

CHAPTER 1

Introduction

“The distinctive function of the banker,”—says Ricardo,
“begins as soon as he uses the money of *others*”; as long as he
uses his own money he is only a capitalist.

—Walter Bagehot (1924: 21)

THE CHALLENGE OF INTERMEDIATION

A fundamental challenge faced by economies for centuries has been to efficiently channel—“intermediate” in economic terminology—society’s aggregate savings toward productive enterprise. Ancient and modern economies alike have devised facilities for intermediating between savers and investors. In ancient Babylon, palace and temple officials loaned their own funds and, in time, deposits that were entrusted to them. In medieval Italy and Spain, institutions were established for the purpose of funneling private savings toward public use, in particular to fund the government’s debt. The roughly contemporaneous development of financial instruments, such as the bill of exchange, helped to channel savings toward the finance of both domestic and international trade. In modern industrial economies, intermediation falls primarily to securities markets and financial institutions.

Perhaps the best way to appreciate the importance of financial intermediators is to consider what the world would look like without them. In their absence, firms seeking finance and savers looking for investment opportunities would have to find each other and negotiate detailed contracts. Will funds be loaned or will they purchase a share of the enterprise? If loaned, for how long, at what interest rate, and against what collateral? If the funds purchase an ownership share, to what fraction of profits and seats on the board of directors will investors be entitled and, if the enterprise fails, how much liability will they bear? A system without financial intermediators would be, to put it mildly, inefficient. The transaction costs involved in seeking out investors and reaching agreements would be prohibitive, and firm managers, in all likelihood, would be forced to rely on retained earnings (i.e., funds generated by the firm but not distributed to owners) and the fortunes of friends and family to finance expansion. Economic growth, as typified by the modern industrialized—or industrializing—economy would seem impossible.

This book focuses on one type of financial intermediary: banks. More specifically, it analyzes the forces that have been responsible for the evolution—sometimes slow, sometimes dramatically rapid—of banking, particularly commercial banking, in the industrialized world during the past two centuries. This should be of interest for at least four reasons.

First, banks are among the oldest extant forms of financial intermediary. They are also among the most familiar. And, despite the growth of alternative intermediators, particularly securities markets, banks remain an integral part of the financial system. In short, the answer to the question posed by Boyd and Gertler (1994), “Are Banks Dead?” is “no.”

Second, if banking is an essential determinant of economic growth, its evolution may affect long-term economic performance and therefore help to account for countries’ differing historical and prospective growth trajectories. A substantial academic literature considers the merits of many aspects of banking systems, including the extent of concentration and branching, the relative sizes of the banking industry, securities markets, and other financial intermediators, and the costs and benefits of different regulatory restrictions on banks. This literature tends to assume—incorrectly—that the key characteristics of the banking system were predetermined at some distant historical date, rather than as the product of a longer-term evolution, potentially biasing its results.

Third, banking stability may have important short-run economic and political consequences. Banking crises have been a common feature of the world economy for the past two centuries and can generate economic fluctuations or exacerbate those already under way. Since certain types of banking systems may be more resistant to crisis than others, understanding the forces that govern the evolution of banking may help to explain differences—both across countries and through time—in the severity of cyclical fluctuations and their political fallout.

Finally, banking systems across the industrialized world have undergone revolutionary changes since the early 1970s. Banking crises and subsequent government rescues, consolidation among banks and between banks and other types of financial institutions, and changes in national, international, and supranational regulatory efforts have dramatically altered the financial landscape in recent years and, given recent financial turmoil, will continue to do so for the foreseeable future. Identifying the forces that historically have been responsible for banking evolution may provide some insight into understanding those at work today.

This book focuses on four types of events that have been both characteristic features of the life cycle of commercial banking systems and have brought about important structural changes in banking: crises, rescues, merger movements, and regulation. The detailed studies of these events

through time and across countries presented in subsequent chapters reveals that they are the result of multiple, and frequently idiosyncratic, causes. Nonetheless, several commonalities emerge.

Banking crises have been a prominent feature of the financial landscape from the earliest days of commercial banking, of which the subprime mortgage crisis is merely the most recent example. Banking crises have three primary causes: (1) “boom-bust” (or macroeconomic) fluctuations; (2) shocks to confidence; and (3) structural weakness. Boom-bust fluctuations have long been a feature of the modern industrialized economy. If banking growth is especially vigorous during a period of robust macroeconomic expansion, the subsequent contraction may strain the banking system to the point of crisis. Shocks to confidence arising from, for example, concerns about the continued convertibility of the currency—raising the specter that deposits may only be redeemable in devalued currency—or the anticipation of an impending war or other event that reduces confidence in the stability of the banking system, may generate panic withdrawals and precipitate a crisis. Finally, the likelihood of crises can be affected by the structure of the banking system. A banking system composed of many small banks or governed by one set of regulations may be more prone to crisis, no matter what the source of disturbance, than one characterized by fewer, larger institutions or governed by a different regulatory framework.

Because crises are costly and may spread if not controlled, private and public actors have incentives to engage in rescue operations. Rescues may be confined to bailing out one troubled institution or may extend to the banking system more generally through lender of last resort operations; they may be of limited scope or may consist of relatively extreme measures, such as wholesale bank nationalization. The consequences of the type of rescue undertaken can have long-lasting effects on the banking system. If a rescue entails sustaining weaker banks, it may encourage other banks to undertake riskier activities secure in the knowledge that they too will be rescued in case of distress, a phenomenon known as “moral hazard.” Thus, rescues may contribute to yet another boom-bust cycle marked by overexpansion and ending in crisis.

Crises typically leave a number of devastated banks in their wake. One consequence of this may be an increase in bank merger activity, as healthy banks absorb troubled institutions. Yet merger movements are not solely the creatures of economic downturns: economic expansion can also stimulate mergers if increased financing requirements are better met by larger financial institutions. The wealth effects of rising equity prices during an economic boom may also increase the willingness of potential acquirers to initiate mergers while rising share prices may make potential targets more willing to be acquired.

Another, usually not quite so immediate, consequence of crises is a change in banking regulations. Government regulation shapes many aspects of banking, ranging from rules governing entry, to those setting capital requirements, reserve requirements, and a variety of balance sheet ratios. Regulations govern whether banks can maintain branches, engage in securities market transactions, and the extent to which shareholders are liable for the losses of the bank. Although promoting banking soundness and stability—and indeed the public interest more generally—is frequently cited as the prime motive for government regulation, the private interests of different sectors of finance, industry, and government also influence the process generating new regulations.

Efforts at stability-promoting reform typically take one of two paths. Stricter regulation, including measures that strengthen safety and soundness requirements, curtail bank powers, and increase the reach and authority of regulators, constrains banks and makes it more difficult for them to get into trouble. An alternative regulatory approach takes a seemingly opposite course: expanding the powers of existing institutions—possibly combined with competition-reducing measures—in an effort to strengthen banks and render them less vulnerable to shocks.

These paths have very different consequences for the banking system and will find support among different private interest groups. Reforms aimed at restricting competition can lead to an excessively constrained banking system, although such reforms may be supported by financial incumbents protecting their private interests (Rajan and Zingales 2003b). Such constraints may eventually render the banking system less able to contribute to economic growth and lead to calls for deregulation. Reforms that allow an expansion of banking powers will have different consequences and will have different advocates. If such regulations are anticompetitive, they may lead to increased—perhaps to an uncomfortably high level—banking concentration. By granting wide-ranging powers, such reforms allow banks to increase risk-taking and may thus lead to greater instability.

In addition to the interplay between public and private interests, regulation will evolve as a result of unintended consequences. For example, regulations aimed at making banking safer by establishing more stringent standards may be so strict that they drive potential bankers to seek ways of avoiding the newly enacted regulations altogether: rather than acquire a government-issued charter, potential bankers might decide to operate as private partnerships or as off-shore enterprises, if doing so allows them to escape the stringent regulation. Bank efforts to avoid regulation lead legislators and regulators to devise new regulations which, in turn, encourage banks to find even newer ways of avoiding them.

