Banking Crises

Richard S. Grossman

May 2016
ABSTRACT

Financial crises have been a common feature of the economic landscape for more than two centuries. The chapter defines banking crises, considers the type of costs that they impose, and outlines the most common causes of banking crises during the past 200 years. The remainder of the chapter considers five distinct historical periods: the nineteenth century, when the “boom-bust” pattern that would typify later crises became established; the inter-war period, which was punctuated by two major sets of crises (post-World War I crisis and the Great Depression); the post-World War II financial lock-down, which was characterized by a complete absence of banking crises; deregulation and the return of crises in the 1970s; and the subprime crisis that emerged in 2008 and the subsequent euro-zone crisis.

Forthcoming in Youssef Cassis, Richard S. Grossman, and Catherine Schenk, eds., The Oxford Handbook of Banking and Financial History, Oxford University Press. The author is Professor of Economics, Wesleyan University, Middletown, CT, Visiting Scholar at the Institute for Quantitative Social Science at Harvard University, Cambridge, MA, a Research Fellow of the Centre for Economic Policy Research, and a Research Network Fellow of CES/Ifo, rgrossman@wesleyan.edu. He thanks Jerry Caprio, Jeff Frieden, and Peter Temin for helpful comments.
Financial crises have been a common feature of the economic landscape for more than two centuries. Although their frequency, severity, and geographic distribution have varied widely since the advent of the earliest modern financial systems in the eighteenth century, with the exception of the period between the end of World War II and the early 1970s, financial crises have never been completely banished from the world stage. At the time of this writing, in the wake of the US subprime meltdown and in the midst of the ongoing European sovereign debt crisis, financial crises are more prominent than any time since the Great Depression. And since, as Kindleberger and Aliber (2011: 8) note, the industry producing books about financial crises is counter-cyclical in nature, the increased financial instability of recent years has led to an outpouring of popular and scholarly writing on historical and contemporary aspects of financial crises.1

This chapter begins by presenting a definition of a banking crisis. It subsequently considers the types of costs that they impose and outlines the most common causes of banking crises during the past 200 years. The remainder of the chapter considers five distinct historical periods: the nineteenth century, when the “boom-bust” pattern that would typify later crises became established; the inter-war period, which was punctuated by two major sets of crises (the post-World War I crisis and the Great Depression); the post-World War II financial lock-down, which was characterized by a complete absence of banking crises; deregulation and the return of crises in the 1970s; and the subprime crisis that emerged in 2008 and the subsequent euro-zone crisis.

1. Definitions and Data

There is no universally accepted definition of financial crisis. Minsky (1982: 13) asserts that a definition is unnecessary, since the major episodes can be identified by pointing. Kindleberger and Aliber (2011: 34) argue that a precise definition of crisis is complicated because the label historically has included a variety of different types of events, suggesting that

---

1 In his presidential address to the Economic History Association, Eichengreen (2012: 289) notes that “[this] has been a good crisis for economic history,” offering in partial support the sharp spike in journalists’ references to the term “Great Depression” following the failure of Lehman Brothers in September of 2008. See Lo (2012) for a review of 21 of the many recent books on the financial crisis.
“. . . the genus ‘crises’ should be divided into species labeled commercial, industrial, monetary, banking, fiscal, financial (in the sense of financial markets), and so on, or into groups called local, regional, national, and international.” In a widely cited book, Reinhart and Rogoff (2009) similarly analyze a variety of different types of financial crises (e.g., banking, currency, sovereign debt). It is, of course, impossible to completely sever these different aspects of financial crises. A currency crisis caused by panic selling of assets in the belief that a currency devaluation (or substantial depreciation of a currency under a floating exchange rate regime) is imminent might lead to a banking crisis if it entails large-scale withdrawals from banks in order to sell the suspect currency, as in the emerging market crises of the 1990s (Furman and Stiglitz (1998), Radelet et al. (1998)). Conversely, a banking sector collapse could raise doubts about the sustainability of an exchange rate and lead to a currency crisis. Similarly, a stock market crash might translate into a banking crisis, or vice versa.

