification of the tests is doubtful. As a consequence, nonstructural test approaches are used. Hamilton and

Whiteman (1985) suggest a test of the order of integration in the time series of interest: If the price-level time series exhibits a higher order of non-stationarity than any of the underlying fundamentals, evidence in favor of speculative bubbles in the inflation process is found. The same order of integrability indicates the absence of bubbles in the hyperinflation process.

Hamilton and Whiteman's approach has the deficit that even if the time series of the price level and the fundamental variables are integrated of the same order, a rational speculative bubble cannot be ruled out. Following Diba and Grossman (1988a), the existence of a cointegrating relationship between prices and fundamentals indicates the absence of bubbles in the price process, while in case of no cointegration, bubbles may exist. In essence, if prices are not more explosive than money supply, then it can be concluded that rational bubbles are not present, since they would generate an explosive component to prices. Diba and Grossman (1988b) provide a theoretical foundation for the application of conventional cointegration tests as these rely only on the expansion phase and ignore the burst of a bubble process. Diba and Grossman show that if bubbles crash, they cannot start again which makes continuously growing and bursting bubbles impossible. Nevertheless, allowing for a more general class of periodically collapsing bubbles or temporary bubbles, the application of conventional cointegration tests may lead to a rejection of the existence of rational bubbles even if periodically collapsing or temporary bubbles are present (Evans 1991). In this vein, Funke et al. (1994) put forward a variant of Hamilton's (1990) Markov-switching technique to analyze rational bubbles during hyperinflationary periods.

Glancing through the literature, the existence of speculative bubble components in hyperinflation processes can be ruled out empirically. Hence, the empirical studies find mostly evidence against self-generating hyperinflations. This finding has policy implications for the ending of hyperinflationary periods.

Ending

Accepting the empirical finding of the absence of self-generating components as drivers of hyperinflation, hyperinflationary periods can be stopped by a combination of economic policies affecting fundamentals, like the exchange rate, the public deficit, the money supply, and the measures for wages and prices. A critical element for the success of stabilization programs is the government's credibility and their capability to convince the public to follow stable policies in the future. All hyperinflations end abruptly, although several attempts to stabilize inflation have failed.

Given the approximate validity of the purchasing power parity, the exchange rate is the crucial link between foreign and domestic prices for goods and service. During hyperinflation, the exchange rate depreciates at approximately the same rate as the domestic price level increases. Hence, stabilizing the exchange rate may stabilize domestic prices as well. After a devaluation of the official exchange rate, stabilization programs require to fix the exchange rate at a level which can be defended by the

central bank. A stable exchange rate, in turn, requires a tight fiscal policy (Dornbusch and Fischer 1986).

The most fundamental step in ending hyperinflation is to eliminate the budgetary problems leading to the hyperinflation (Sargent 1982). Successful stabilization programs require a reorganization of government finances and an ending to the reliance of the government to finance its expenditures through the issue of money. Solving budget problems is a precondition to gain control over monetary growth so that hyperinflationary periods can be stopped by ending monetary accommodation. The step consists of an increase in political independence of the central bank to refuse money creation. Moreover, improvement of the budget situation requires fiscal austerity and a rise of tax revenues. This involves a comprehensive tax reform and the elimination of government subsidies (Morales and Sachs 1989).

The above considerations concentrate on an orthodox mix of fiscal and monetary policies. The combination of orthodox policies with measures aimed at wages and prices is called heterodox stabilization policies to ending hyperinflations (Bruno et al. 1988). The heterodox approach argues that tight fiscal and monetary policies are not sufficient to stop hyperinflation. Heterodox stabilization programs justify the direct interventions into wage and price settings with the argument of a large component of inflation inertia emanating mainly from long-term labor contracts and wage indexation based on past inflation. Wage and price controls at the beginning of the reform interrupt inflation inertia and bring down the public inflation expectations. In addition, peoples' confidence in the stabilization program is promoted.

Conclusion

In an important sense, the present survey reaches a conclusion similar to the one highlighted by Laidler and Parkin (1975) over 40 years ago. Inflation, if fully anticipated, produces modest social costs. We are no closer to knowing what is "optimal" inflation except that low and stable inflations come closest to reducing the loss of purchasing power of money. Because prices of goods and services incorporate elements that are difficult to measure precisely, we cannot even be sure what the actual level of inflation really is. Hence, what is deemed low may well differ across countries and across time. Nevertheless, avoiding inflation is not only desirable because it represents a form of taxation without representation, but in theory at least, low and stable inflation ought to be more easily forecasted, thereby reducing the likelihood of large and persistent forecast errors.

The common feature of all hyperinflations is a drastic increase in the money supply which results from the need to finance huge budget deficits. Hyperinflations emerge only in regimes of fiat money. Lack of independence of the central bank which can be forced to finance the government budget deficit via money printing is another precondition. Moreover, hyperinflations emerge during or in the aftermaths of war, civil war, revolution, or social unrest. Even under the enormous monetary stress, money demand functions are stable. During hyperinflationary periods,

speculative bubbles can be ruled out. Stopping hyperinflations requires a combination of policies concerning the exchange rate, public budget, money supply, and, in some cases, direct measures aimed at wages and prices as well as public confidence in the package of measures.

In terms of monetary policy, generally inflation performance represents a critical determinant of central bank credibility. Once lost, credibility is difficult to reestablish. Unfortunately, low and stable inflations need not imply that it is easier to forecast. In spite of great advances in theory and empirical evidence, the economics profession remains far away from providing a conclusive account of the dynamics of inflation. It is hoped that, by the time the next survey of inflation is written, we will have successfully addressed this issue.

References

Aarstol M (1999) Inflation, inflation uncertainty, and relative price variability. South Econ J 66:414–423

Abel A, Dornbusch R, Huizinga J, Marcus A (1979) Money demand during hyperinflation. J Monet Econ 5:97–104

Baglan D, Yazgan M, Yilmazkuday H (2016) Relative price variability and inflation: new evidence. J Macroecon 48:263–282

Barro R (1976) Rational expectations and the role of monetary policy. J Monet Econ 2:1–32

Benati L (2008) Investigating inflation persistence across monetary regime. Q J Econ 123:1005–1060

Benati L, Lucas R, Nicolini J, Weber W (2016) Long-run money demand redux. Working paper

Berg C, Jonung L (1999) Pioneering price level targeting: the Swedish experience, 1931–1937. J Monet Econ 43:525–551

Bernanke B (2002) Deflation: making sure "It" doesn't happen here. Remarks before the National Economist Club, Washington, DC, 21 November 2002

Bernanke B (2004) The great moderation. Speech at the Meeting of the Eastern Economic Association, Washington, DC, February 20

Bernanke B (2008) Outstanding issues in the analysis of inflation. The Federal Reserve Bank of Boston's 53rd Annual Conference, 9 June

Bernanke B, Mishkin F (1992) Central Bank behavior and the strategy of monetary policy. In: Blanchard OJ, Fischer S (eds) NBER macroeconomics annual 1992, vol 7. The MIT Press, Cambridge, MA, pp 183–238

Blanchard O, Simon J (2001) The long and large decline in U.S. output volatility. Brook Pap Econ Act 1:135–164

Bordo M (1981) The classical gold standard: some lessons for today. Econ Rev Fed Reserv Bank St Louis 63:2–17

Bordo M (1999) The gold standard and related regimes: collected essays. Cambridge University Press, Cambridge/New York

Bordo M, Orphanides A (2013) The great inflation: the rebirth of modern central banking. University of Chicago Press, Chicago

Bordo M, Schwartz A (eds) (1984) A retrospective on the classical gold standard. Chicago University Press, Chicago

Bordo M, Siklos P (2016) Central Bank credibility: an historical and quantitative exploration. In: Bordo M, Eitrheim Ø, Flandreau M, Qvigstad J (eds) Central banks at a crossroads. Cambridge University Press, Cambridge, MA, pp 62–144

Bordo M, Siklos P (2018) Central Banks: evolution and innovation in historical perspective. NBER working paper 23847, to appear in Sveriges Riksbank 350 Years: a Central Bank in a World of Central Banks. Cambridge University Press, Cambridge, MA, forthcoming
Borio C, Erdem M, Filardo A, Hofmann B (2015) The costs of deflation: a historical perspective. BIS Q Rev 20:31–54
Bresciani-Turroni C (1937) The economics of inflation: a study of currency depreciation in post-war Germany, 1914–1923. Allan & Unwin, London
Bruno M, Di Tella G, Dornbusch R, Fischer S (1988) Editors, inflation stabilization: the experiences of Israel, Argentina, Brazil, Bolivia, and Mexico. Cambridge University Press, Cambridge
Burdekin R (2008) China's monetary challenges. Cambridge University Press, Cambridge
Burdekin R, Siklos P (2004) Editors, deflation: current and historical perspectives. Cambridge University Press, Cambridge
Cagan P (1956) The monetary dynamics of hyperinflation. In: Friedman M (ed) Studies in the quantity theory of money. University of Chicago Press, Chicago, pp 25–117
Capie F (1986) Conditions in which rapid inflation has appeared. Carnegie Rochester Conf Ser Public Policy 24:115–168
Charbonneau K, Evans A, Sarker S, Suchanek L (2017) Digitalization: a review of the literature. Bank of Canada Staff Analytical Note 2017-20
Ciccarelli M, Mojon B (2010) Global inflation. Rev Econ Stat 92:524–535
Côté D, De Resende C (2008) Globalization and inflation: the role of China. Bank of Canada working paper 2008-12
Debelle G, Lamont O (1997) Relative price variability and inflation: evidence from US cities. J Polit Econ 105:132–152
Diba B, Grossman H (1988a) Explosive rational bubbles in stock prices? Am Econ Rev 78:520–530
Diba B, Grossman H (1988b) Rational inflationary bubbles. J Monet Econ 21:35–46
Dornbusch R, Fischer S (1986) Stopping hyperinflations past and present. Weltwirtschaftliches Arch 122:1–47
Eichengreen B (1992) Golden fetters. Oxford University Press, Oxford
Engle R, Granger C (1987) Co-integration and error correction: representation, estimation and testing. Econometrica 55:251–271
Evans G (1991) Pitfalls in testing for explosive bubbles in asset prices. Am Econ Rev 81:922–930
Fielding D, Mizen P (2000) Relative price variability and inflation in Europe. Economica 67:57–78
Flood R, Garber P (1980) Market fundamentals versus price-level bubbles: the first test. J Polit Econ 88:745–770
Frenkel J (1976) A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. Scand J Econ 78:200–224
Frenkel J (1977) The forward exchange rate, expectations, and the demand for money: the German hyperinflation. Am Econ Rev 67:653–670
Friedman M (1956) The quantity theory of money – a restatement. In: Friedman M (ed) Studies in the quantity theory of money. University of Chicago Press, Chicago, pp 3–21
Friedman M (1970) The counter-revolution in monetary theory. Institute of Economic Affairs, London
Friedman B, Kuttner K (1992) Money, income, prices, and interest rates. Am Econ Rev 82:472–492
Funke M, Hall S, Sola M (1994) Rational bubbles during Poland's hyperinflation implications and empirical evidence. Eur Econ Rev 38:1257–1276
Gorodnichenko Y, Tavalera O (2017) Price setting in online markets: basic facts, international comparisons, and cross-border integration. Am Econ Rev 107:249–282
Hamilton E (1977) The role of war in modern inflation. J Econ Hist 37:13–19
Hamilton J (1990) Analysis of time series subject to changes in regime. J Econ 45:39–70
Hamilton J, Whiteman C (1985) The observable implications of self-fulfilling expectations. J Monet Econ 11:247–260
Hanke S, Bushnell C (2016) Venezuela enters the record book. The 57th Entry in the Hanke-Krus World Hyperinflation Table, working paper