BANKING AND ECONOMIC GROWTH

An important motive for studying banking evolution is its potential influence on economic growth. The theoretical literature on the interaction between the financial system, particularly banking, and economic growth is long and has a good pedigree. In his classic work, *The Wealth of Nations*, Adam Smith (1776 [1994]: 349) highlights the importance of banks in mobilizing capital:

It is not by augmenting the capital of the country, but by rendering a greater part of that capital active and productive than would otherwise be so, that the most judicious operations of banking can increase the industry of a country. That part of capital which a dealer is obliged to keep by him unemployed, and in ready money, for answering occasional demands, is so much dead stock, which, so long as it remains in this situation, produces nothing either to him or his country. The judicious operations of banking enable him to convert this dead stock into active and productive stock; into materials to work upon, into tools to work with, and into provisions and subsistence to work for; into stock which produces something both to himself and to his country.

Henry Thornton (1802 [1939]: 176) quibbles with Smith's characterization, suggesting that banks are important not only because they *mobilize* credit, but because, in the course of supplying paper money, they also *create* credit: "Whether the introduction of the use of paper is spoken of as turning dead and unproductive stock into stock which is active and productive, or as *adding* to the stock of the country is much the same thing."

The proper role of banks in creating credit by increasing issues of money was frequently debated in nineteenth-century Britain. Two instances stand out: the Bullionist controversy following Britain's suspension of the gold standard in 1797, and the clash between the Currency School and the Banking School, which arose around the time of the reorganization and recharter of the Bank of England in 1844. The crux of these arguments was the proper regulation of note issue. According to the Bullionists and the Currency School, in order to prevent an inflationary over-issue of currency, the note issue should vary "exactly as it would have done were it wholly metallic."¹ The Anti-Bullionists and the Banking School argued that as long as issuers maintained sufficient metallic reserves to insure convertibility, the note issue ought to vary with the needs of trade. If notes were issued only against the security of actual commercial transactions—the "real bills doctrine," which became a pillar of nineteenth-century banking orthodoxy in Britain and elsewhere—an

¹Morgan (1943: 130). See also Andréadès (1966) and Fetter (1965) on these debates.

inflationary over-issue would not be possible. An implication of the Anti-Bullionist/Banking School view is that the inability to issue notes and create credit in response to an increased demand for real commercial transactions would retard economic growth.

Joseph Schumpeter (1934: 74) argues that bankers play a key role in economic growth, explicitly highlighting their role in credit creation:

The banker, therefore, is not so much primarily a middleman in the commodity of “purchasing power” as a *producer* of this commodity. However, since all reserve funds and savings to-day usually flow to him, and the total demand for free purchasing power, whether existing or to be created, concentrates on him, he has either replaced private capitalists or become their agent; he has himself become the capitalist par excellence. He stands between those who wish to form new combinations and the possessors of productive means. He is essentially a phenomenon of development, though only when no central authority directs the social process. He makes possible the carrying out of new combinations, authorizes people, in the name of society as it were, to form them. He is the ephor of the exchange economy.

John Hicks (1969: 78, 94–97) juxtaposes the above approaches by describing three stages through which banking develops. In the first stage, banks operate solely as trusted middlemen: borrowing, re-lending, and profiting on the spread between borrowing and lending rates. In the second stage, the banker begins to accept deposits that can be withdrawn on demand. At this point, the banker becomes a principal in the transaction, accepting the risk of a bank run if liquid assets are not sufficient to redeem demand liabilities. In the third stage, deposits subject to withdrawal become transferable, either by check or note. According to Hicks, “. . . it is at this point that the bank becomes able to create what is in effect money.”

A related strand of literature focuses not directly on the credit-creating powers of banks, but on banks more generally in the realm of firm finance. Gurley and Shaw (1955) characterize earlier forms of firm finance as “self-finance,” in which firms finance investment projects from undistributed profits, and “direct finance,” in which firms borrow directly from savers. The next stage of financial development, termed “indirect finance,” occurs when banks issue their own debt and use the proceeds to finance investment projects. Gurley and Shaw argue that failure to develop this type of indirect finance through financial intermediators can retard economic development.²

²Nonetheless Gurley’s (1967: 953) review of Cameron (1967) argues that: “. . . recent experience strongly suggests that banking systems as intermediaries are not highly essential to the growth process.”

Alexander Gerschenkron (1962) also takes up the idea of investment finance, arguing that the importance of financial intermediators rose with the level of economic “backwardness.” He asserts that in Britain, the most developed economy of the eighteenth and nineteenth centuries, industrialization was financed primarily with the retained earnings of industrialists. In moderately backwards economies, exemplified by Germany, banks played a crucial role in mobilizing savings during industrialization—a role that had not been necessary in economically developed Britain. In the most backwards of countries, exemplified by Russia, he argues that financing industrialization was too great a task for banks to accomplish, and so the leading role had to be undertaken by the state.

Yet not all economic theorists have viewed finance as quite the lynchpin that these others seem to think it is. Joan Robinson (1952: 86) argues that:

... [t]here is a general tendency for the supply of finance to move with the demand for it. It is true, of course, that at any moment there are many excellent ideas which cannot be implemented because those that conceive them cannot back them up with finance. But, by and large, it seems to be the case that where enterprise leads finance follows.

And, according to Robert Lucas (1988: 6), economists “badly over-stress” the role of financial factors in economic growth. Further, Ross Levine (1997: 688) argues that economists who study the developing world express their skepticism about the role of the financial system in promoting growth “by ignoring it.”

Empirically, there is no doubt that finance and economic development move together. However, the question remains: does causality run primarily from finance to development or vice versa? Certainly, during financial crises, disintegration of the financial sector leads to a short-term deterioration in economic performance. Does the reverse hold? And does it hold in the longer run?

Raymond Goldsmith (1969) provides the first systematic empirical attempt to discern the relationship between financial system development and economic growth. This work provides detailed data on the financial systems of thirty-five countries at twenty-year intervals from as early as 1860. Goldsmith calculates the ratio of financial claims to national wealth, what he terms “the financial inter-relations ratio” (FIR), in order to determine the importance of the financial superstructure in economic growth. Goldsmith (1969: 390–91) states the task in the following way:

One of the most important problems in the field of finance, if not the single most important one, almost everyone would agree, is the effect that financial structure and development have on economic growth.

Does it make a measurable difference in the speed or character of economic growth how large a country's financial superstructure is compared to its national wealth and product; how rapidly and regularly the financial superstructure expands in nominal and real terms; how much of a country's investment is financed externally through the issuance of financial instruments rather than internally out of investors' own saving or through involuntary transfer like taxation; what types of financial instruments exist; what are their relative importance and penetration throughout the economy; what are the character, the methods of operation, the degree of specialization and concentration, the geographic distribution of financial institutions; and whether financial institutions are owned or operated privately, cooperatively, or by the government?

He then elaborates on the difficulty in finding an answer:

These questions and many others directed toward the nature of the financial superstructure and of changes in it are easy to pose, but most questioners are unaware how difficult and precarious answers are, both methodologically and factually. . . .

To assess the role of financial development and structure in economic growth we may turn, depending on our philosophical predilections, to economic theory or to economic history. We shall unfortunately find that in the present state of the theory of economic growth and with the present lack of sufficiently intensive historical studies of financial development we cannot get definite answers from either discipline. The reader who is looking for simple, unambiguous, and enduring solutions may as well, therefore, stop here.

Goldsmith finds that the FIR increased—up to a certain point—with the level of development. He also argues that, given the wide range of experiences across countries and the relatively low frequency of the data, it is impossible to definitively establish causality.