A more precise definition of financial crisis is given by Eichengreen and Portes (1987: 10), who characterize it as “… a disturbance to financial markets, associated typically with falling asset prices and insolvency among debtors and intermediaries, which ramifies through the financial system, disrupting the market’s capacity to allocate capital…” With the appropriate emphasis on “insolvency among intermediaries,” the above description is suitable for this essay’s focus on commercial banking crises. Nonetheless, even this definition leaves the precise parameters of a banking crisis vague. How extensive must the level of insolvency among banks be to constitute a crisis? Clearly, the failure of one or two banks need not constitute a full blown crisis.2 On the other hand, there is no agreement on the level of banking distress that constitutes a crisis. Lindgren, Garcia, and Saal (1996) argue that a “sound” banking system is one in which “most banks” have positive net worth (i.e., assets exceed liabilities). Bordo et al. (2001), however, argue that a banking crisis occurs when the “most or all” of the aggregate net worth of the banking system is eroded.

Given the paucity of detailed data on bank failures and bank net worth from the nineteenth and early twentieth century, catalogues of banking crises that rely on quantitative

2 Grossman (1993) models the macroeconomic consequences of bank failures during the National Banking Era in the United States and finds that the number of failures must cross a numerical threshold to constitute a “crisis.” Turner (2014), in his analysis of banking stability in the United Kingdom, considers both bank failure and the behavior of bank stock prices. Unlike many of the authors considered here, Friedman and Schwartz (1963) do not define a banking crisis on the basis of the number of insolvent or failed banks, but on the currency-to-deposit ratio, arguing that a banking crisis leads to a rise in that ratio, as depositors withdraw deposits in favor of currency.
definitions are liable to be inconclusive. Grossman (1994) defines a banking crisis as occurring when: (1) a high proportion of banks fail (e.g., the United States during the Great Depression, Austria following 1873); (2) an especially large or important bank failed (e.g., France’s Union Générale in 1881, Scotland’s City of Glasgow Bank in 1878, Austria’s Credit Anstalt in 1931); or (3) failures of the type described in (1) or (2) were prevented only by extraordinary and direct intervention by the government or some other actor, through the declaration of a bank holiday, or a reorganization or nationalization of the banking sector (e.g., Italy’s banking reorganization following the crisis of 1931, the rescue of Baring Brothers in England in 1890). Other catalogues of banking and financial crises abound, although they are generally not comparable because the authors typically focus on different eras and types of crises (Kindleberger and Aliber (2011), Caprio and Klingebiel (2003), Bordo et al. (2001), Grossman (2010), Reinhart and Rogoff (2009), Turner (2014), and Lindgren, Garcia, and Saal (1996)).

2. Costs of Banking Crises

Banking crises have substantial and far-reaching consequences. The costs of banking crises are borne by four distinct entities: shareholders, liability holders, taxpayers, and markets.

Among the most directly, and most rapidly, affected are shareholders of failed banks. Since most banks today operate with limited liability, the worst that can happen to a shareholder of a failed bank is to see a decline in the value of those shares, possibly to zero. When shareholder liability was not limited, the costs to shareholders could be greater. In the United States during the late nineteenth and early twentieth century, banks chartered by the federal government and those chartered by a number of states operated with double (or even triple) liability, under which shareholders of failed banks, in addition to seeing the value of their shares decline to zero, could be called on to pay an amount equal to (twice) their initial investment in order to satisfy creditors. In Britain, liability of banks and non-bank firms was also often extended by the issuance of “uncalled capital,” that is, shares with a nominal value that exceeded the amount paid in by shareholders; shareholders could be called upon and, in the case of a failure, would be called upon to pay in the uncalled capital. Grossman (2001b) and Grossman and Imai (2013) find that double liability and extended liability often reduced bank risk-taking and with it, the risk of banking crises. Elsewhere, unlimited liability was the rule, meaning that shareholders’ personal wealth could be called on to pay off the bank’s debts. Following the
failure of the City of Glasgow Bank in 1878, owners of each £100 share were called on to pay in a total of £2,750 to settle the bank’s outstanding debts, leading to the insolvency of more than 85 percent of its shareholders, including one bank that had inadvertently acquired four shares (Checkland (1975: 471)).