Hanke S, Krus N (2012) World hyperinflations. Working paper
Hanke S, Kwok A (2009) On the measurement of Zimbabwe's hyperinflation. Cato J 29:353–364
Hercowitz Z (1981) Money and the dispersion of relative prices. J Polit Econ 89:328–356
James H (2016) Making the European monetary union. Harvard University Press, Cambridge, MA
Jaramillo C (1999) Inflation and relative price variability: reinstating parks' results. J Money Credit Bank 31:375–385
Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254
Kamin S, Marazzi M, Schindler J (2004) Is China exporting deflation? International Finance Discussion Paper 791, Board of Governors of the Federal Reserve System
Keynes J (1924) A tract on monetary reform. Macmillan, London
Kindleberger C (1987) Marshall plan days. Allen & Unwin, New York
King M (1997) The inflation target five years on. Speech at the London School of Economics, 29 October
Konieczny J, Skrzypacz A (2005) Inflation and price setting in a natural experiment. J Monet Econ 52:621–632
Laidler D, Parkin M (1975) Inflation: a survey. Econ J 85:741–809
Lastrapes W (2006) Inflation and the distribution of relative prices: the role of productivity and money supply shocks. J Money Credit Bank 38:2159–2198
Lucas R (1973) Some international evidence on output–inflation tradeoffs. Am Econ Rev 63:326–334
Lucas R, Nicolini J (2015) On the stability of money demand. Federal Reserve Bank of Minneapolis working paper 718, September
Mahadeva L, Sterne G (2000) Monetary policy frameworks in a global context. Routledge, New York
Meltzer A (2009) Inflation nation. New York Times. May 4
Morales J, Sachs J (1989) Bolivia's economic crisis. In: Sachs J (ed) Developing country debt and economic performance. Country studies – Argentina, Bolivia, Brazil, Mexico. NBER. University of Chicago Press, Chicago
Muth J (1961) Rational expectations and the theory of price movements. Econometrica 29:315–335
Nautz D, Scharff J (2012) Inflation and relative price variability in the euro area: evidence from a panel threshold model. Appl Econ 44:449–460
Pang K, Siklos P (2016) Macroeconomic consequences of the real-financial Nexus: imbalances and spillovers between China and the US. J Int Money Financ 65:195–212
Parkin M (1980) Oil push inflation? Banca Nazionale Lav Rev 133:163–185
Rotemberg J (1983) Aggregate consequences of fixed costs of price adjustment. Am Econ Rev 73:433–436
Sargent T (1977) The demand for money during hyperinflation under rational expectations. Int Econ Rev 18:59–82
Sargent T (1982) The ends of four big inflations. In: Hall R (ed) Inflation: causes and effects. University of Chicago Press, Chicago
Sargent T, Wallace N (1973) Rational expectations and the dynamics of hyperinflation. Int Econ Rev 14:328–350
Sheshinski E, Weiss Y (1977) Inflation and costs of price adjustment. Rev Econ Stud 44:287–303
Siklos P (1999) Inflation target design: changing inflation performance and persistence in industrial countries. Rev Fed Reserv Bank St Louis 81:47–58
Siklos P (2013) Sources of disagreement in inflation forecasts: an international empirical investigation. J Int Econ 90:218–231
Siklos P (2017) Central banks into the breach: from triumph to crisis and the road ahead. Oxford University Press, Oxford
Siklos P (2018) What has publishing inflation forecasts accomplished? Central Banks and their competitors. In: Mayes D, Siklos P, Sturm J-E (eds) Handbook of the economics of central banking. Oxford University Press, Oxford, forthcoming

Siklos P, Barton A (2001) Monetary aggregates as indicators of economic activity in Canada: empirical evidence. Can J Econ 34:1–17
Siklos P, Zhang Y (2010) Identifying the shocks driving inflation in China. Pac Econ Rev 15:204–223
Silver M, Ioannidis C (2001) Intercountry differences in the relationship between relative price variability and average prices. J Polit Econ 109:355–374
Svensson L (2009) Inflation targeting. In: Durlauf S, Blume L (eds) The new Palgrave dictionary of economics, 2nd edn. Palgrave Macmillan, London, pp 1237–1302
Taylor M (1991) The hyperinflation model of money demand revisited. J Money Credit Bank 23:327–351
Taylor J (1993) Discretion versus policy rules in practice. Carnegie Rochester Conf Ser Public Policy 39:195–214
Woodford M (2003) Interest and prices. Princeton University Press, Princeton

Deflations in History

41

Richard C. K. Burdekin

Contents

Abstract

Even though experiences with falling prices have been rare during the postwar period, deflation was widespread during the 1930s, and recorded historical episodes extend back to ancient and medieval times. With deflation having resurfaced as a major policy concern in the years following the global financial crisis, this chapter compares the properties of the earlier deflations with more recent episodes in both Europe and Asia. In focusing upon the determinants of deflation, its impact upon the economy as a whole, and the role of monetary policy, we see that even though deflation remains, in essence, a monetary phenomenon, combatting its effects, remains far from straightforward. A striking feature of the twenty-first-century deflations, for example, has been the discrepancy between consumer and producer price movements seen after 2008 that occurred in conjunction with sharp declines in commodity prices. Policymakers both past and present have had to contend with a variety of downward pressures on the money supply as well as complications arising from supply shocks and other negative forces.

R. C. K. Burdekin (✉)
Robert Day School of Economics and Finance, Claremont McKenna College, Claremont, CA, USA
e-mail: Richard.Burdekin@claremontmckenna.edu

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2_49

Keywords
Deflation · Gold standard · Monetary policy · Price scissors · Supply shocks

> Hence followed a scarcity of money, a great shock being given to all credit, the current coin too, in consequence of the conviction of so many persons and the sale of their property, being locked up in the imperial treasury or the public exchequer . . . The facilities for selling were followed by a fall of prices, and the deeper a man was in debt, the more reluctantly did he part with his property, and many were utterly ruined . . . (Roman historian Tacitus' description of the events in 33 AD, quoted in Temin 2013, pp. 141–142)

Introduction

Experiences with falling prices have been rare since the 1930s. However, deflation not only has a long history going back to ancient and medieval times but also has resurfaced as a major policy concern in the years following the global financial crisis. This chapter compares the properties of the earlier deflations with more recent episodes in both Europe and Asia, focusing upon the determinants of the deflation, its impact upon the economy as whole, and the role of monetary policy. Co-movement between different price series cannot be taken for granted, and the post-2008 Chinese, Japanese, and Spanish cases have been marked by a large discrepancy between consumer and producer price movements. The role of supply-side factors, in particular the sharp oil price declines seen after the global financial crisis, must be considered in addition to purely monetary factors. Indeed, whereas earlier deflations like those of the 1930s occurred in the face of monetary contraction, the post-2008 experience suggests that even more expansionary policy cannot be counted on to offset downward pressures on the price level. The years since the global financial crisis have also been marked by a record disconnect between the stance of central bank policy and the rates of expansion achieved with regard to the actual amount of money in circulation. Although deflation remains, in essence, a monetary phenomenon, the historical record confirms that combatting its effects with monetary policy remains a tricky proposition. Policymakers both past and present have had to contend a variety of downward pressures on the money supply and complications arising from exchange rate constraints, supply shocks, and other negative forces.

In the days of commodity money, bullion shortages in fifteenth-century Europe were associated with deflation that ended only with "a new wave of devaluations combined with heavy emissions of . . . copper currency" (Day 1978, p. 47). The collapse of the Roman Empire represented another negative monetary shock as coinage supplies became dependent upon the Germanic tribes' ability to operate the old Roman mints (Horesh 2014, p. 86). Similarly, the deflations of the late nineteenth century in the United Kingdom and United States occurred when growing scarcity of gold put downward pressure on the money supply and prices alike under the classical gold standard (Friedman 1992). The most well-known historical

episode of deflation during the Great Depression of the 1930s also occurred in conjunction with monetary contraction, including the astonishing one-third drop in the US money supply between 1930 and 1933 (Friedman and Schwartz 1963). More recently, monetary tightening by the Bank of Japan in 1989 sets the stage for Japan's stock market crash and ongoing "Lost Decade" and subsequent deflation concerns. Even briefer, and less well-known, deflationary episodes like that experienced by the Confederate States of America in early 1864 can be readily linked to monetary contraction. In this particular case, the April 1864 currency reform that reduced the money supply by one-third in the Eastern Confederacy induced a brief deflation that stood in stark contrast to the extensive inflationary trend both before and after this event (Burdekin and Weidenmier 2001).

In contrast to these earlier episodes of deflation, worldwide concerns with deflation emerged in the aftermath of the global financial crisis even in the face of substantial monetary easing. In the case of the United States, for example, the Federal Reserve's policy response seemingly could not have been more different from the early 1930s case. The Federal Reserve quadrupled the size of the monetary base between 2008 and 2016 and, like many other central banks around the world, engaged in successive rounds of quantitative easing after the onset of the global financial crisis. However, even the maintenance of near zero interest rate targets by the Federal Reserve and European Central Bank was not enough to stop inflation falling close to zero in many countries and actually dropping into negative territory in such Eurozone economies as Greece and Spain. Deflation concerns also reemerged in both China and Japan notwithstanding expansionary central bank policy and currency depreciation. Indeed, some central banks, like Sweden's Riksbank and the Swiss National Bank, were prompted to pursue not only quantitative easing but also negative interest rates. In Sweden's case, this unprecedented move was in part a reaction to the strengthening of the Swedish krona against the euro following the European Central Bank's own policy loosening (Spence 2015). This is in a small way reminiscent of the experience of the 1930s, when waves of competitive devaluations led to a "race to the bottom" after the abandonment of the international gold standard.