Around the same time, Rondo Cameron and coauthors (1967, 1972) undertook case studies on the role played by banks in the economic development of about a dozen countries. The results of these studies are mixed: in some countries (Belgium, Japan, Scotland, and the United States) the authors find that banking systems contributed to growth; in others (France, Italy, Russia, and Spain) the banking systems are found either to have responded passively to economic development or, where government intervention was misguided, to have hindered development by misallocating funds.

Starting in the 1990s, economists began to bring more sophisticated statistical techniques to bear on the relationship between finance and eco-

conomic growth (Demirgüç-Kunt and Levine 2001). The data sets analyzed typically include both greater numbers of countries and higher frequency data than previous efforts. King and Levine (1993) examine the relationship between economic growth and various measures of financial development across eighty countries during 1960–89 and find financial development to be a good predictor of economic growth over the subsequent ten to thirty years. De Gregorio and Guidotti (1995), Levine, Loayza, and Beck (2000), and Levine and Zervos (1998) reach similar findings for financial intermediary development—both banks and stock markets—and economic growth during various subperiods from the 1960s through the 1990s. Rajan and Zingales (1998) find that industrial sectors that were more highly dependent on external finance benefited more from greater financial sector development and Beck, Demirgüç-Kunt, Laeven, and Levine (2008) find that small firms benefited disproportionately from financial sector development.

Rousseau and Wachtel (1998) take a slightly different approach, analyzing data from fewer countries (Canada, Norway, Sweden, the United Kingdom, and the United States) over a longer period (1870–1929). They conclude that in most cases causality runs from financial sector development to real economic activity, although their findings are stronger for some countries than for others. Rousseau and Sylla (2003) examine a sample seventeen countries during 1850–1997 and find a robust correlation between financial factors and economic growth “. . . consistent with a leading role for finance.” Other empirical studies of periods of a century or more for Sweden (Hansson and Jonung 1997) and Canada (Wilson 2001) suggest that, for at least several subperiods and possibly for different stages of economic development, the direction of causality between financial sector development and economic growth is more ambiguous.

The literature on the causal relationship between financial sector development in general—and banking in particular—and economic development is inconclusive. Modern studies focusing on three decades of annual data on large panels of countries of varying levels of economic development suggest that financial sector development leads to economic growth. Despite the sweeping geographic scope of these studies, they consider relatively short periods of time: if the relationship between finance and growth changes over time or with the level of development—or if the development path of countries differs—studies with limited time horizons may not detect a relationship.

Although this book will touch on the relationship between banking and economic growth, its main goal is to discover the forces that drive the evolution of banking systems. If finance leads economic growth, as a number of studies suggest, then banking may well have an important effect on long-run economic growth. Even if no causal connection to

long-run growth is found, if banking structure and economic stability are causally connected, the shape of the banking system may have important consequences for short-term economic fluctuations.

SECURITIES MARKETS, BANKS, AND OTHER INTERMEDIATORS

Although the focus of this book is on commercial banks, it will be useful to place them in the context of the broader financial system. Economists ascribe a variety of functions to the financial system that are important in economic development (Levine 1997). These include facilitating trade in goods and services, providing liquidity services, mobilizing savings, allocating resources, facilitating the trading of risk, collecting and communicating information, and providing a means for monitoring managers and exerting corporate control.

Among the most basic services provided by the financial system are those that enable trade. As the use of coins spread from Lydia and Ionia throughout the ancient Mediterranean world, one of the earliest banking functions to develop was money changing, that is, the trading of moneys of different realms, or what we would today call the foreign exchange business.³ Another early banking function involved facilitating the transfer of funds from one individual to another—possibly separated by some distance—via book transfers and bills of exchange.

The economic impact of the development of money changing and remitting facilities was considerable, permitting the expansion of inter-regional and international commerce, which in turn promoted regional specialization in production. Specialization contributed to greater technological progress and increased consumption possibilities. In the words of Adam Smith (1776 [1994]: 18):

Among men . . . the most dissimilar geniuses are of use to one another; the different produces of their respective talents, by the general disposition to truck, barter, and exchange, being brought, as it were, into a common stock, where every man may purchase whatever part of the produce of other men's talents he has occasion for.

A second function of the financial system is the provision of liquidity services. Liquidity is “the ease and speed with which agents can convert assets into purchasing power at agreed prices” (Levine 1997: 692). By

³Davies (1994: 60ff.). Strictly speaking, money changing is distinct from banking, although money changers often subsequently became bankers. The development of money, which we do not discuss in detail here, is a crucial prerequisite for economic exchange. Without some form of money, all trade must take place through barter.

issuing banknotes and demand deposits, banks *create* liquidity. Hicks (1969) asserts that the provision of liquidity services was vital to industrial development, since industrialization requires large sums of fixed capital for extended periods of time. Enabling firm owners to sell a portion of their equity via securities markets, or to borrow on the security of these holdings from banks, reduces the liquidity risk to investors of tying up a large fraction of their wealth in one project and allows more investment projects to be funded.

A further function of the financial system is to accumulate the aggregate savings of an economy and channel it to areas where it can be put to productive use. This is particularly important if large quantities of capital are required, as is generally the case in industrialized and industrializing economies. The point is brought home by Postan (1935: 2), who argues that:

By the beginning of the eighteenth century there were enough rich people in the country to finance an economic effort far in excess of the modest activity of the leaders of the Industrial Revolution. It can, indeed, be doubted whether there had ever been a period in English history when the accumulated wealth of landlords and merchants, of religious and educational institutions would have been inadequate for that purpose.

The financial system also provides a mechanism for the trading and pooling of risk. For example, the owner of a company who sells shares on the stock exchange is effectively selling some of the accompanying risk to investors. Since individual investors can each purchase a small proportion of the shares offered, they are able to diversify their portfolios by purchasing a variety of assets with different risk characteristics. Similarly, banks pool risk by lending accumulated depositor funds to a variety of projects, hence reducing the risk to depositors from the failure of any one project.

Financial intermediators also provide information services and a mechanism for monitoring managers and exerting corporate control. Banks acquire information on the credit-worthiness of borrowers, the soundness of the collateral, and the likelihood of success of proposed projects prior to making loans. They may also be able to exercise oversight as to how the money is spent. This benefits depositors, since they do not have to gather information—either before or after the loan is made—on those to whom their money is lent. Securities markets gather and channel information. The up- or downwards movement of securities prices reflect the flow of information about the prospects for a given private or public debt or equity issuer. Further, because shares may be publicly traded, the opportunity exists for investors to buy a large enough share of a

company to effect a change in the management—this prospect may be higher if the share price has fallen due to poor management or some other factor (Mørck, Shleifer, and Vishny 1988, 1989).

We have, until now, spoken mostly about intermediation and the financial system in broad non-specific terms, noting the presence of both financial institutions and securities markets, but not dwelling excessively on their different characteristics. Although they perform many of the same functions, it is necessary to distinguish the two.

Securities markets, in their most basic form, are markets for debt (IOUs) and equity (ownership shares). Because of the generally poor quality of information, the earliest securities markets were primarily debt markets: traded debt consisted almost entirely of government debt, presumably with a good reputation for repayment,⁴ or by debt backed by inventories of goods (i.e., trade credit), which could be seized if the borrower did not pay off the debt.⁵ The number and types of securities markets have proliferated, and today include markets for types of debt, equity, and derivative securities that were not contemplated even a few years ago.

In a theoretical economic sense, securities markets provide the optimal form of intermediation: those who need funds for investment projects can offer their wares in the financial marketplace and savers can shop for the financial product with the attributes (e.g., maturity, risk, return, collateral) that suit their preferences (Rajan and Zingales 2003a,c). If the market is active, transaction costs are low, and all parties have sufficient information, buyers and sellers will get the best price possible. Market prices will continuously adjust to reflect changes in supply and demand for securities which, in turn, reflect new information about the underlying investments. If circumstances warrant, adjustments can take place through the market: firms and other demanders of funds can purchase their own debt or equity if the funds are no longer needed; savers can sell securities at any time and gain access to their funds.