Holders of demand liabilities—demand deposits and, historically, banknotes—are also among the first to suffer when banks fail. Because depositors and note holders have a higher priority claim to a failed bank’s assets than other creditors and shareholders, they should fare better in a bankruptcy than other claimants. Depositors in U.S. banks lost $1.3 billion during the Great Depression; this exceeded shareholder losses of $0.9 billion but as a percentage of total deposits was far smaller than that experienced by shareholders (Friedman and Schwartz 1963: 351). Note holders of the City of Glasgow Bank were paid in full. Even if depositors and note holders were eventually able to get their funds following liquidation or takeover, being without access to their money for a considerable period of time could have forced them to default on obligations of their own. Since the Great Depression, crisis-related losses to demand liability holders have been minimized by the adoption of deposit insurance schemes around the world (Demirgüç-Kunt and Kane (2002)) and the provision of state guarantees in countries such as Finland, Norway (Englund and Vihriälä (2003)), and Ireland (Honohan (2012)), with the attendant risk of moral hazard.

Governments, quasi-government agencies, and the taxpayers who support them frequently bear some of the costs of banking crises by assisting failing or failed banks. This assistance can take at least three forms. First, the government or deposit insurance agency can close the affected institutions, sell off the assets, and pay off depositors. This was one of the approaches taken by the US Federal Deposit Insurance Corporation and Resolution Trust Corporation during the financial crisis that lasted from the 1980s through the early-1990s (Federal Deposit Insurance Corporation (1998)). If a crisis leads to panic selling of assets, the amount realized from their sale will decline and the net cost to the taxpayer will rise. Second, the government may arrange—either through payment, the provision of regulatory capital, or moral suasion--another, presumably healthier institution to take over the liabilities of a failed bank. In Japan, for example, healthy banks had long been “encouraged” by the authorities to absorb failing banks, contributing to the severity of the country’s “lost decade” (Hoshi (2002)). The eventual cost of such an operation depends on the deal struck by the government and the
purchasing bank, and any subsequent change in the market value of the assets acquired. Third, the government may act more directly by contributing capital (with or without taking a management role), making deposits, or nationalizing the bank. This approach was undertaken in Germany during the Great Depression when, by 1932, the government had acquired 50 percent of the capital of the large commercial banks (Allen et al. (1938)), as well as in Switzerland and Romania. In Italy, the Istituto Mobiliari Italiano (IMI) and Istituto Ricostruzione Industriale (IRI) were created to take over the industrial participations, loan portfolios, and, in some cases, the lending operations of distressed banks (Allen et al. (1938), Gerbi (1956)). In Sweden, A/B Industrikkredit was created in the wake of the Kreuger financial crash to perform a similar function. In the United States, the government-established Reconstruction Finance Company both loaned funds to and purchased preferred shares of troubled banks (Butkiewicz (1995)). In a survey of banking distress during the last two decades of the twentieth century, Caprio and Klingebiel (2003) estimate the fiscal costs of bank recapitalizations as 8 percent of GDP in Norway (1987–93), 11.2 percent in Finland (1991–94), and 24 percent in Japan (1990–2003); Blass and Grossman (2001) estimate the fiscal cost of the Israeli bank shares crisis of the 1980s at about 20 percent of GDP.

Finally, banking crises can have disruptive effects on securities and currency markets, as well as macroeconomic costs. Doubts about the solvency of a country’s banking system can lead investors to pull their money out of the country. Such “capital flight” can lead to substantial volatility in exchange markets. Similarly, widespread fears about the stability of the banking system can lead to a disruption in securities markets, both because of the decline in bank share prices and because of distress selling by banks of their securities portfolios. Thus, banking instability that may be limited in scope can both develop into a full-fledged crises spread via contagion (Bordo and Murshid (2000), Kaminsky, Reinhart, and Végh (2003)). The accumulated effects described above, combined with the increased cost of credit intermediation (Bernanke (1983)), can lead to macroeconomic costs as high as 20 percent of GDP (Grossman (1993)).

Crises can have substantial effects on the structure of banking systems as well. These include a decrease in the number and aggregate assets of financial institutions, along with a corresponding increase in banking concentration, with attendant effects on competition. Crises can also lead to a greater role for the government in the management of affected institutions.
Finally, crises often lead policy makers to make changes to banking regulations. Reforms aimed at making the banking system more resistant to crises are politically popular in the weeks and months following crises; at the same time, interest groups take advantage of the sentiment in favor of reform to advance their own agendas. And even when crises do not lead to changes in the rules and regulations, they can often lead to changes in accepted practices as bankers seek to reassure depositors and shareholders in a post-crisis environment.