Besides the stance of monetary policy, another apparent contrast between earlier deflationary episodes and the post-2008 experience concerns the relationship between consumer and producer prices. Although not necessarily moving in lock-step, consumer price declines generally went hand in hand with declines in producer prices. A different element has emerged in the post-2008 Chinese case, however, with mild consumer price deflation being accompanied by substantially greater deflation in producer prices. This gap between consumer and producer price movements approached 8% in 2016, a phenomenon never present in any of the earlier deflation episodes. Moreover, growing gaps between consumer and producer prices were also seen in Japan and Spain after the global financial crisis. Again mild consumer price deflation was accompanied by much sharper declines in producer prices. The Spanish deflation began only in 2013 and, like the Greek case, reflects pressures associated with membership in the Eurozone and inability to enhance competitiveness via currency depreciation. Complicating the Greek situation, of course, is the additional role played by ongoing social unrest and

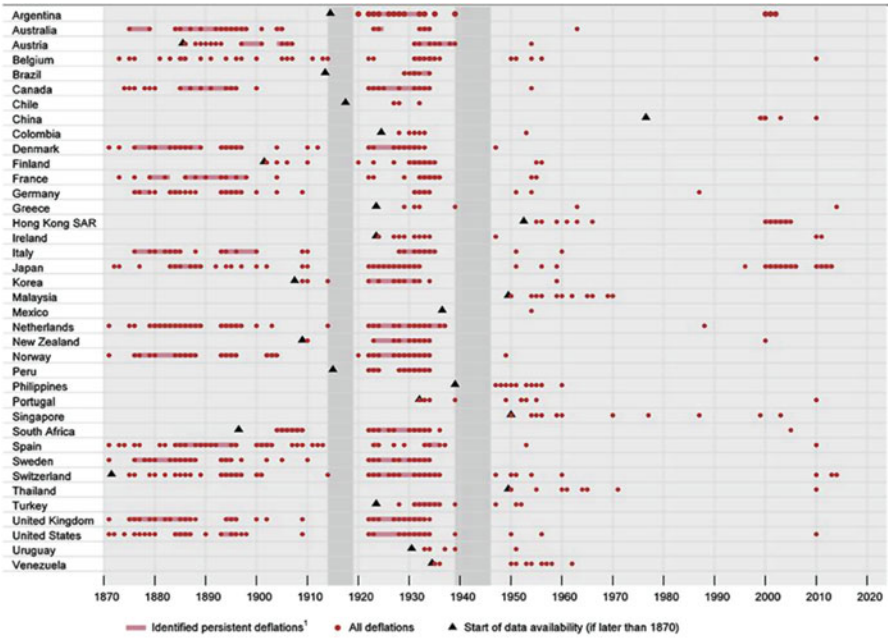


Fig. 1 Timeline of deflations

other long-standing problems – leading Hetzel (2014, p. 275) to conclude: "Greece was dealt a bad hand [but] it has also played poorly the hand it was dealt."

The Historical Record

Deflationary episodes are heavily concentrated in the Great Depression of the 1930s and the latter part of the nineteenth century. Figure 1, which reproduces the deflation timeline depicted in Borio et al. (2015, p. 34), makes clear the contrast between the number of countries experiencing deflation during those earlier time periods as compared to the very limited postwar deflationary episodes. Burdekin and Siklos' (2004) international evidence concerning the frequency of deflationary episodes in 20 countries further suggests that, although deflations were frequent prior to 1945 in many countries, with the exception of the United Kingdom and France, they did not occur as often as in the United States. Over 200 years of consumer price data, episodes of inflation and deflation are almost evenly divided if the definition of deflation includes zero inflation. Otherwise, years with falling prices account for

roughly 25% of the total. When wholesale prices are used, regardless of whether cases of zero or near zero inflation are excluded, the frequency of deflationary episodes is essentially unaffected. There remain only very slight differences between the degree of inflation persistence for the full 200 years sample and the portion of the sample comprised of deflationary episodes (including cases of zero inflation). However, Burdekin and Siklos (2004) find, via simple autoregressive estimation, that inflation is easier to predict, based on its past history, than is deflation. Such inability to anticipate price drops certainly appears to have been evident during the Great Depression, when futures markets continued to bet on commodity price increases even as prices of corn, cotton, oats, and rye were falling by 35% or more during the early 1930s (Hamilton 1992).

Although the deflations of the 1930s and 1890s were themselves accompanied by monetary contraction, this did not so much reflect policy intent as the confines of the exchange rate constraint under the gold standard. In the late 1890s, gold became scarcer in the absence of new gold discoveries, and, for currencies linked to gold, greater purchasing power for gold could be achieved only via a decline in the prices of goods and services. The deflationary pressures did not ease until the invention of the cyanide process, and new gold discoveries in South Africa, provided for renewed expansion in the world's gold supply (Friedman 1992). Meanwhile, in the early 1930s, central banks were largely handcuffed by the fact that unilateral monetary expansion would erode the balance of payments, triggering a gold outflow that would eventually render infeasible the existing parity with gold. Only after the abandonment of the gold standard, beginning with the United Kingdom in 1931 and later the United States in 1933, was there any real scope for sustained expansionary policy (c.f., Eichengreen 1992).

Even unilateral exchange rate commitment can have similarly restrictive effects, leading Argentina to abandon its fixed exchange rate, and float its currency, in December 2001 after being mired in a 4-year recession. At that time, Domingo Cavallo, an architect of Argentina's currency board system that was enacted in 1990 to help quell hyperinflation, acknowledged the extent of the policy dilemma: "I always thought that stopping falling prices would be easy ... But I realize now that stopping hyperinflation was the easy part" (see Luhnow and Druckerman 2001, p. A6). In general, the most effective way to fight deflation is always going to be to keep it from starting or at least to keep it from lasting long enough to become entrenched in expectations. With reference to today's Eurozone, Eichengreen and Temin (2013) foresaw that the weaker nations' attempts to remain part of the single currency were likely to give rise to similarly dangerous constraints with austerity measures being forced onto already weakening economies. Spain's post-2013 deflation is far from the only issue here, with major concerns surrounding the other members of the so-called "PIIGS" group (Greece, Ireland, Italy, and Portugal).

Even though evidence of a link between deflationary episodes and overall declines in economic activity may be less than clear-cut outside the Great Depression (Borio et al. 2015), one reason to fear deflation in consumer prices is that, if it is expected that such price declines will continue in the future, there is an incentive to delay purchases. This then leads to a further reduction in aggregate demand, putting further downward

pressure on prices and suggesting that deflation could be at least partially self-sustaining. Reinforcing the contractionary effects of postponed consumption, the Mundell-Tobin effect implies that deflation encourages more money hoarding owing to rising real rates of interest. Mundell (1963) shows that, with the money rate of interest responding less than proportionately to innovations in inflation, even fully anticipated inflation would reduce real money balances and reduce wealth. The converse that deflation adds to the real value of money holdings and increases total real wealth is seen by Tobin (1965) as potentially having the destabilizing effect of diverting savings away from capital formation. Although the implied decline in capital's share in total wealth, and associated rise in its yield, would eventually reverse this process, "they may do so too little and too late" (Tobin 1965, p. 684).

Fisher's (1933) debt-deflation mechanism represents another potential threat, whereby deflation adds to real debt burdens, in turn triggering bankruptcies and further declines in economic activity. Although typically thought of in terms of the 1930s experience, a precursor appears to have been seen in ancient Rome in 33 AD. At that time, the collapse of a housing boom forced senators to sell land to raise money to pay off their creditors. This, in turn, not only caused land prices to fall even more but also raised debt burdens in relation to senators' property values and exacerbated the decline in overall spending (Temin 2013, p. 142). The dependence on land as collateral itself naturally has parallels with Japan's "Lost Decade," with falling land values playing a key role in triggering defaults on loans after 1989.

Balance sheet effects can be another important factor (Bernanke 1983). When debtors who default forfeit their assets to banks, plunging prices of the forfeited assets then hurt bank balance sheets and threaten the solvency of the banking sector. If banks curtail lending as a result, firms dependent on bank credit will, in turn, face a credit crunch as seen in 2008. If left unchecked, this could easily lead to further production cutbacks, intensifying the deflationary spiral. Firms' borrowing difficulties will be further exacerbated as declines in asset prices reduce the value of their loan collateral. In the US case, the unprecedented debt buildup during the 1920s was followed by a sudden shift from a stable price environment to a deflationary environment at the beginning of the 1930s. This meant that firms and households faced an unexpected rise in their debt service costs at the very time that their ability to fund their debt diminished. There is little doubt that the unanticipated nature of the rising real debt burden at the beginning of the 1930s fueled the operation of a debt-deflation process in the United States at this time (Fackler and Parker 2005).

Although Japan's post-1990 problems never approached those experienced during the Great Depression, balance sheet and collateral concerns again emerged in the aftermath of the initial asset price collapse. The dramatic fall in land prices seen in Japan after 1989 was of particular concern because land formed the collateral for so many loans. This was followed by an ongoing series of net worth shocks for both financial intermediaries and entrepreneurs that lowered lending rates and investment activity alike. Hirakata et al. (2016) identify overwhelmingly negative values for net worth shocks in Japan that persist from the early 1990s through the aftermath of the global financial crisis. The failure of repeated attempts at

expansionary monetary and fiscal policy to reverse the negative trends partly reflects the fact that Japanese banks were seeking to bolster their own balance sheets, which had been hurt not only by nonperforming loans but also by the sharp drop in the market value of their equity holdings. Indeed, the sharp drop in the Japanese money multiplier in the 1990s (Hutchison 2004) served as somewhat of a precursor for the massive drop in the US money multiplier after 2008. The banks' reduced willingness to lend and to circulate the new money being created by the Bank of Japan meant that these funds were often simply held within the banking system, thereby doing little to fuel new spending and combat ongoing deflation.

Delays in bank recapitalizations, as well as the lack of major structural reforms, undoubtedly are among the factors accounting for the length of the downturn in Japan. Although positive technology shocks may have played some part in the downward pressure on prices (Saxonhouse 2005), the overall performance of the Japanese economy since the beginning of the "Lost Decade" implies that any "good deflation" was outweighed by the bad in this case. In this regard, Hoshi and Kashyap (2015) point to some worrying parallels with the post-global financial crisis situation in countries like France, Italy, and Spain. The emergence of declining inflation, and outright deflation in the case of Spain, has only added to the difficulties. With falling inflation adding to debt burdens, there is the very real possibility of a debt-deflation spiral emerging at the national government level. Indeed, some estimates suggest that, should inflation remain at 1% or lower, this could be enough to make Italian government debt unsustainable (Stewart 2013).