These desirable qualities of securities markets are highlighted in Walter Bagehot's (1924: 12–13) classic description of the London money market: “English capital runs as surely and instantly where it is most

⁴Sovereign borrowers were not always credit-worthy, although they were likely more reliable than private borrowers. England's Charles II unilaterally stopped all payment of principal to his creditors in 1672, during the so-called “Stop of the Exchequer.” North and Weingast (1989) argued that by placing national finances under Parliamentary control, the Glorious Revolution of 1689 increased the credibility and creditworthiness—hence lowering borrowing costs—of the British government. See, however, Sussman and Yafeh (2004).

⁵The ability of the lender to seize collateral depends crucially upon a variety of institutional details, including property rights and enforcement mechanisms. LaPorta et al. (1997) spawned a substantial literature that focuses on the rights of creditors under Anglo-Saxon and continental legal systems.

wanted, and where there is most to be made of it, as water runs to find its level.”

Despite the many theoretical advantages of securities markets, the conditions necessary to secure those advantages—an active market with many buyers and sellers, low transaction costs, and ample information—have historically been rare. Even with today’s deep and liquid securities markets, the transaction and information costs (e.g., securing a bond rating, complying with government and exchange requirements) for small firms or firms in less developed countries may render access to securities markets impractical. Historically, the obstacles were much higher: active securities markets were few and far between, high transaction costs limited market access to only the largest and most credit-worthy borrowers, and information flows were poor by contemporary standards. Without solutions to the information problem, capital may not flow in sufficient quantities to where it is most wanted, but to where high transaction costs and imperfect information direct (Stiglitz and Weiss 1981).

Although both securities markets and banks are vulnerable to imperfect information and high transaction costs, banks may be superior to securities markets in coping with these defects, particularly in the early stages of development (Sylla 1998: 84). As lenders, banks can screen individual loan applicants and evaluate the credit-worthiness of proposed projects, thus making funds available to borrowers that are neither very large nor particularly well-known. Banks can also monitor the progress of projects, assess the need for additional funding, and evaluate the behavior of the managers, sometimes through a seat on the board of directors (DeLong 1991). In the course of these operations banks may develop expertise in evaluating firms in a particular industry or geographic region, lowering both transaction and information costs.

As borrowers, banks may also offer some transaction and information cost advantages over securities markets. Banks can mobilize substantial sums relatively cheaply by collecting deposits from large and small savers alike. By contrast, high minimum share prices, such as the £100 set by England’s 1844 Joint Stock Banking Act, discouraged the participation of less well-heeled investors. Additionally, a bank’s capital—committed to depositors in case of failure—and reputation may ameliorate the information asymmetry between banks and their depositors, which might otherwise discourage individuals from entrusting a bank with their savings.

THE SCOPE OF THIS BOOK

In some ways, the scope of this book is narrow; in others it is quite broad. It is narrow in that it focuses on financial institutions and largely ignores

securities markets, the other major component of modern financial systems. This narrowing is not as complete as one might think, however: many types of financial institutions exist, including commercial banks, savings banks, investment banks, life insurance companies, and credit cooperatives, among others. In 1963 the United States and most western European countries had between sixteen and twenty-four different types of financial institutions; Japan had as many as thirty-three. These numbers are not entirely the result of twentieth-century developments: as early as 1900, Belgium, Canada, Germany, Britain, Italy, Japan, the Netherlands, Norway, Sweden, and the United States each had more than ten different types of financial institutions (Goldsmith 1969: 354).

This book focuses on banks, with a particular emphasis on shareholder-owned incorporated banks that make loans financed by the issue of liquid liabilities, such as demand deposits and currency, which the bank must be prepared to redeem at short notice. Depending on the country, these institutions may be known as commercial banks, joint stock banks, chartered banks, credit banks, trading banks, national banks, or ordinary banks.

This narrow focus is warranted for several reasons. First, commercial banks have typically been among the largest and most influential of intermediaries, and in many countries have taken the lead in financing trade and industry, crucial factors in economic growth and development. Second, commercial banks have relied to a large extent on demand liabilities, highlighting their role both in overcoming the information asymmetries that may retard early economic growth and as providers of a large proportion of the means of payment.⁶ This reliance on liquid liabilities to fund less liquid loans has subjected commercial banks to greater risk of panic-induced withdrawals, in turn leading to financial crises, than other financial institutions (Diamond and Dybvig 1983). Finally, narrowing the scope makes the analysis far more tractable: because each type of financial institution has its own legal and economic origins, it is impossible to make an in-depth comparison of all types of financial institutions across multiple countries.

Commercial banks were not the only institutions to make loans financed by the issuance of money-like instruments: government banks and private banks frequently undertook these tasks long before commercial banks emerged. Government banks, however, were a special case of commercial bank: although frequently created as ordinary commercial banks, they typically had a special relationship with the state—both in terms of rights and responsibilities—that distinguished them from other commercial banks. The quasi-public role of these banks increased

⁶Savings banks, credit cooperatives, and other thrift institutions have also relied heavily on deposits; however, primarily on less liquid savings deposits.

during the late nineteenth and early twentieth centuries as they evolved into modern central banks. Although private bankers were sometimes important members of the financial system (e.g., the Rothschilds), most were typically smaller than commercial banks. As incorporated commercial banking progressed, private banks became, in aggregate, much less important.

An additional practical hindrance to a systematic study of private banks is the lack of data. Commercial banks were typically brought into existence with government sanction in the form of a charter and were required to provide periodic accounts to government authorities. Private banks were generally under no such obligation, and our knowledge of them is much more limited, coming primarily from company histories which rely on archival sources.⁷

The narrow focus on commercial banks comes at a cost, however. Guinnane's (2002) analysis of the nineteenth-century German financial system demonstrates that restricting such an analysis to commercial banks omits several crucial intermediators, including savings banks, mortgage banks, and credit cooperatives. Savings banks, both government-sponsored and private, have long been important parts of the financial system, particularly in Germany and the Scandinavian countries, in some cases exceeding the aggregate size of the commercial banking sector. Although these institutions were often important intermediators, their role was usually quite different from that of commercial banks. Savings banks typically collected funds from small depositors and made long-term loans to households, often for housing finance, or purchased long-term government securities, the proceeds of which were loaned by government-sponsored institutions. Although savings banks have played an important role in many countries' financial systems, because they have not been a major source of funds for trade and industry they will not take center stage here.

Geographically and chronologically, the scope of this book is quite broad, covering western Europe, Australia, Canada, Japan, and the United States from the beginnings of industrialization. These countries were both among the first to develop modern commercial banking systems and the first to industrialize, typically accomplishing both before the end of the nineteenth century. Hence, they all have relatively long continuous histories of both banking and industrial development. The broad coverage is warranted on two counts. First, since banking systems typically evolve slowly, determining the forces generating such evolution

⁷See, however, Pressnell (1956) on English country banks in the eighteenth and nineteenth centuries and Jonker (1996a) on a variety of private Dutch financial institutions in the first half of the nineteenth century. See also Joslin (1954) and Temin and Voth (2008).

can only be discerned by analyzing a long time period. Consider, for example, banking regulation. Limiting the study to the last quarter of the twentieth century constrains its focus to *deregulation*. Since the historical experience has included—and, given the recent turbulence in the financial system, the future certainly will include—both periods of increased regulation, as well as deregulation, a complete picture of the forces driving regulatory changes can only be discerned by examining a longer time period. Furthermore, because the cause-and-effect relationship between banking regulation and other economic and political events is complicated, examining a longer time series will make it easier to disentangle the channels of causality. Second, since the path of regulation and deregulation across countries has been highly idiosyncratic, a systematic assessment can be made only by examining the experiences of a large number of countries.