3. Fear and Greed

Banks’ vulnerability to insolvency and, hence, crises, arises from the nature of their balance sheets, that is, their assets, what they own or are owed, and their liabilities, what they owe. Cash and deposits at the central bank are a bank’s safest assets. These pose little risk of loss, since they are unlikely to decline in value and are liquid, meaning that they can be quickly and cheaply employed to meet depositor withdrawals. Other categories of assets, such as securities (usually debt but, in some jurisdictions, equity as well) and loans, typically yield higher returns, but are subject to two types of risk. First, unlike cash, the price of debt and equity will rise and fall with the fortunes of the issuing entities, as well as with overall market conditions. For example, an economic downturn that leads to the default of borrowers can severely impair the value of a bank’s loans, while an increase in market interest rates can lead to a capital loss on its debt securities. Second, neither securities nor loans are as liquid as cash or cash-like assets. Although organized secondary markets exist for many types of securities, allowing them to be converted into cash, these markets do not exist for all types of securities and are largely absent for loans, hence the costs involved in converting them into cash can be prohibitively high.

The largest category on the liabilities side of the balance sheet is deposits. Today, deposits consist mostly of demand deposits, which have to be redeemed on demand by the bank. Historically, when banks issued their own notes, this circulating currency was an important component of demand liabilities, redeemable in metallic money, notes of the central bank, or some other liquid asset at 100 percent of its value upon demand. Because a substantial portion of bank liabilities are payable on demand, while most of their assets cannot immediately be converted to cash, banks are subject to “runs,” substantial withdrawals of deposits or requests for note redemption that may exceed the bank’s ability to satisfy these requests (Diamond and
Dybvig (1983), Calomiris and Kahn (1991), Allen and Gale (2007), Goldstein and Razin (2013). Although the likelihood of a run depends on both the asset and liability sides of the bank balance sheet, bank closures are typically precipitated by an inability to meet withdrawals of demand liabilities.

The ability of a bank to withstand shocks depends in large part upon the composition of its balance sheet. In making balance sheet decisions, bankers are confronted by two impulses, which can be conveniently summarized by the terms “fear” and “greed.” The higher a proportion of a bank’s assets held in cash, the greater the likelihood that it will be able to meet depositor demands under any circumstances. The main reason for bankers to hold more cash is fear that they will have insufficient reserves to meet depositor demands. On the other hand, because cash yields no return bankers have an incentive—call it “greed”—to increase their holdings of earning assets such as loans and securities. And, indeed, higher yielding assets do carry a greater risk of suffering a substantial decline in value—or even default—than cash. Thus, bankers have a strong motivation to hold a higher proportion of their balance sheet in earning assets and to bias their distribution toward riskier categories of assets.

Another important element of a bank’s balance sheet is capital, which plays several roles. First, it provides a buffer against a shortfall in cash flow. In a Modigliani and Miller (1958) world, with no taxes, bankruptcy costs, or agency costs, firms will be indifferent between funding their operations with debt or equity. Equity has some advantages from the standpoint of banking stability, however. Unlike debt, which must be serviced regularly and promptly, dividends on equity can be suspended without existential consequences, freeing up money to pay depositors and other creditors. Second, if a bank is forced to close, capital serves as a reserve that can be called on to liquidate unpaid debts. Third, higher holdings of capital can encourage banks to undertake less risk: because capital is at risk in case of failure, banks have an incentive to avoid risks that might put them out of business. Fourth, because banks are better informed about the soundness of their operations than investors (i.e., information asymmetry), the decision to hold more capital and to subject owners to a greater loss in case of failure signals to depositors and potential investors that the bank will undertake less risk than it otherwise might.