The situation may actually be even worse than it appears given that upward biases in conventional consumer indexes mean that a measured inflation rate of 1% or less could well be masking an actual deflationary trend – notwithstanding the improvements made in response to the recommendations of the Boskin Commission (Boskin 2005). With Spain having experienced declining consumer prices since 2013, this put the Spanish government's already heavily extended fiscal position (as with other PIIGS countries like Greece) at substantial risk. Public debt not only rose by approximately 60% of GDP in the face of the global financial crisis but also continued to increase afterward (Munevar 2016), leaving Spain with the unattractive combination of one of the highest nominal debt burdens in the Eurozone plus outright deflation in both consumer and producer prices.

The apparent disconnect between the extent of the expansionary policies since 2008 and the continued deflationary concerns may be partly due to structural problems. But it is also important to take into account the extent to which expansionary central bank policies have translated into actual monetary expansion. Just as Japan was plagued by a sharp decline in the money multiplier in the 1990s, European Central Bank and Federal Reserve expansion of the monetary base has been offset by sharp decreases in commercial bank lending rates. Table 1 compares the extent of base money issuance in China, Japan, the Eurozone, the United Kingdom, and the United States to overall monetary expansion over the 2008–2016 periods. Monetary base expansion exceeded 100% in all five cases, ranging from a low of 103.1% for the European Central Bank to a high of 402% for the Federal Reserve. Yet, except for the 212.7% increase seen in China, actual growth in the overall money supply

Table 1 Comparative monetary trends, 2008–2016

Money supply	China	Japan	Eurozone	United Kingdom	United States
Total monetary base growth from September 2008 to January 2016	147.5	284.6	103.1	402.0	316.9
Total bank money growth from September 2008 to January 2016	235.5	−5.9	8.9	−0.9	−6.5
Overall money supply growth from September 2008 to January 2016	212.7	20.3	18.0	16.2	9.9

Note: Monetary base growth refers to high-powered money produced by central banks; bank money is the remaining portion of the money supply produced by commercial bank lending; and overall money supply growth is in terms of broad money (M3 or equivalent)
Source: Hanke (2016)

otherwise ranged between just 9.9% for the United States and 20.3% in Japan because of a huge drop in the velocity of money.

The rise in Chinese bank lending for the most part involved funds going to the government's state-owned enterprises, with the surging real estate markets serving as another catalyst for loan expansion. Outside of China, bank lending stagnated and commercial banks largely sat on the new funds generated by central bank open market purchases. Hanke (2016) blames this phenomenon on overly zealous tightening of bank regulation and supervision, arguing that these policies have been "ultra-tight and procyclical." Regulatory stringency remains an understandable reaction to the excesses of the earlier 2000s, however, and there are obvious parallels between the Dodd-Frank Act in the United States and the earlier imposition of the Glass-Steagall Act in 1933. There is also the question of whether bank lending has been held back not by excessive regulation but rather by continued weakness in the commercial banks own balance sheets. Although substantial recapitalization was undertaken in the United States, the European banks still seemed to stand in need of more recapitalization rather than being in a position to benefit from regulatory rollbacks alone. Bank lending aside, the data do call attention to the fact that base money issuance by the European Central Bank has lagged well behind its counterparts, however. The fact that its quantitative easing was roughly one-third the scale of the Federal Reserve's may not be the only reason for relatively greater deflationary concerns in the Eurozone than in the United States, but this could hardly have helped matters.

The actual extent of deflation concerns should itself not be assessed on the basis of consumer prices alone. There is always the possibility of relative price changes wherein some components of the price level fall, while others keep rising. Moreover, even though discussions about the role of monetary policy tend to focus on the behavior of headline price indices, deflation in less-heralded areas such as producer prices should not be ignored given the potential for eventual transmission to consumer prices. Wage rigidities add to the dangers associated with any such price drop given that pushing the real wage above equilibrium must sooner or later give rise to unemployment. This concern can be contrasted with the old costs of production view

that saw falling prices as a sign of lower production costs and a reflection of improvements in transportation and technology. In this vein, Dickey (1977) argues that the US deflation of 1869–1896 was primarily of the "good" variety since relative price changes, profit expectations, and bond yields all suggest that supply-side influences dominated price movements over demand-side effects. In analyzing both the Canadian and US experiences prior to World War I, Bordo and Redish (2004) argue for a more mixed picture, however. In particular, they reject any simple demarcation between "good" and "bad" deflation and attribute the price declines to a combination of both negative money supply shocks and positive supply shocks.

Price-Level Targeting

As a possible means of achieving price-level stability without the mechanical constraints incumbent under gold standard or fixed exchange rate arrangements, price-level targeting has been proposed. In 1898 Swedish economist Knut Wicksell made a pathbreaking case for price-level stabilization in his address to Sweden's *Nationalekonomiska Förening* (reprinted in English in Wicksell 1958, pp. 67–89). Wicksell's proposal actually relied upon interest-rate targeting and, as noted by Jonung (1979), could not have been implemented under the gold standard because adjusting interest rates to their noninflationary level would have required independent control of the money supply. After Sweden left the gold standard for good on September 27, 1931, however, "the norm that Wicksell presented at the turn of the century became, some thirty years later, the official foundation for Swedish monetary policy" (Jonung 1979, p. 468). Meanwhile, Irving Fisher (1913, 1920) argued in favor of a monetary regime based on a standard of commodity prices. Whereas price-level targeting could offer a means of signaling the policymaker's commitment to lowering real interest rates even after nominal interest rates have already been cut to zero, the postwar era has seen central banks adopt inflation targeting instead Following its enshrinement in the Reserve Bank of New Zealand Act 1989, the Bank of Canada, the Bank of England, the European Central Bank, and many others adopted formal inflation targets – with the Federal Reserve joining the party in January 2012.

In an inflationary world, inflation targeting will generally be less restrictive than price-level targeting because it does not require all past price increases to be reversed. On the other hand, in a deflationary world, price-level targeting implies a stronger commitment to reversing downward pressures because of the need for additional expansion to reverse past price declines. As shown in Fig. 2 (depicting the reverse case of the positive inflation shock illustrated by Hatcher and Minford 2014), price-level targeting implies a more expansionary response to a deflationary shock so as to return the price level to its original trajectory. This means that there is mean reversion under a price-level target, which, in turn, implies less uncertainty about future prices than with an inflation target. The scope for providing a clearer signal to wage setters by firmly anchoring price expectations must be balanced against the risk of producing unemployment in the face of the required

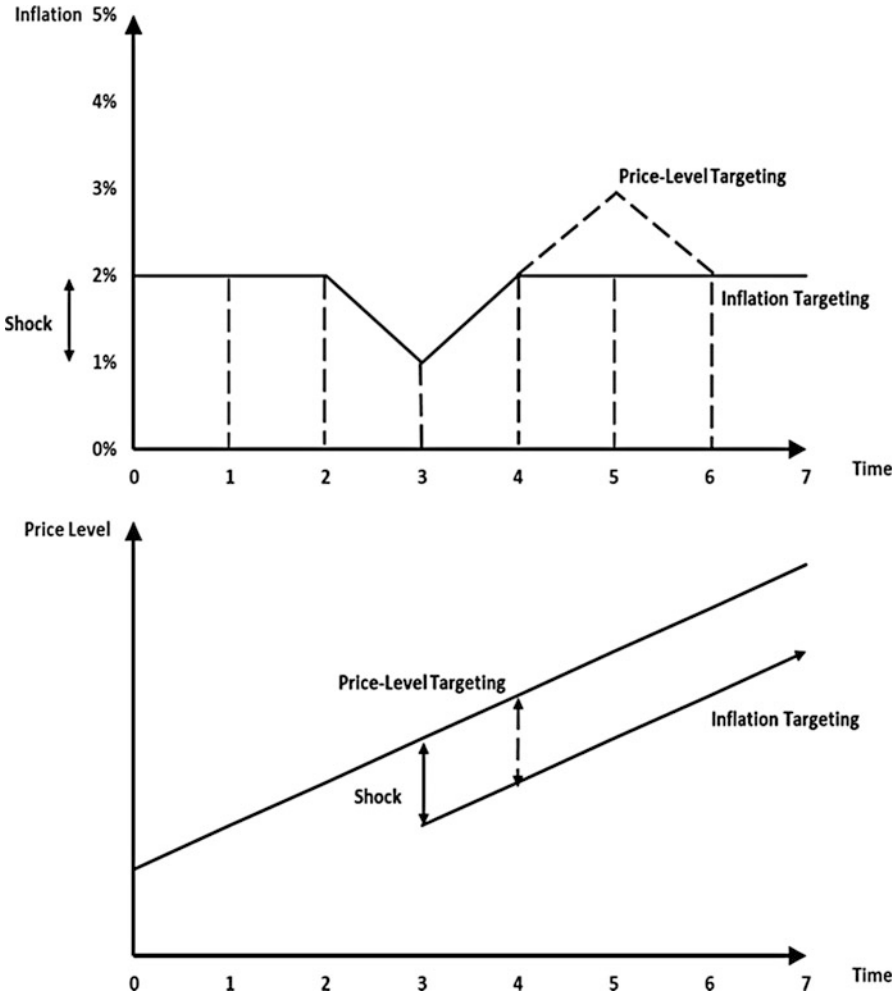


Fig. 2 Price-level versus inflation targeting in the face of a deflationary shock

contractionary response to an inflation shock, however – in cases where wages are rigid downward. This leads Andersson and Berg (1995), for example, to conclude that an inflation target offers the less risky option in an initial fight against inflation even though a price-level target would be still be preferred in the long run after more complete credibility has been achieved. Whereas the importance both of commitment and forward-looking, rational expectations to the success of price-level targeting are further emphasized by Ambler (2009), Svensson (2001) advocates the long-run objective of price-level targeting as the only policy that can guarantee the constancy of the purchasing power of money and conceivably reproduce the stationary long-run price level seen under the classical gold standard.