Given the above argument in favor of breadth, it would seem logical to expand the study to include a longer time period and more countries. Doing so would be problematic for two reasons. First, commercial banking only emerged in force in the industrialized world in the nineteenth century. Banking in the seventeenth and eighteenth centuries consisted primarily of government banks and largely undocumented private banks. Second, given the close relationship between banking and industry, it is reasonable to conclude that banking evolution differed systematically between the industrialized and the non-industrialized worlds.

The approach of the book is thematic and comparative, focusing on the forces generating important structural changes in banking across countries. The book is by no means exhaustive or encyclopedic: even the three chapters on individual countries do not present comprehensive banking histories. Rather, the book considers the major forces—crises, bailouts, merger movements, and regulation—that have shaped banking systems in a comparative framework. The bulk of the book is focused on the period from the beginnings of industrialization through the Great Depression of the 1930s. A final chapter assesses developments from the Depression through the turn of the twenty-first century.

THE ARGUMENT

A main thesis—and organizing principle—of this book is that the four types of structure-altering events discussed above (crises, bailouts, merger movements, and regulation) combined to generate an archetypical life cycle in the commercial banking systems of the industrialized world from their earliest days through the first third of the twentieth century. A stylized view of that life cycle is presented in figure 1.1. Although the pat-

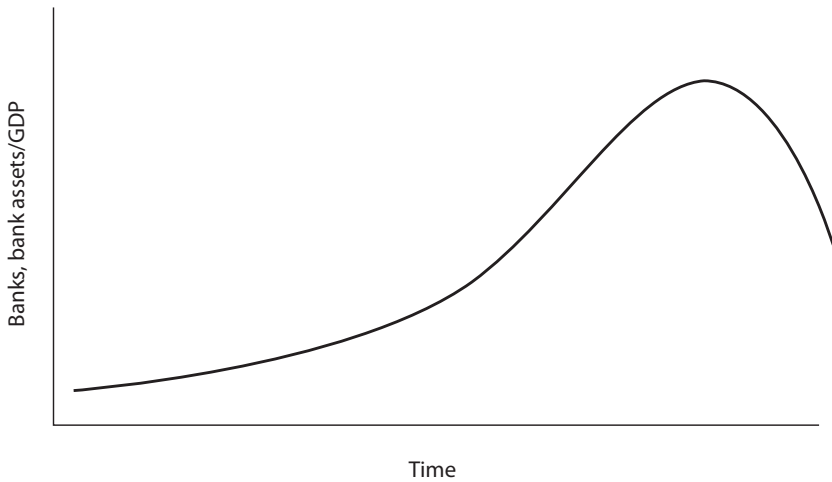


Figure 1.1. A stylized view of the life cycle of banking systems.

tern presented shows a smooth rise and fall in the number of banks, in fact, national experiences differed substantially and the pattern outlined below was rarely smooth. Nonetheless, a broadly similar pattern holds across countries: a period of expanding bank population, followed by a peak, followed by a more or less rapid decline. A theoretical and empirical literature on product life cycles posits a similar pattern in competitive industries.⁸ Although the life cycle posited by this literature is similar to the banking pattern presented here, there are a number of important differences, which are highlighted in the description that follows.

The introduction of commercial banking, as with any successful industrial innovation, leads to rapid growth at first. The speed and strength of this early growth depends positively on the demand for banking services, derived from the current and expected future financing needs of commerce and industry, and negatively on the severity of the regulatory constraints placed on banks and competition from other intermediators. Commercial banking growth may be strengthened if it takes over some of the functions of preexisting financial intermediators (Lindgren 2002). This early growth is susceptible to a variety of economic and regulatory shocks and, as noted above, will not be in any sense smooth or constant. Although figure 1.1 implies that paths of both the numbers and assets of banks move together, this need not be the case.

⁸Gort and Klepper (1982), Klepper and Graddy (1990), Klepper (1996, 2002), Klepper and Simons (2005), and Jovanovic and MacDonald (1994).

The general upward trend continues until a turning point is reached. The turning points in the number and assets of banks may not be contemporaneous in any given country, nor need they take place at the same stage in any particular country's economic development, but usually occur in close proximity to a banking crisis or a merger movement. Both of these events reduce the numbers of banks: crises typically lead to a reduction in aggregate banking assets, while mergers may be accompanied by a continued increase in banking assets. This stands in contrast to the product life cycle literature, in which the turning point coincides with a shakeout, often connected to the adoption of new technology.

Crises generate two reactions, both of which can have profound effects on banking structure. In the short run, government or private actors may attempt to rescue some or all of the failed banks. In the slightly longer run, governments may impose more stringent regulations in an attempt to bolster banking stability. Mergers, if they result in a substantial increase in banking concentration, may also encourage regulators to intervene in order to reduce the likelihood of a monopolized banking sector. Government efforts to rein in banks via regulation—whether due to crises or heightened merger activity—will be met with increasingly inventive means of circumventing such regulation, which itself may lead to yet further regulation. Edward Kane (1977) aptly refers to these cycles of moves and counter-moves as the “regulatory dialectic.”

Figures 1.2 through 1.10 present country-specific data on the number and assets of commercial banks from the nineteenth and early twentieth centuries. Before discussing the figures in detail, several warnings are in order. First, the data vary widely in quality: some series come from official sources, others from the financial press; some were compiled contemporaneously, others were assembled retrospectively. Second, they vary in coverage: some series contain gaps, others consist of two or more distinct series spliced together, and still others include more than one type of financial institution. Finally, for some countries comprehensive data are not available at all. Although future researchers will, no doubt, compile superior data, the current series are sufficient to illustrate the argument: banking systems experience a period of expanding population, followed by a peak, followed by a more or less rapid decline. In considering this pattern, I divide the life cycle into three phases: early expansion, later expansion and peak, and contraction.

Although the relationship between economic growth and financial development is not precisely understood, the data suggest that banking and economic development move together in the early phase of the banking life cycle. As noted above, the initial growth in banking depends positively on the demand for increased banking services brought about by economic growth, the rapidity with which commercial banking takes

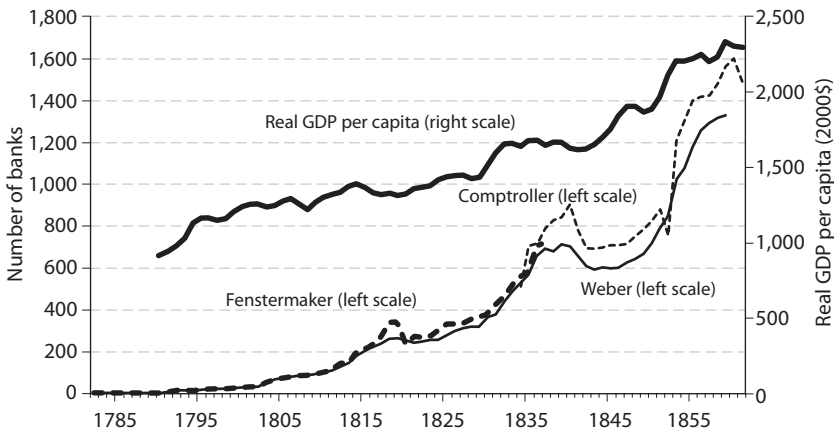


Figure 1.2a. Number of banks and real GDP per capita in the United States, 1782–1861. *Sources:* Number of banks: Fenstermaker (1965: 111), U.S. Comptroller of the Currency (1876: 94), and Weber (2005); real GDP per capita: Johnston and Williamson (2008).

over the functions of preexisting financial institutions, and negatively on the restraining hand of regulation and the presence of competing intermediators that retard the development of commercial banking.