---

3 This assumes that the interests of bank managers and shareholders are aligned, in other words, that there is no principal-agent problem (see Glassman and Rhoades (1980)).
Holding capital may impose a cost on bank shareholders, however. In “normal” times—when there are no extraordinary deposit withdrawals and returns on earning assets exceed the interest paid to depositors—it is in the shareholders’ interest to fund as much of the bank’s earning assets with deposits as possible: the less dispersed the ownership, the fewer shareholders with whom to share the profits. Although much has been made of the costs and benefits of higher levels of bank capital—and the distribution of those costs (Admati et al. (2011), Haldane (2011), and Miles, Yang, and Marcheggian (2013))—from the enactment of the first commercial banking codes in Britain (1844) and Sweden (1846), through the establishment of the Basel (1988), Basel II (2004), and Basel III (2011) accords, regulators have established minimum capital levels in the belief that enhanced levels of capital encourage bank “soundness and stability” (Basel Committee on Banking Supervision (2004)).

4. Historical Pattern of Crises before World War I

According to Kindleberger and Aliber (2011: 26), “History views each event as unique. In contrast, economists search for the patterns in the data, and the systematic relationships between an event and its antecedents. History is particular; economics is general.” Although the genesis of each banking crisis is, in fact, unique, the vast majority of crises during the past two centuries can be traced to “boom-bust” macroeconomic fluctuations. Macroeconomic fluctuations have played a key role in models of financial crises at least as far back as Irving Fisher’s classic Booms and Depressions (Fisher (1932)). Fisher, who received Yale’s first Ph.D. in economics in 1891, was one of the first modern economic theorists to take an analytical approach to financial crises, arguing that financial crises result from the cyclical nature of real economic activity. In this view, economic expansion leads to a growth in the number and size of bank loans (and possibly also the number and size of banks) and an increase in the relative indebtedness of non-bank firms. The initial macroeconomic expansion might be

---

5 Other causes discussed in the literature include shocks to confidence (Friedman and Schwartz (1963: 308)) and structural weaknesses of the banking system (Grossman (2010: 63)). The Belgian banking crisis of 1838–39 illustrates both of these causes. This crisis was precipitated by the breakdown of diplomacy between Belgium and the Netherlands and the accompanying fear that a war might break out between the two countries. The crisis was further exacerbated by a bank war between Belgium’s two largest banks, Société Générale and the Banque de Belgique. Chlepner (1943).
generated by any one of a variety of factors: “…the outbreak of war or the end of a war, a bumper harvest or a crop failure, the widespread adoption of an invention with pervasive effects—canals, railroads” (Kindleberger and Aliber (2011: 28)). The boom is typically fed by an expansion of credit. During the period when gold and silver were important components of the money supply, credit expansion could be fed by the discovery of new sources of metallic metal or substantial gold imports resulting from a balance of payments surplus. As bank notes and bank deposits became more important sources of the credit supply, the expansion of these sources—or of central bank reserves, which serve as a backing for deposits—could generate the expansion in the credit supply.

As the economy continues to expand, investors continue to seek out more investment opportunities. International capital flows reinforce the development of the investment boom (Chinn and Frieden (2012)). Once the most profitable investments projects are funded, optimism about continued economic growth leads investors to fund less credit-worthy investment projects, resulting in an unsustainable buildup of debt during cyclical upswings. This increasing indebtedness leads Fisher to lament: “If only the (upward) movement would stop at equilibrium!” (Fisher (1932: 33)). But, of course, it does not. When the expansion ends, marginal firms—those that were the last to receive funding—are unable to meet their debt service obligations. The resulting loan defaults loans and declines in security values and the price level further exacerbates the distress. The debt-deflation spiral feeds on itself: loan defaults lead to the failures of banks and other intermediaries, exacerbating the macroeconomic downturn already underway.

The pattern of boom bust cycles had long been observed and was already more than a half-century old over 150 years ago when the British financial journalist D. Morier Evans (1859 [1969]: 1) observed:

Within the last sixty years, at comparatively short intervals, the commercial world has been disturbed by a succession of those terrible convulsions that are now but too familiar to every ear by the expressive name ‘panic.’ Each separate panic has its own distinctive features, but all have resembled each other in occurring immediately after a period of apparent prosperity, the hollowness of which it has exposed. So uniform is this sequence, that whenever we find ourselves under circumstances that enable the acquisition of rapid fortunes, otherwise than by the road of plodding industry, we may almost be justified in arguing that the time for panic is at hand.
The above scenario aptly describes many pre–World War I crises. For example, England suffered banking crises in 1825, 1836–39, 1847, 1857, 1866, and 1890. Each of these was preceded by several years of rapid economic growth, frequently fuelled by abundant harvests and gold inflows, and accompanied by increased speculation. The object of speculation varied from crisis to crisis and included at different times Latin American investments, limited liability companies, joint stock banks, railroads, and grain. The United States similarly endured banking crises on a regular basis during the late nineteenth and early twentieth century, with severe crises in the United States in 1837, 1857, 1873, 1893, and 1907. Additional examples from Australia, Canada, Western Europe, and Japan abound (Grossman (2010: 297-313)). The admittedly sketchy quantitative data available for the 30 years before World War I from nearly a dozen countries supports the boom-bust theory. Comparing business cycles that were marked by banking crises with those that were free from them indicates that crisis-cycles were preceded by faster economic growth, more inflation, and a greater percentage increase in the number of banks and quantity of bank assets than cycles that did not culminate in banking crises (Grossman (2010: 69)).

As the nineteenth century progressed, crises grew more international in scope. Kindleberger and Aliber (2011: 166) cite McCartney (1935), who argues that the crisis of 1873 was the first truly international crisis, erupting in Austria and Germany in May, spreading to Italy, Holland, and Belgium, and then to the United States in September, and later to Britain, France, and Russia. This was particularly true of the major crises of 1873, 1893, and 1907 and, as we will discuss below, of the banking crises of the Great Depression.

5. The Interwar Period

Banking crises were common before World War I. Britain and the United States each experienced about one crisis per decade during the nineteenth and early twentieth century and virtually every country that had a banking system had some sort of banking crisis during this period. The interwar period, however, stands out for both the frequency and severity of its banking crises. In the United States approximately 11,000 banks suspended in 1933 alone; although many of these banks reopened relatively quickly, the total number of commercial banks in operation fell by more than 40 percent between 1929 and 1933. And in Austria, Belgium,
France, Germany, Italy, Norway, and Switzerland one or more of the country’s largest banks collapsed, along with many smaller ones.

Whatever other factors may have been at work generating the banking crises of the interwar period, macroeconomic fluctuations were by far the most important source of banking instability. This is straightforward to demonstrate: the interwar period was subject to two sharp “boom-bust” macroeconomic cycles, each of which was accompanied by a spate of banking crises. The first began with the end of World War I and the subsequent rebuilding boom. The collapse of this boom around 1921 led to banking crises in Belgium, Denmark, Italy, the Netherlands, Norway, and Sweden. The collapse of the second boom-bust coincided with the onset of the Great Depression in 1929–30. The causes and extraordinary severity of the Great Depression and the role of banking crises in its propagation are the subjects of a voluminous literature, much of it still inconclusive.8 Although the answers provided by this literature are far from definitive, it is clear that the banking crises of the Great Depression were driven in large part by cyclical factors: the expansion of real economic activity in countries that subsequently endured banking crises was about 1 percent per year greater in the decade between the end of World War I and the beginning of the Great Depression than among those countries that did not experience banking crises.

In addition to macroeconomic factors, banking crises were precipitated by shocks to confidence. In the United States, Friedman and Schwartz (1963: 308) argue that bank failures in Missouri, Indiana, Illinois, Iowa, Arkansas, and North Carolina led to widespread attempts to convert bank deposits into currency, a “contagion of fear” that led to an intensification and spread of bank suspensions. Shocks to confidence also contributed to banking crises in Austria and Germany when their large outstanding short-term foreign debts led to doubts about the continued convertibility of their currencies, capital flight, and collapse of leading financial institutions.

Weaknesses in banking structure also played a role in crises during the interwar period. Although all countries were hit by a severe macroeconomic downturn, some banking structures were more resistant to banking crises than others. Countries that experienced banking crises