Although the case for price-level targeting over inflation targeting is seen as remaining model dependent, Hatcher and Minford (2016) emphasize that price-level targeting becomes more appealing in the face of deflationary shocks when central banks are confronted by the zero interest rate lower bound. In this case, expectations of higher than usual inflation aimed at recapturing the price-level target should "induce negative real interest rates which stimulate economic activity out of the recession, putting an end to deflation and ending lower bound episodes rapidly" (Hatcher and Minford 2016, p. 348). Although no central bank has yet made the switch to a price-level target, the Bank of Canada seriously considered the case for price-level targets in preparation for its 2011 policy agreement with the government (Kahn 2009). Interestingly, Ruge-Murcia (2014) finds that, unlike other inflation targeting central banks, the Bank of Canada's policies over the 1996–2013 periods were already observationally equivalent to price-level targeting. Specifically, in the Canadian case, shocks to the price level over the sample period were reversed, and there was none of the "drift" evident in the other cases (comprising the central banks of Australia, New Zealand, Sweden, and the United Kingdom). There have also been calls for price-level targeting in the United States both before and after the 2012 announcement of an inflation target. San Francisco Federal Reserve Bank President John Williams presented the case for this policy change in May 2017, pointing to the advantage of allowing inflation to remain above the Federal Reserve's existing 2% inflation target to make up for periods when it was too low (Williams 2017). Similar sentiments were previously expressed by Federal Reserve Bank of Chicago President Charles Evans in October 2010 (Evans 2010).

The potential benefit of price-level targeting in anchoring price expectations is supported by Fregert and Jonung's (2004) account of how wages were able to adjust downward under the price-level target introduced after Sweden left the gold standard in September 1931 (see also Berg and Jonung 1999). The introduction of price-level targeting admittedly did not preclude continued uncertainty among policymakers and the public about the future course of the Swedish price level. Nevertheless, in Fisher's (1934, p. 331) view, the ensuing stability of the internal purchasing power of the Swedish krona in conjunction with this new stabilization policy "stand out as among the remarkable facts of the depression."

Fisher himself called for an adjustable commodity-based dollar that would automatically offset any upward or downward pressure on its purchasing power. Whenever prices fell below target by 1%, for example, the dollar value of the resource unit would be raised by 1% – at the same time automatically lowering the number of resource units in the dollar.

Fisher's proposal involved adjusting the resource content of the monetary unit in proportion to any deviations from the (fixed) target price level. Following Hall (2005, p. 95), this relationship can be written as:

$$x_t = 1/(p_o r_t), \tag{1}$$

x_t is the value of the currency expressed in resource units, r_t is the value of one of the resource relative to the cost-of-living bundle, and p_o is the target price level. Hall

(2005) points out that Chile's bank-issued bearer certificate operates according to Fisher's principle of a self-stabilizing monetary unit. This Chilean monetary unit, the *Unidad de Fomento* (UF), had its peso content adjusted each day in line with the estimated cost of living so as to maintain its purchasing power.

Burdekin et al. (2012) further suggest that a stripped-down form of Fisher's commodity price standard could have helped Austria maintain price stability during the gold standard era. When increased scarcity of gold produced lower goods prices under the classical gold standard, Austrian data imply that investors saw silver prices moving on a one-to-one basis with the aggregate price level as Austria bought and sold gold on the London market to maintain the targeted exchange rate. This implies that a silver peg could potentially have allowed for inflation targeting with this one commodity used as a proxy for overall prices. More recently, while not involving any explicit price-targeting strategy, the 1934 Silver Purchase Act was a significant part of the more expansionary policies adopted after the US exit from the gold standard in 1933. The empirical importance of the silver effect detailed in Burdekin and Weidenmier (2009) serves as a reminder that government bond purchases are not the only means of achieving monetary expansion. And it is not just in the 1930s that commodity purchase programs may warrant further consideration if very low interest rates, and liquidity-constrained banks, call into question the effectiveness of more conventional policies.

Consumer Prices Versus Producer Prices

During the Swedish application of price-level targeting, the policy focus was consumer prices, and the Riksbank began calculating a weekly consumer price index after the new policy was launched in September 1931. Although the Swedish Minister of Finance announced in January 1932 that stabilization of consumer prices did not necessarily preclude rising wholesale prices (Fregert and Jonung 2004, p. 117), consumer prices and other product price series followed quite similar patterns – with both consumer and wholesale prices turning upward from 1933 onward. Similarly, when UK and US consumer prices began to recover after 1933, producer prices followed suit. No such co-movement is seen in the recent Chinese experience, however, where a gap of over 7% emerged between consumer and producer prices in the years following the onset of the global financial crisis. Although consumer price deflation remained minimal, producer prices exhibited much more drastic declines – culminating in a sizeable gap between the two (Fig. 3). Analogous, albeit less pronounced gaps were seen in Japan after the global financial crisis and also during the post-2013 Spanish deflation (Figs. 4 and 5).

In China, a prior bout of short-lived consumer price deflation arose in the midst of extremely tight monetary policy aimed at combating an inflationary spike in 1993–1994. While renewed consumer price deflation in the aftermath of, first, the Asian financial crisis, and, second, the global financial crisis had also been only intermittent in nature, the post-2008 drop in producer prices was much more pronounced. Kohler (2016) actually argues that the problem was of China's own

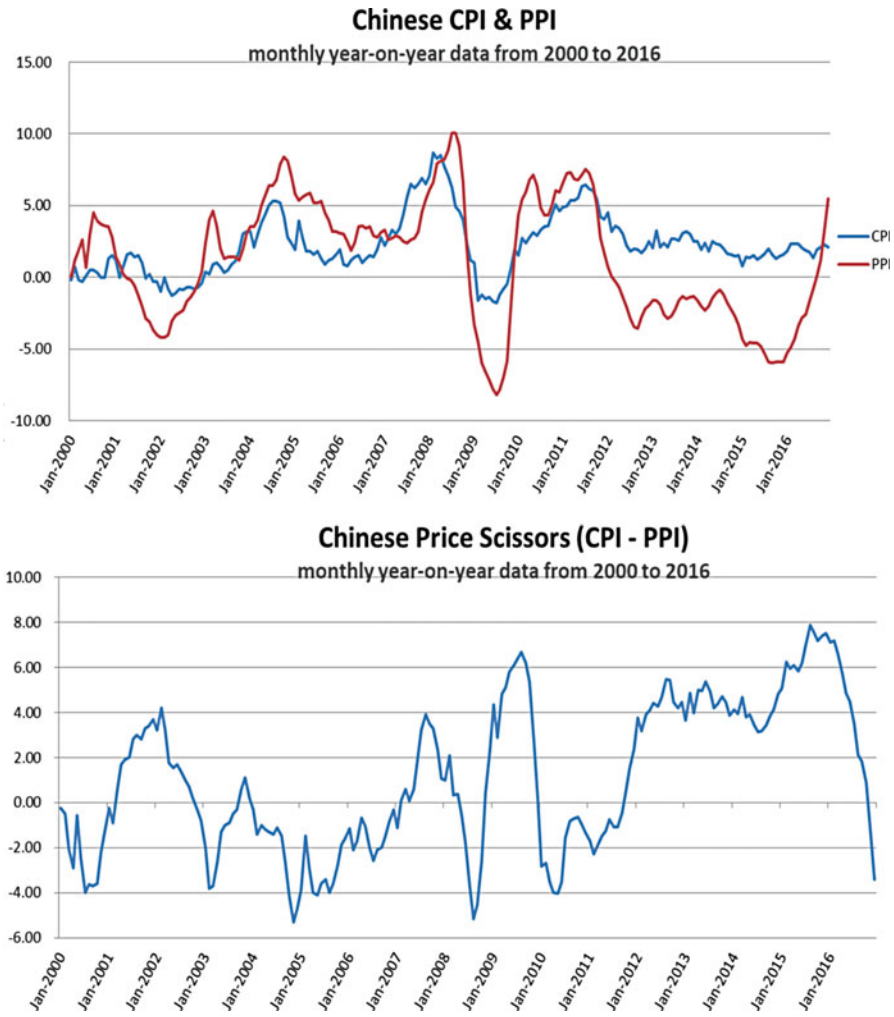


Fig. 3 Consumer versus producer prices in China, January 2000–December 2016

making insofar as China's massive stimulus program launched in November 2008 led to oversupply of steel and other raw materials. Irwin (2017) also points to low oil prices as a source of deflationary pressures, along with "more worldwide capacity for major commodities like steel and aluminium than there is demand, in part because of China's sheltering of state-run enterprises from the vicissitudes of the marketplace." There may actually be something of a vicious circle here in that excess supply in China puts downward pressure on global commodity prices, in turn causing China to be faced by imported deflation that only exacerbates the original problem (Wan 2015). Liu et al. (2017) also place blame on the November 2008 stimulus package, arguing that, in artificially bolstering the ratio of investment to consumption, this

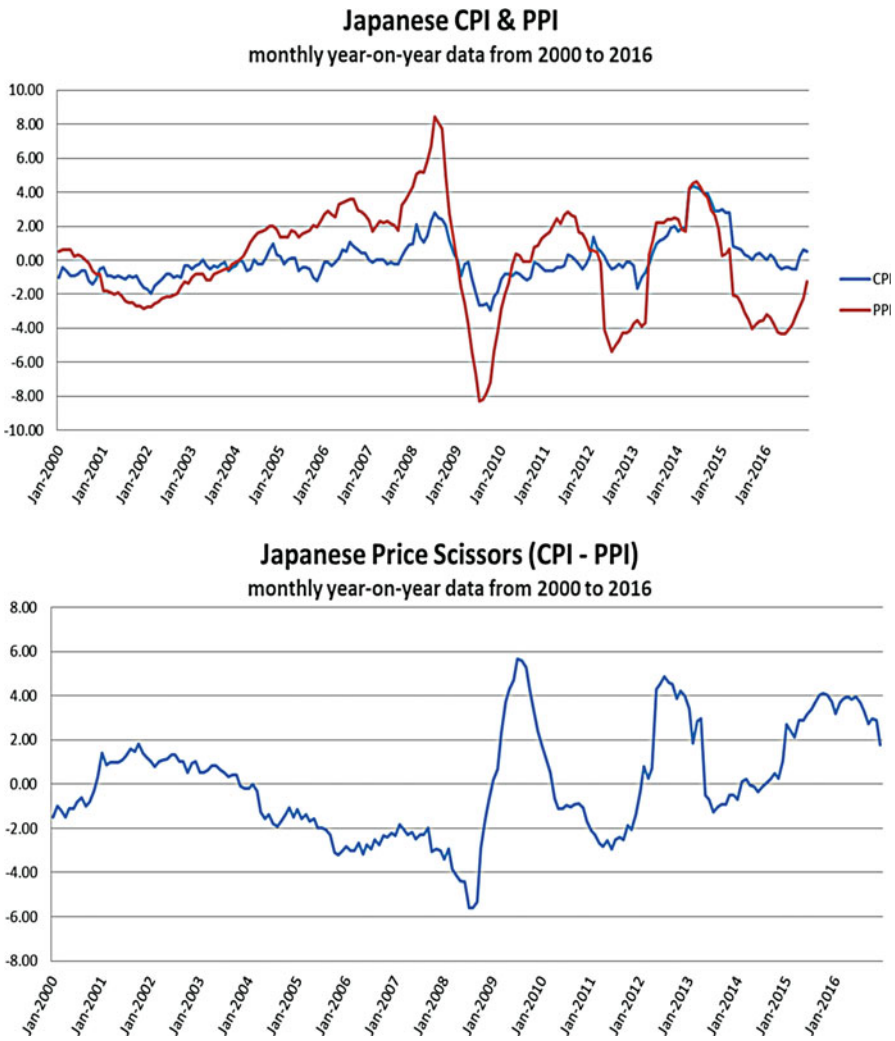


Fig. 4 Consumer versus producer prices in Japan, January 2000–December 2016

worsened structural imbalances in China's economy that induced the rising price scissors. Liu et al. therefore see the problem as deeper than a simple overcapacity issue and argue that the only way to eliminate the price scissors is through structural reform on the demand side that would boost PPI by encouraging greater private investment.