Figure 1.2a presents data on the numbers of banks in the United States from the early expansion phase of U.S. commercial banking, 1782 to 1861, as well as data on the level of real gross domestic product per capita (i.e., total economic output per person) from 1790 to 1861. Banking data from 1862–1939 are presented in figure 1.2b. The data presented in figure 1.2a are suggestive: declines in the number of banks in the late 1810s, early and late 1840s, and around 1860 coincide with or are preceded by declines in real GDP per capita; rapid banking growth during the 1830s and 1850s coincide with strong economic growth. Although this correlation is notable, it does not answer the question of whether banking expanded to meet the needs—contemporary or anticipated—of commerce and industry, or if real economic activity expanded because of the increase in the numbers and activities of banks.

Similar patterns can be seen in the growth of Australian and Canadian banking during the first half of the nineteenth century (figure 1.3). Banking expansion was particularly strong during boom periods (e.g., the Australian gold rush of the early 1850s); declines occurred during severe economic downturns (e.g., Canada 1837, Australia 1841–43). The first half-century or so of Nordic banking also demonstrates the positive relationship between economic growth and banking development (figure 1.4). Denmark and Sweden, the most prosperous and rapidly growing of

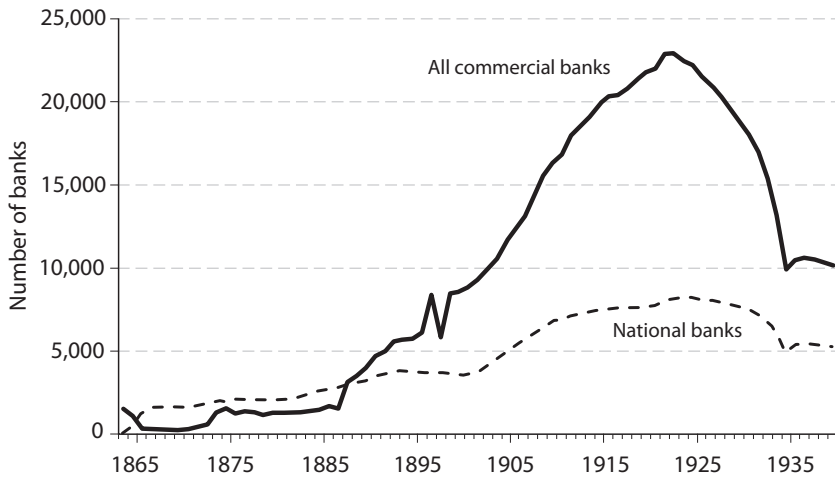


Figure 1.2b. Number of banks in the United States, 1862–1939. *Source:* U.S. Department of Commerce (1989).

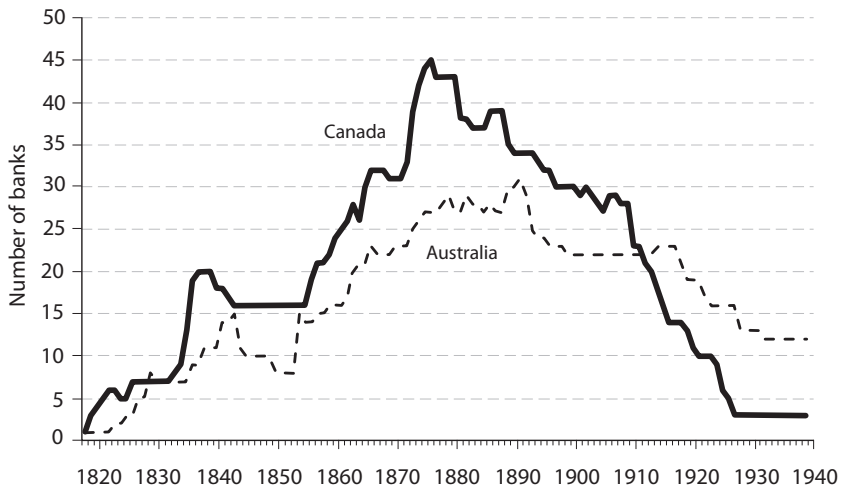


Figure 1.3. Number of banks in Australia and Canada, 1817–1939. *Sources:* Australia: Butlin, Hall, and White (1971) and Butlin (1986); Canada: Breckenridge (1910), *Canada Year Book* (various), Curtis, Taylor, and Michell (1931: 3–4), and League of Nations (1938).

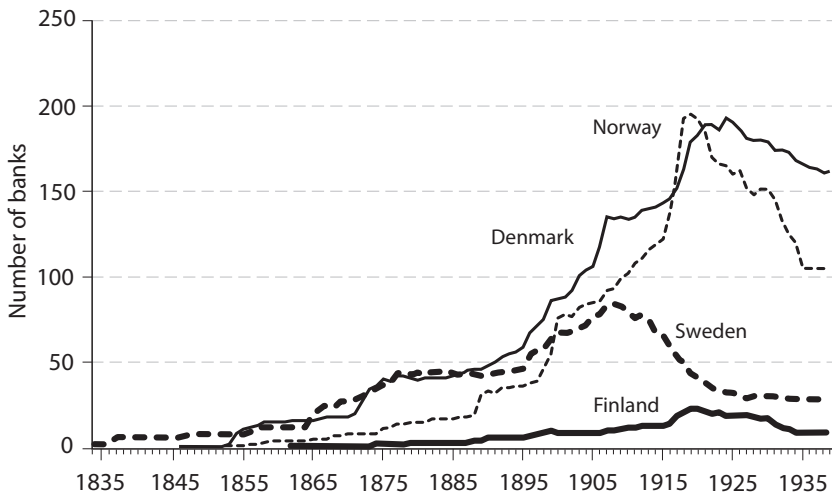


Figure 1.4. Number of banks in the Nordic countries, 1834–1939. Sources: Denmark: Danmarks Statistik (1969) and Johansen (1985); Finland: *Suomen Tilastollinen Vuosikirja* (various); Norway: Norges Offisielle Statistikk (1948); Sweden: Sveriges Riksbank (1931) and League of Nations (1938).

the Nordic countries, experienced more rapid growth in banking than relatively less developed Norway. Finland, which was the least economically developed of the four, saw the slowest and latest banking growth.

Economic growth—or lack thereof—was not the only factor limiting banking expansion during this early phase. Government restrictions sometimes retarded banking growth. The most obvious example is England,⁹ where legislation prevented the establishment of incorporated banks other than the Bank of England for more than a century prior to 1826, despite accelerating economic growth during the preceding several decades (figure 1.5).¹⁰ The removal of this restriction had dramatic results: more than one hundred such banks were established between 1826 and 1836.

⁹The terms “England” and “England and Wales” are used interchangeably throughout this book. Banks in England and Wales were governed by the same laws during the period under study and contemporaries generally did not distinguish between them (e.g., *Economist Banking Supplement*). The banking data employed are for England and Wales; crisis data (appendix 3.1 and 3.2) are labeled as England, even though other parts of the United Kingdom were frequently also affected; other data employed (e.g., GDP, population) are for Great Britain or the United Kingdom, as available. The words “Britain” and “British” are also used in a variety of circumstances (e.g., Parliament).

¹⁰Crafts (1983) and Harley (1982) note that economic growth during this period was slower than scholars had previously estimated. Nonetheless, they both acknowledge at least a gradual acceleration in the rate of economic growth starting in the last two decades of the eighteenth century.