Although higher producer prices were seen in early 2017, Chinese consumer prices failed to respond in the way they had in the past. Ge (2017) attributed this primarily to overcapacity in China's downstream industries that worked against passing on higher producer prices into manufactured goods prices. Indeed, Ge

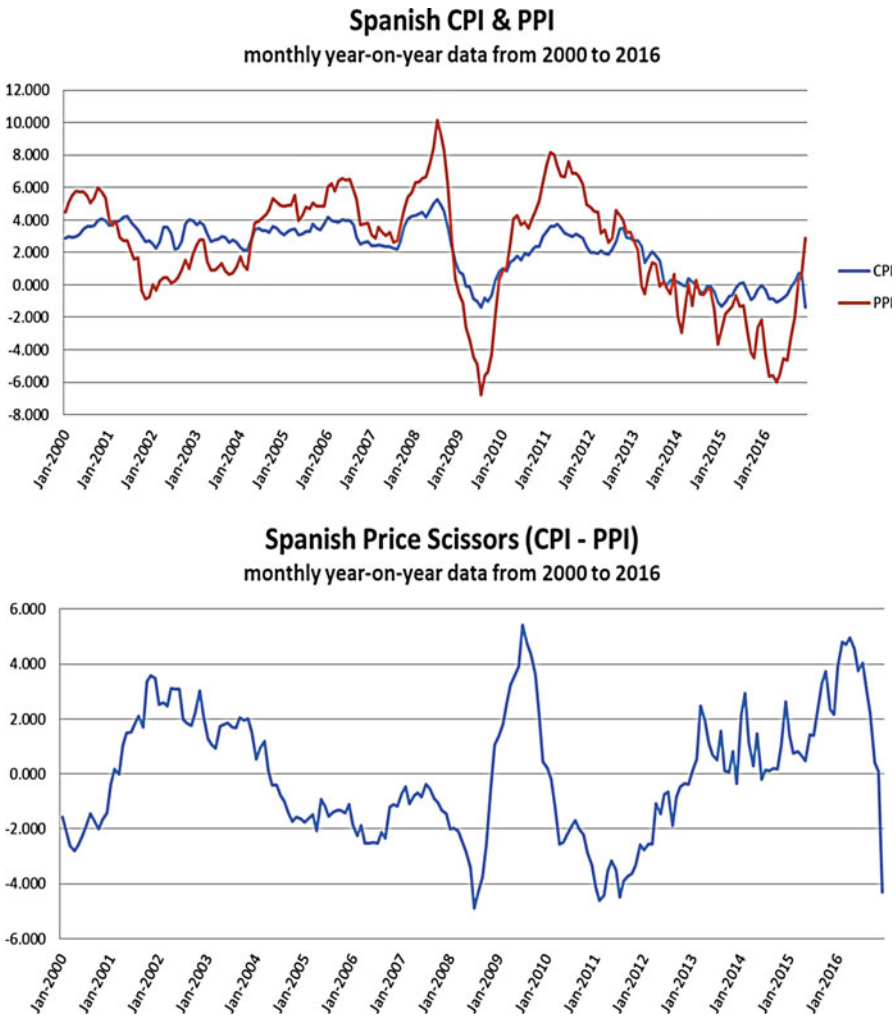


Fig. 5 Consumer versus producer prices in Spain, January 2000–December 2016

(2017) argues that this was likely to give rise to a negative feedback loop as higher producer prices were associated with lower profit margins that led downstream industries to reduce their investment spending. Meanwhile, the relative strength of consumer prices relative to producer prices cannot reasonably be attributed to wage gains, given that China's former rising wage trends reversed after the global financial crisis (Fig. 6). Nor can deflationary pressures seemingly be attributed to productivity gains. As shown in Fig. 7, whereas Chinese total factor productivity accelerated in the aftermath of the massive November 2008 stimulus package, a slowdown sets in after 2010 that coincided quite closely with the opening of the price scissors.

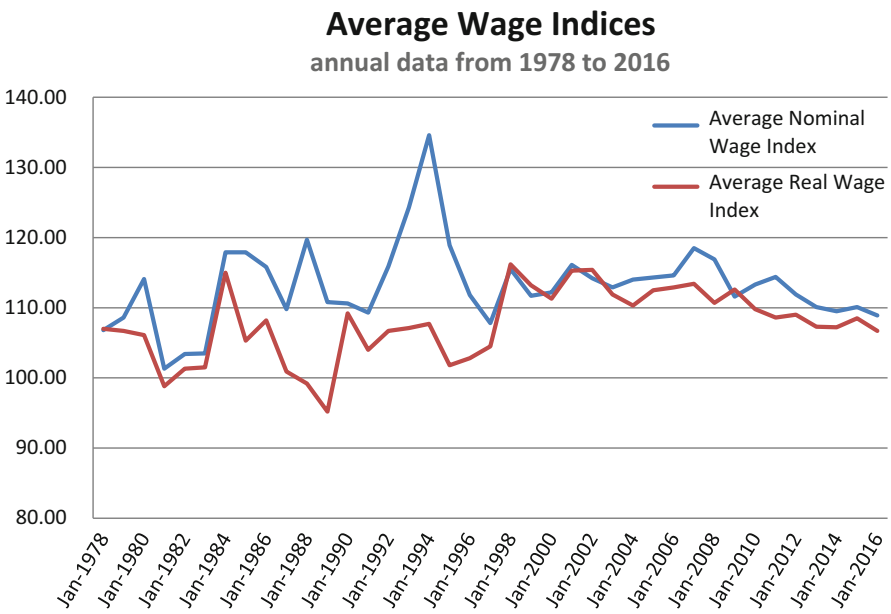


Fig. 6 Chinese wage trends over the post-1978 reform era. (Source: WIND terminal)

Fig. 7 Chinese total factor productivity growth, 2001–2013. (Source: Wu 2016)

Table 2 Components of Chinese total factor productivity, 2001–2013

Year	Technical change	Technical efficiency change	Economies of scale	Efficiency in resource allocation	Total factor productivity growth
2001	0.75	−0.39	0.77	0.83	1.96
2002	0.64	−0.25	0.79	0.79	1.97
2003	0.58	−0.17	0.84	0.83	2.08
2004	0.50	−0.05	0.89	0.75	2.09
2005	0.48	−0.01	0.90	0.83	2.2
2006	0.47	0.02	0.90	0.82	2.21
2007	0.44	0.03	0.91	0.75	2.13
2008	0.37	0.05	0.88	0.70	2.00
2009	0.36	−0.05	0.97	0.75	2.03
2010	0.35	0.20	0.95	0.79	2.29
2011	0.27	0.19	0.94	0.64	2.04
2012	0.19	0.04	0.89	0.65	1.77
2013	0.05	0.02	0.86	0.56	1.49
Average	0.42	−0.02	0.88	0.75	2.03

Source: Wu (2016)

Wu (2016) links this overall slowdown to across-the-board declines in the contributions arising from technical change, technical efficiency, economies of scale, and efficiency in resource allocation (Table 2).

The recent discrepancy between producer and consumer prices echoes a phenomenon that existed in China before price reform began in the 1980s. The controlled prices under the command economy had "allowed a form of price scissors to open by allowing raw materials prices, including agricultural prices, to increase more slowly than consumer prices" (Wedeman 2003, p. 49). The resulting resource transfer from agricultural areas to the urban and coastal areas itself had parallels with Preobrazhensky's theory of primitive socialist accumulation in the former Soviet Union, whereby resource transfers from the agrarian sector were to fund growth in the industrial sector (see also Burdekin 1989). It is ironic that a similar price scissors, previously thought to be merely a relic of the old command economy, reemerged in the predominantly market-based system in place today. Moreover, some degree of analogy with the old price scissors remains insofar as expansion of the services sector at the expense of the agricultural sector has helped boost consumer prices relative to producer prices in the new millennium (Lu and Wang 2015). The differential between consumer and producer prices has itself been found to be linked with stock market performance in China, and Guo and Liu (2015) find evidence of a long-run relationship between the CPI-PPI differential and the Shanghai A-share index over the 2001–2011 period. They attribute this linkage to the implications of growing CPI-PPI differentials implying larger profit margins and stronger company performance – albeit without taking into account the role of policy-driven factors like bank deposits and bank loans (Liang and Willett 2015).

Guo and Liu's (2015) emphasis on the price scissors' impact on profit margins points to the phenomenon potentially having at least some positive effects. And the fact that the Shanghai market continued to rise even as PPI deflation accelerated after 2010 is not inconsistent with Guo and Liu's premise. On the other hand, there is the risk that declining producer prices will eventually pull consumer prices down as well, leading to the traditional concerns of consumers putting off nonessential purchases and wages needing to have downward flexibility in order to avoid the risk of workers being priced out of a job. Indeed, Zou (2016) argues that the scale of the recent producer price declines could lead to deflation becoming entrenched in China and potentially unleash a debt-deflation cycle. This is itself far from just a Chinese concern as similar price scissors developed in other nations facing deflationary pressures after the global financial crisis. Indeed, the price scissors reached 4–5% in Japan over the 2009–2016 period and also approached 5% in Spain, steadily increasing through 2016 after deflation in consumer prices set in during 2013.

Traditionally, it would indeed be expected that, even if consumer prices initially remained reasonably stable, sharp producer price declines would sooner or later translate into significant consumer price deflation. This is consistent with the unidirectional causality from producer prices to consumer prices identified in such studies as Caporale et al. (2002), who find this pattern to consistently apply across all the G7 countries over the 1976–1999 period. Some controversy surrounds the question of whether this standard direction of causality has held in the Chinese case, however. Fan et al. (2009), for example, instead find evidence of causality running from consumer to producer prices over the 2001–2008 period in China. Although they attribute this to a dominant role of demand-side factors over supply-side factors during this period, their sample predates the more recent oversupply concerns that emerged after 2008.

Japan and Spain represent the two largest economies, other than China, to have so far experienced outright deflation in the aftermath of the global financial crisis. As with the Chinese situation, structural concerns apply in each case. However, worries about underinvestment in Japan contrast with the overinvestment that was a hallmark of China following the launch of the massive November 2008 stimulus package. Meanwhile, Spain's delays in bank recapitalization have been compared to Japan's own inadequate policy responses as it entered its first "Lost Decade" in 1990 (Hoshi and Kashyap 2015). Although it may seem that Spain's mildly positive growth since deflation began in 2013 implies a modern-day instance of "good" deflation, Munevar (2016) points to not only large-scale job losses but also to so many people losing their homes that Spain alone accounted for nearly 30% of the entire Europe-wide stock of empty houses by 2016. Moreover, Spain's limited GDP growth has relied upon highly expansionary fiscal policy that is likely non-sustainable with continued deflation pushing real debt burdens up.

Notwithstanding the different structural challenges facing the Chinese, Japanese, and Spanish economies after the global financial crisis, one deflationary catalyst they all faced was the worldwide decline in commodity prices since 2008, highlighted by a more than 70% drop in crude oil prices. Insofar as this affected producer prices more than consumer prices in Japan and Spain, like in China, we have a potential

common element giving rise to the price scissors phenomenon in three otherwise quite different economies. Short-run effects of oil price changes are generally confined to such oil-related consumption areas as petrol and heating oil. The initial impact on non-oil consumption will be constrained by the time required for increases in firms' input costs to be passed through – and, in the case of Spain, Álvarez et al. (2017) stress the slow speed and limited strength of this mechanism. Similarly limited transmission between oil price movements and consumer prices in the Eurozone as a whole is suggested by Castro et al. (2016), who conclude that deflation in consumer prices is unlikely to ensue from oil price declines alone. Limited effects on consumer prices, combined with more substantial and direct effects on producer prices, would, however, be very much in line with the widening gap between consumer prices and producer prices that emerged when oil prices plummeted after 2008.

Burdekin and Hu's (2018) vector autoregression (VAR) analysis of post-2000 data from China, Japan, and Spain offers some support for a supply-side, commodity-based impetus to the recent deflationary pressures insofar as oil price movements (proxied by West Texas Intermediate crude) exert effects on PPI that are significant at the 99% confidence level for all three countries. Oil prices also significantly Granger-cause the CPI in Japan and Spain at the 99% confidence level. Meanwhile, weaker oil price effects on Chinese CPI help explain the more dramatic price scissors in China than in the other two countries – with lower commodity prices seemingly driving down both price series in the other countries but definitively impacting only PPI in China. Oil price shocks were previously found to play a similarly important role in determining Chinese producer price movements over the 1996–2009 period by Cai and Wang (2012). Cai and Wang's structural VAR analysis identified significant effects of global commodity prices on China's PPI, whereas there was no evidence of any such transmission to China's CPI. Tang et al. (2014) find further evidence of global commodity price movements being transmitted to Chinese PPI but not CPI using data from the 2000 to 2011 period. Although Tang et al. (2014) point to profits in downstream industries accordingly being squeezed when commodity prices are rising, the other obvious implication is that falling commodity prices should be profitable for such industries, at least in the short-run – in line also with Guo and Liu's (2015) finding of positive stock market effects arising from any widening of the price scissors.

The danger of a widened price scissors, however, is that falling PPI eventually gives rise to larger CPI declines, in turn raising fears of negative effects on consumption, potential debt-deflation, and also pressure for wage cuts. Analysis of the price scissors itself in Burdekin and Hu (2018) shows the lagged change in oil prices to be significant at the 99% confidence level for China and Japan and at the 95% confidence level for Spain, with the expected negative sign. Thus, falling commodity prices appear to lower PPI more than CPI, hence exacerbating the extent of the price scissors. Insofar as supply-side deflationary pressures have indeed outweighed the effects of expansionary demand-side policies in China, Japan, and Spain, it has to be of considerable concern that the extensive expansionary policy initiatives undertaken since 2008 apparently proved insufficient to reign in deflationary pressures.

An Alternative Commodity-Based Standard?

To the extent that post-2008 deflationary pressures can indeed be considered a supply-side phenomenon, the price scissors developing in countries like China, Japan, and Spain would simply be signaling how such effects initially manifest themselves more in PPI than in CPI. This raises two rather troubling implications, however. First of all, even massive monetary stimulus seems to have been insufficient to ward off ongoing deflationary pressures since the global financial crisis. Second, it seems highly likely, if not inevitable, that sharply falling PPI will sooner or later translate into significant consumer price deflation if left unchecked. In light of this, might central banks consider striking at the root of the problem and target commodity prices directly? Although a formal commodity-based standard may not be a likely contender today, serious consideration was given after World War II to buffer stocks of raw materials being incorporated under the new International Monetary System (see Burdekin 2007). In a more limited application of the idea of commodity-backing, Frankel (2003) and Frankel and Saiki (2002) suggest that policymakers in commodity-dependent nations should tie monetary policy to the nation's primary commodity price. That is, stimulate monetary expansion whenever the nation is faced by declining demand for the commodity (or commodities) in question. Meanwhile, less specialized producers could enjoy analogous benefits by fixing the price of a basket of export commodities in terms of local currency (Frankel 2005).

Although the case for pegging the export price will generally be stronger when exports are large relative to total GDP, bigger economies could still benefit from such a commodity-based strategy, at least in comparison to the more popular alternative of pegging to the US dollar. For example, had Argentina adopted a wheat-peg rather than a dollar-peg during 1991–2001, the Argentinean peso would have been pushed down by falling commodity prices in contrast to the sharp appreciation actually seen over the latter part of this period (Frankel 2003). The dollar peg was deflationary for Argentina just as the gold peg was deflationary for the United States and many other countries in the late nineteenth century and during the Great Depression – and the authorities in each case ended up maintaining a fixed single-asset price, while almost all other prices declined. For Latin American countries as a whole, Frankel (2011) stresses the advantage of targeting PPI rather than CPI as a basis for monetary stabilization policy. Frankel demonstrates how, under this policy, falling commodity export prices would, for example, automatically trigger an expansionary response and currency depreciation – whereas CPI targeting would not. Meanwhile, in the case of rising import prices, CPI targeting would induce the central bank to tighten monetary policy and appreciate the currency to offset the rise in local currency prices, thereby actually contracting in the face of a negative shock.

Tying monetary expansion to commodity prices rather than consumer prices alone certainly seems like a valid proposal today for any countries facing expanding price scissors of the type seen in China, Japan, and Spain – regardless of whether the economies concerned are necessarily as commodity-dependent as those of Latin America. Indeed, under inflation targeting, the existing policy focus is even narrower in that the target is typically set in terms of a "core inflation" measure. Core inflation is

intended to represent the long-term trend in prices, which generally involves excluding food and energy prices on account of their greater volatility. The problem is that, under this strategy, even sustained movements in oil prices such as those seen in the aftermath of the global financial crisis will fail to elicit any policy response. Quite aside from the question of price level versus inflation targeting, a broader measure of the policy target seems warranted in the post-global financial crisis world.

Deflation: Always and Everywhere a Monetary Phenomenon?

In essence, declining prices must imply insufficient monetary expansion, i.e., too little money chasing too many goods. This premise seems to receive ample support from the deflationary experiences of the nineteenth century and 1930s. Each episode was marked by clear declines in the money supply and a largely passive policy, limited at the time by the constraints of the international gold standard. On the other hand, post-2008 deflationary pressures have emerged despite record rates of base money expansion around the world and central banks pushing interest rates down close to zero or even beyond. Although it may not seem reasonable to label such herculean efforts as being insufficient, the reality is that the expansionary push has been offset by other factors reversing the positive effects on the price level. The first has been a decline in the money multiplier and cutback in bank lending, dwarfing in size the movements in the same direction that were seen during the 1930s – as well as in Japan in the 1990s. The second key consideration concerns how the much more limited expansion in overall money supply has seemingly been offset by supply-side effects like the sharp declines in the price of oil and other commodities after the global financial crisis. More worrying still, these effects were initially felt primarily in producer prices rather than consumer prices in countries like China, Japan, and Spain.

Post-2008 deflationary concerns have already drawn renewed attention to the case for major structural reforms in China, Europe, and Japan. Some old remedies may warrant more attention as well, however. Price-level targeting would require central banks to reverse any price declines that have already occurred, thereby potentially providing a more solid anchor for expectation formation and wage setting as seen in the Swedish example of the 1930s. At the same time, the price scissors phenomenon, under which producer price declines can, in the short run, greatly outstrip the fall in consumer prices, warrants policymakers' attention too. Giving at least some weight to commodity prices in the design and implementation of monetary stabilization may be appropriate, again building upon proposals and theories that arose during deflationary episodes of the past. In this respect, it is worth bearing in mind the old adage: "Those who cannot remember the past are condemned to repeat it."