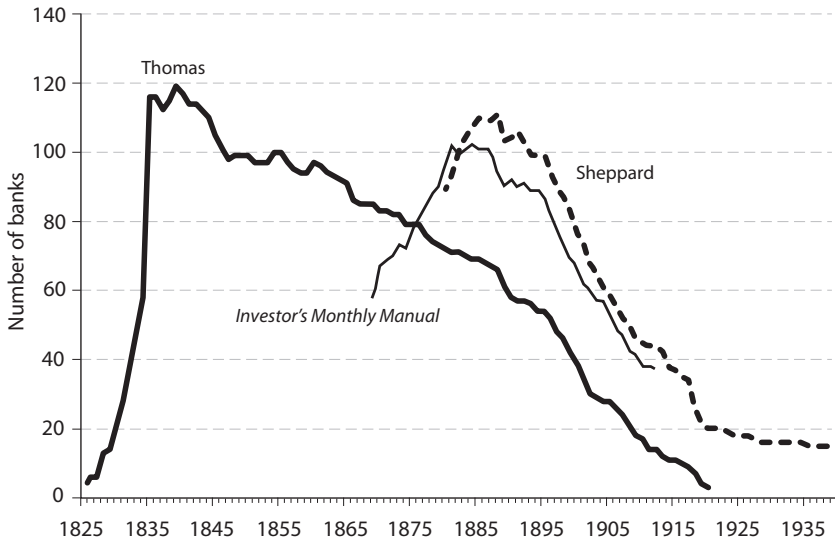


Figure 1.5. Number of banks in England, 1826–1939. Sources: *Investor's Monthly Manual* (various), Sheppard (1972), and Thomas (1934: appendix M).

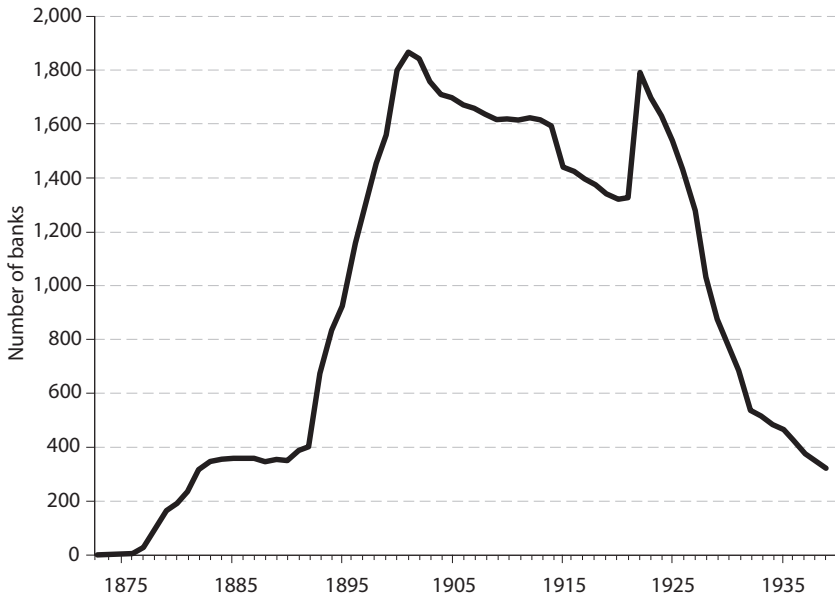


Figure 1.6. Number of banks in Japan, 1873–1939. Source: Bank of Japan (1966a).

Similarly, in Japan, regulations constrained growth during the early phase of commercial banking (figure 1.6). Under the 1872 National Banking Regulations (patterned on the U.S. National Banking Acts of 1863–64), commercial (“national”) banks were empowered to issue banknotes—redeemable in specie (i.e., gold or silver)—against the security of government bonds. Regulations promulgated in 1876 removed the requirement to redeem national banknotes in specie and, not surprisingly, the number of national banks and the size of the national banknote issue expanded dramatically. As a result of the over-issue, in 1879 the government prohibited the chartering of new banks, halting bank expansion for the next twenty years, and in 1883 it restricted the note issue to the Bank of Japan. Once new banks were again allowed to be chartered, under the Ordinary Banking Regulations, which took effect in 1893, the number of banks rose from 400 to more than 1,800 in less than a decade.

The number of institutions provides a good measure of the spread of banking during the early phase of commercial banking development, since most of the increase results from the entry of new firms. As banking becomes more established, however, a larger proportion of the growth comes through the growth of existing institutions, rather than from new entrants. Hence, in considering the later expansion, it will be useful to also consider measures of the size, as well as the number, of banks. Unfortunately, information on aggregate bank assets is even more fragmentary than data on bank population.

Banking growth during the later expansion phase occurs primarily through the expansion of existing banks rather than from the entry of new banks. In Canada, for example, the peak number of banks was reached in 1875; however, bank assets as a fraction of total economic output continued to rise until 1927 (figure 1.7). Similarly, the number of banks in Sweden peaked in 1907, although bank assets as a share of total output did not reach a peak until 1922 (figure 1.8). And in England the number of commercial banks peaked around 1890, while bank assets continued to grow until the 1930s (figure 1.7). In each of these countries, aggregate banking assets grew despite a decline in the number of banks, indicating that the continued growth resulted in part from an increase in bank merger activity. Heightened merger activity also led to the growth of branch networks, as newly acquired banks became part of the branch networks of the banks that purchased them.

In other countries, such as Denmark, Finland, and Norway (figure 1.8), where both the number and assets of commercial banks reached peaks shortly after World War I and prior to postwar slumps, and in Japan and the United States (figure 1.9), where peaks occurred prior to the depths of the Great Depression, the turning points in numbers and

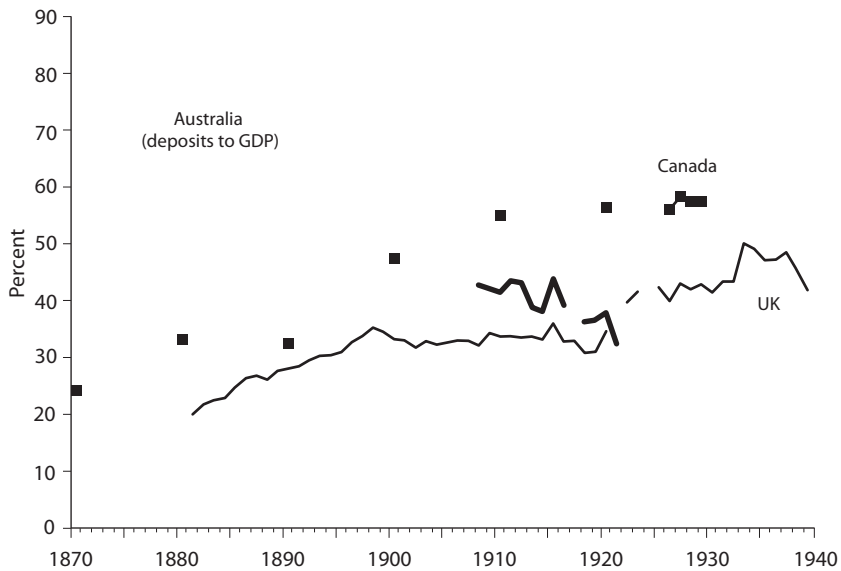


Figure 1.7. Ratio of bank assets to GDP: Australia, Canada, and England and Wales, 1870–1939. *Sources:* See figures 1.3 and 1.5 and Urquhart and Buckley (1965). *Note:* The ratio for England and Wales consists of assets of bank in England and Wales divided by GDP for the entire United Kingdom.

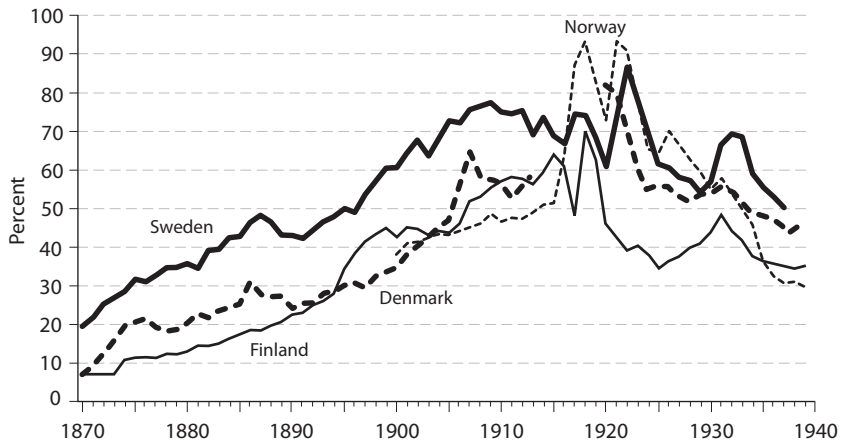


Figure 1.8. Ratio of bank assets to GDP: The Nordic countries, 1870–1939. *Sources:* Bank assets: see figure 1.4; GDP: Mitchell (1978) and Hjerpe (1989).