Cross-References

► International Monetary Regimes: The Gold Standard
► International Monetary Regimes: The Interwar Gold Exchange Standard

References

Álvarez LJ, Sánchez I, Urtasun A (2017) The effect of oil price fluctuations on Spanish inflation. Banco de España, Madrid, May 11. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/ArticulosAnaliticos/2017/T2/files/beaa1702-art14e.pdf

Ambler S (2009) Price-level targeting and stabilisation policy: a survey. J Econ Surv 23:974–997

Andersson K, Berg C (1995) The inflation target in Sweden. In: Haldane AG (ed) Targeting inflation: a conference of central banks on the use of inflation targets organized by the Bank of England. Bank of England, London, pp 207–225

Berg C, Jonung L (1999) Pioneering price level targeting: the Swedish experience 1931–1937. J Monet Econ 43:525–551

Bernanke BS (1983) Non-monetary effects of the financial crisis and the propagation of the Great Depression. Am Econ Rev 73:257–276

Bordo MD, Redish A (2004) Is deflation depressing? Evidence from the classical gold standard. In: Burdekin RCK, Siklos PL (eds) Deflation: current and historical perspectives. Cambridge University Press, New York, pp 191–217

Borio C, Erdem M, Filardo A, Hoffman B (2015) The costs of deflation: a historical perspective. BIS Q Rev 31–54

Boskin MJ (2005) Causes and consequences of bias in the consumer price index as a measure of the cost of living. Atl Econ J 33:1–13

Burdekin RCK (1989) Preobrazhensky's theory of primitive socialist accumulation. J Contemp Asia 19(3):297–307

Burdekin RCK (2007) Nontraditional monetary policy options and commodity-based stabilization policy. Int Econ Finance J 2:1–18

Burdekin RCK, Hu X (2018) Deflationary pressures today: the Chinese, Japanese and Spanish cases compared. Chin Econ, forthcoming

Burdekin RCK, Siklos PL (2004) Fears of deflation and the role of monetary policy. In: Burdekin RCK, Siklos PL (eds) Deflation: current and historical perspectives. Cambridge University Press, New York, pp 1–27

Burdekin RCK, Weidenmier MD (2001) Inflation is always and everywhere a monetary phenomenon: Richmond vs. Houston in 1864. Am Econ Rev 91:1621–1630

Burdekin RCK, Weidenmier MD (2009) The development of 'non-traditional' open market operations: lessons from FDR's silver purchase program. In: Atack J, Neal L (eds) The origins and development of financial markets and institutions: from the seventeenth century to the present. Cambridge University Press, New York, pp 319–344

Burdekin RCK, Mitchener KJ, Weidenmier MD (2012) Irving Fisher and price-level targeting in Austria: was silver the answer? J Money Credit Bank 44:733–750

Cai Hong-bo, Wang Jun-hai (2012) International commodity price and China's economic fluctuation: empirical analysis based on different crude oil price Shocks [in Chinese]. Int Bus (Univ Int Bus Econ) [国际商务(对外经济贸易大学学报)] (1):54–64

Caporale GM, Katsimi M, Pittis N (2002) Causality links between consumer and producer prices: some empirical evidence. South Econ J 68:703–711

Castro C, Jerez M, Barge-Gil A (2016) The deflationary effect of oil prices in the euro area. Energy Econ 56:389–397

Day J (1978) The great bullion famine of the fifteenth century. Past Present 79:3–54

Dickey GE (1977) Money, prices and growth: the American experience 1869–1896. Arno Press, New York

Eichengreen B (1992) Golden fetters: the gold standard and the Great Depression, 1919–1939. Oxford University Press, New York

Eichengreen B, Temin P (2013) Fetters of gold and paper. In: Crafts N, Fearon P (eds) The Great Depression of the 1930s: lessons for today. Oxford University Press, New York, pp 429–448

Evans CL (2010) Monetary policy in a low-inflation environment: developing a state-contingent price-level target. Remarks delivered before the Federal Reserve Bank of Boston's 55th

economic conference on October 16. http://www.chicagofed.org/webpages/publications/speeches/2010/10_16_boston_speech.cfm
Fackler JS, Parker RE (2005) Was debt deflation operative during the Great Depression? Econ Inq 43:67–78
Fan G, He L, Hu J (2009) CPI vs. PPI: which drives which? Front Econ China 4(3):317–334
Fisher I (1913) A compensated dollar. Q J Econ 27:213–235
Fisher I (1920) Stabilizing the dollar. Macmillan, New York
Fisher I (1933) The debt-deflation theory of Great Depressions. Econometrica 1:337–357
Fisher I (1934) Stable money: a history of the movement. Adelphi, New York
Frankel J (2003) A proposed monetary regime for small commodity exporters: peg the export price. Int Finance 6:61–88
Frankel JA (2005) Peg the export price index: a proposed monetary regime for small countries. J Policy Model 27:495–508
Frankel JA (2011) Comparison of product price targeting and other monetary anchor options for commodity exporters in Latin America. Economia 12:1–57
Frankel J, Saiki A (2002) A proposal to anchor monetary policy by the price of the export commodity. J Econ Integr 17:417–448
Fregert K, Jonung L (2004) Deflation dynamics in Sweden: perceptions, expectations, and adjustment during the deflations of 1921–1923 and 1931–1933. In: Burdekin RCK, Siklos PL (eds) Deflation: current and historical perspectives. Cambridge University Press, New York, pp 91–128
Friedman M (1992) Money mischief: episodes in monetary history. Harcourt Brace Jovanovich, New York
Friedman M, Schwartz AJ (1963) A monetary history of the United States, 1857–1960. Princeton University Press, Princeton
Ge Feng [葛丰] (2017, March 20) PPI, CPI price scissor enlarged, adding difficulty to macroeconomic control [in Chinese]. China Economy Weekly [中国经济周刊], No. 11. http://www.ceweekly.cn/2017/0320/184609.shtml
Guo Sanshua [郭三化], Liu Jun [刘俊] (2015) Empirical study of the relationship between the CPI-PPI differential and the stock index [in Chinese]. J Financ Dev Res [金融发展研究] (7): 72–76
Hall RE (2005) Controlling the price level. Am J Econ Sociol 64:93–112
Hamilton JD (1992) Was the deflation during the Great Depression anticipated? Evidence from the commodity futures market. Am Econ Rev 82:157–178
Hanke SH (2016) State money and bank money: lifting the fog around QE. Cent Bank, June 6. https://www.cato.org/publications/commentary/state-money-bank-money-lifting-fog-aroundqe
Hatcher M, Minford P (2014) Inflation targeting vs. price-level targeting: a new survey of theory and empirics. VOX, CEPR's Policy Portal, May. http://voxeu.org/article/inflation-targeting-vs-price-level-targeting
Hatcher M, Minford P (2016) Stabilisation policy, rational expectations and price-level versus inflation targeting: a survey. J Econ Surv 30:327–355
Hetzel RL (2014) Should Greece remain in the Eurozone? Econ Q 100(Third Quarter):241–278. Federal Reserve Bank of Richmond
Hirakata N, Sudo N, Takei I, Ueda K (2016) Japan's financial crises and lost decades. Jpn World Econ 40:31–46
Horesh N (2014) Chinese money in global context: historic junctures between 600 BCE and 2012. Stanford University Press, Stanford
Hoshi T, Kashyap AK (2015) Will the U.S. and Europe avoid a lost decade? Lessons from Japan's postcrisis experience. IMF Econ Rev 63(1):110–163
Hutchison MM (2004) Deflation and stagnation in Japan: collapse of the monetary transmission mechanism and echo from the 1930s. In: Burdekin RCK, Siklos PL (eds) Deflation: current and historical perspectives. Cambridge University Press, New York, pp 241–268
Irwin N (2017, April 26) The low-inflation world may be sticking around longer than expected. New York Times. https://www.nytimes.com/2017/04/26/upshot/the-low-inflation-world-may-be-sticking-around-longer-than-expected.html

Jonung L (1979) Knut Wicksell's norm of price stabilization and Swedish monetary policy in the 1930s. J Monet Econ 5:459–496
Kahn GA (2009) Beyond inflation targeting: should central banks target the price level? Econ Rev (Third Quarter):35–64. Federal Reserve Bank of Kansas City
Kohler A (2016) The disaster of negative rates. The Australian, April 12. http://www.theaustralian.com.au/business/opinion/alan-kohler/the-disaster-of-negative-rates/news-story/9d7f93a8a0e1188dff29b31b4d70f372
Liang P, Willett TD (2015) Chinese stocks during 2000–2013: bubbles and busts or fundamentals? Chin Econ 48(3):199–214
Liu Fengliang, Zhang Xiaomeng, Yu Ze (2017) High investment, structural imbalance, and the CPI & PPI deviation [in Chinese]. J Financ Res [金融研究] (2):54–69
Lu Jie, Wang Gaowang (2015) A structural explanation on the 'divergence' between the CPI and PPI [in Chinese]. Jingji Yanjiu/Econ Res J 50(4):136–149
Luhnow D, Druckerman P (2001, December 4) Long hailed as hero, reformer in Argentina sees his dream sour. Wall Street Journal, pp A1, A6
Mundell R (1963) Inflation and real interest. J Polit Econ 71:280–283
Munevar D (2016) Spain: the politics of austerity and deflation. Institute for New Economic Thinking, July 4. https://www.ineteconomics.org/perspectives/blog/spain-the-politics-of-austerity-and-deflation
Ruge-Murcia F (2014) Do inflation targeting central banks implicitly target the price level? Int J Cent Bank 10:301–326
Saxonhouse G (2005) Good deflation/bad deflation and Japanese economic recovery. IEEP 2:201–218
Spence P (2015, September 27) How Sweden's negative interest rate experiment has turned economics on its head. The Telegraph. http://www.telegraph.co.uk/finance/economics/11895084/How-Swedens-negative-interest-rates-experiment-has-turned-economics-on-its-head.html
Stewart H (2013, November 13) Deflation fears stalk Eurozone as Spain reports fall in prices. The Guardian. https://www.theguardian.com/business/2013/nov/13/deflation-fears-eurozone-spain-fall-prices
Svensson LEO (2001) Price stability as a target for monetary policy: defining and maintaining price stability. In: Deutsche Bundesbank (ed) The monetary transmission process: recent developments and lessons for Europe. Palgrave, Chippenham, pp 60–102
Tang K, Wang C, Wang S (2014) China's imported inflation and global commodity prices. Emerg Mark Finance Trade 50:162–177
Temin P (2013) The Roman market economy. Princeton University Press, Princeton
Tobin J (1965) Money and economic growth. Econometrica 33:671–684
Wan Zhao [万钊] (2015, August 10) Chinese PPI-CPI scissor enlarged, road ahead for structural adjustment will be long [in Chinese]. Reuters. http://cn.reuters.com/article/cn-ppi-colum-idCNKCS0QG0J420150811
Wedeman AH (2003) From Mao to market: rent seeking, local protectionism, and marketization in China. Cambridge University Press, New York
Wicksell K (1958) In: Lindahl E (ed) Selected papers on economic theory. Harvard University Press, Cambridge, MA
Williams JC (2017) Preparing for the next storm: reassessing frameworks and strategies in a low R-star world. FRBSF Economic Letter 2017-13, May 8
Wu Qiong [邬琼] (2016) Computation and analysis of Chinese factor productivity [in Chinese]. State Information Center [国家信息中心], 2 Sept. http://www.sic.gov.cn/News/455/6841.htm
Zou Jingxian (2016) CPI, PPI diversion and deflation [in Chinese]. Int Econ Rev [国际经济评论] (4):129–139

Index

S. Battilossi et al. (eds.), *Handbook of the History of Money and Currency*,
https://doi.org/10.1007/978-981-13-0596-2

J