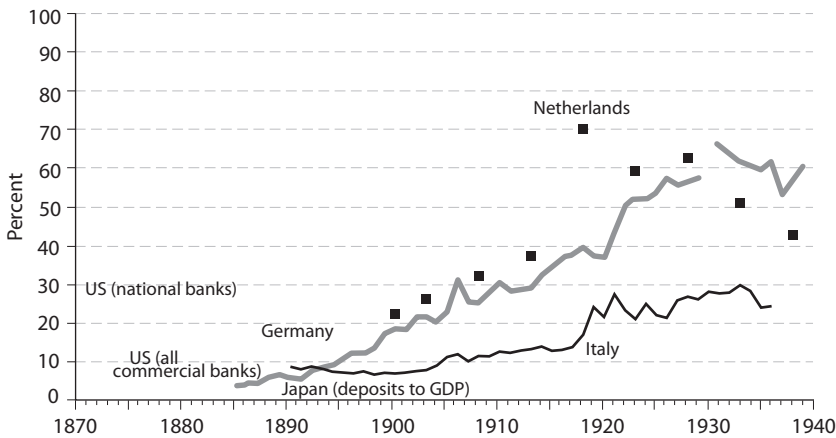


Figure 1.9. Ratio of bank assets to GDP: Germany, Italy, Japan, the Netherlands, and the United States, 1870–1939. *Sources:* Bank assets: Deutsche Bundesbank (1976), Cotula et al. (1996), Bank of Japan (1966a), Nederlandsche Bank (1986), and U.S. Department of Commerce (1989); GDP: Mitchell (1989, 1995) and Balke and Gordon (1986).

assets were chronologically close. In these countries, the peak typically coincided with the onset of a banking crisis.

The later expansion is followed by a decline—either rapid or not so rapid—in the banking population. This decline generally occurs as the result of a banking crisis or a merger movement, and can be augmented by the growth in popularity of alternative intermediators, such as savings banks. Banking crises lead to a reduction in both the number and the aggregate assets of the banking sector; mergers thin the ranks of banking institutions, but they need not lead to a reduction in aggregate bank assets. Banking crises may also lead to bank mergers, as failing banks are absorbed by healthier institutions, and so increases in mergers and bank failures may occur at the same time.

More fragmentary bank population data, starting with the later expansion phase, are presented for Belgium, Germany, Italy, and the Netherlands in figure 1.10. Each of these countries follows the general life cycle outlined above. Germany's banking population reached a peak in 1908, amidst a gathering bank and industrial merger movement.¹¹ The banking populations of Belgium, Italy, and the Netherlands reached peaks in the

¹¹Riesser (1911: 602ff.) ascribes the merger movement to a variety of factors, including an industrial merger movement, the growth of “communities of interest” in banking, as well as measures which hindered firm access to the stock exchange and hastened the need for even larger banks.

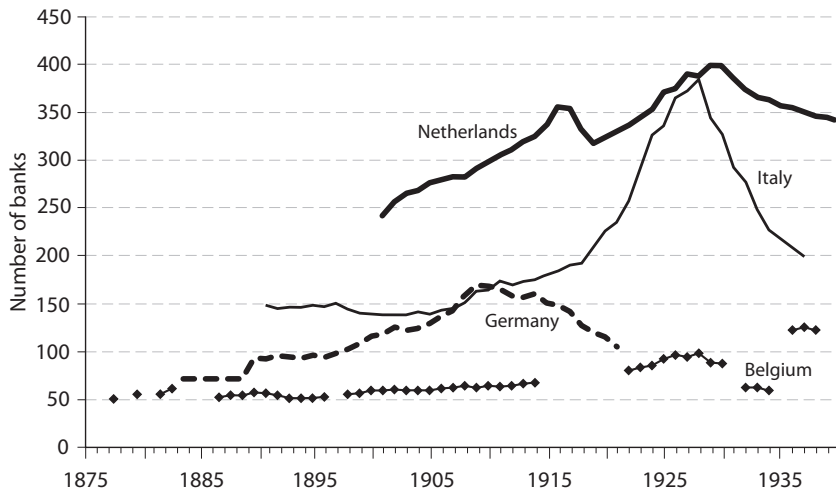


Figure 1.10. Number of banks in Belgium, Germany, Italy, and the Netherlands, 1877–1939. Sources: *Moniteur des Intérêts Matériels* (various), Deutsche Bundesbank (1976), Cotula et al. (1996), and Nederlandsche Bank (1986).

later 1920s, just prior to the onset of the Great Depression.¹² The relatively late—and dramatic—rise in the number of Dutch banks resulted from the fact that banks in the Netherlands had faced tough competition during much of the later nineteenth century from an alternative funding source: the well-developed money market. The decline of this market, especially during World War I, spurred a rapid growth in banking.

No matter where countries were within the life cycle described above, the Great Depression and World War II halted the process. The enormity of the financial devastation caused by both of these events led countries to impose rules and regulations that severely constrained banks. It was hoped that this financial lockdown would usher in a period of greater banking stability. And, in fact, these constraints, combined with the healthy economic environment of the post–World War II period, led to nearly three decades of uninterrupted banking stability. This stability, however, came at a cost. Even though banking grew and prospered, the strict constraints under which it operated ensured that there would be no substantial innovations in the business of banking. When market interest rates rose in the late 1960s and early 1970s, the post–World War II constraints, which had been a relatively minor inconvenience until that

¹²The increase in the number of banks in Belgium in 1935 resulted from a law that forced commercial banks to divest themselves of their investment banking enterprises and spin them off into new entities.

point, made it impossible for the banking system to adapt to the new environment. This initiated a process of deregulation in the 1970s which continued, for the most part uninterrupted, until the early twenty-first century. The last quarter of the twentieth century and early years of the twenty-first have seen a resurgence of crises, bailouts, merger movements, and regulatory reform.

CHAPTER OUTLINE

The life cycle of banks described above provides the roadmap for this book. The next chapter describes the rise of commercial banking. Starting in antiquity, it discusses the evolution of the different elements that allowed for the emergence of commercial banking in the eighteenth and nineteenth centuries: private banking, bills of exchange, public debt, and government banks.

Chapters 3 to 6 take a comparative approach to the evolution of banking structure, comparing the forces leading to banking crises, bailouts, merger movements, and regulatory reforms across the industrialized world. Broadly speaking, this is what economists call cross-section analysis: comparing similar phenomena across geographic and other categories. The strength of this approach is that the same phenomena can be compared in different settings, making it possible to isolate the common factors behind them. Its weakness is that it ignores an important time-series element and thus may miss some of the interactions between these phenomena over time. For example, if crises lead to regulatory reforms, which lead banks to engage in new practices in order to avoid these reforms, which then lead to further reforms, comparing crises or regulatory evolution across countries may not catch this dynamic. In order to capture the interplay between the different factors that affect banking structure, the subsequent three chapters each focus on one country with a long banking history: England, Sweden, and the United States.

The final chapter concentrates on the last two-thirds of the twentieth century and the beginnings of the twenty-first, starting with the regulations imposed in the aftermath of the Great Depression and World War II that essentially ended the life cycle pattern outlined above. It considers the forces behind the trend toward deregulation beginning in the 1970s, and continues with a discussion of the crises, bailouts, mergers, and regulatory reforms that have followed. The great variety of developments during this period truly demands a book of its own: this is clearly not that volume. Nonetheless, given the possibility that the longer historical record may shed some light on more recent events, this book concludes with an abbreviated look at contemporary history.