---

8 Prior to becoming chairman of the Federal Reserve Ben Bernanke (1995: 1) wrote: “To understand the Great Depression is the Holy Grail of macroeconomics… We do not yet have our hands on the Grail by any means…” Recent research has highlighted the importance of the gold standard in intensifying and propagating the Great Depression. Eichengreen (1992).
during the early 1920s were less likely to experience crises in the 1930s. For example, Denmark, the Netherlands, and Sweden all experienced crises in the early 1920s but had relatively stable banking systems during the 1930s. By contrast, several countries that experienced banking crises in the 1930s (e.g., Belgium, Switzerland) had been spared the worst of the post–World War I boom-bust banking crises. This may be accounted for by the fact that structurally weak banking systems in 1920 suffered crises during the post–World War I boom-bust episode, but the exit of many poorly performing banks left the surviving banking system more resistant to crises. Banking systems that had been more structurally sound at the end of World War I avoided the banking crises of the early 1920s and the accompanying shakeout of the banking system. It may be that a banking system that had been sound enough to survive a post–World War I-sized shock was not necessarily resilient enough to survive the even larger Great Depression-sized shock. Alternatively, it may be that countries that did not experience a post–World War I banking crisis grew too rapidly during the subsequent decade, increasing their vulnerability to crisis.

Data on a variety of aspects of banking structure bank branching across countries during the Great Depression further supports a role for bank structure in determining banking stability (Grossman (1994)). Banks in countries that did not have crises had, on average, substantially more branches per bank than their counterparts in crisis countries. Banks in England, Canada, and Australia were among the most extensively branched, although those in the non-crisis countries of Sweden, Czechoslovakia, and the Netherlands also had more branches per bank than all but one of the crisis country banking systems. The Canadian experience, especially as compared with that of the United States, lends support to the argument that branch banking contributed to stability. Macroeconomic performance during the Depression was dismal in both the United States and Canada, although the extensively branched Canadian banking system survived intact while approximately 5,000 U.S. banks, primarily unit banks, failed.

6. The Great Lockdown

The financial turbulence of the inter-war years initiated a period of increased regulation of the banking system, as governments strove to bolster banking stability. Prior to World War I, only a handful of countries (e.g., Canada, England, Finland, Japan, Sweden, and the United States) had comprehensive banking codes that limited entry into banking, set capital requirements, and spelled out a variety of other balance sheet requirements. Elsewhere, banks
typically came into existence and operated under general commercial law or other ad hoc arrangements (Grossman (2001a)). In the wake of the post-World War I crises and the Great Depression, Austria (1924), Belgium (1935), Denmark (1919), Germany (1931-34), Italy (1926), and Switzerland (1934) enacted their first ever banking codes; comprehensive banking law was not enacted until even later in France (under the Vichy regime in 1941) and the Netherlands (1952).  

Constraints on banks were further tightened under the exigencies of World War II, when countries enacted rules and regulations that allowed them to harness the resources of the banking system to finance the war effort. Many of these constraints persisted long after the conflict had ended. In February 1942, the Commonwealth Bank of Australia was given powers to set maximum interest rates as a wartime measure. At the war’s end, these powers were continued by the Banking Act 1945 and lasted until the 1980s (Australia. Parliament. House of Representatives. Standing Committee on Finance and Public Administration (1991: 20-21)). Similarly, in Germany, interest rates on deposits during World War II had been set by the German bankers’ association in “authoritative consultation with the Reich Control Office for Banking” (Irmler (1956: 322)). Following the war, the job was taken over by West German state governments, which maintained control until interest rate deregulation in 1967. In Japan, the government enhanced the role of banks during the 1930s, at the expense of securities markets, in order to better direct the flow of national savings toward military ends (Hoshi and Kashyap (2001: 15-89)). Sweden introduced credit controls at the beginning of World War II and interest rate controls at the war’s end, both of which remained in force for nearly a half-century (Jonung (1993)). And in the United States, where banking regulations by state and federal government had for decades been stringent relative to Europe, laws enacted during the 1930s established deposit interest rate ceilings, which persisted until the 1980s, and enforced a separation between commercial and investment banking which survived until the 1990s.

The constraints discussed above were not exceptional, but were the rule among industrialized countries. Interest rate regulation is indicative of the overall posture toward banking regulation. In 1960, interest rates in all twenty-two industrialized countries of the

---

9 Grossman (2010: 140-141), Myers (1949). Even Canada, which had endured the Great Depression with minimal banking instability, adopted a major reform: the establishment of the Bank of Canada. According to Bordo and Redish (1987: 417), the bank was founded because of the government’s desire to be perceived as doing something in response to the Great Depression, rather than from any pressing economic need for a central bank.
Organization for Economic Cooperation and Development (OECD) were set by official government policy or by state-sanctioned bankers’ cartels. By the end of 1980, interest rates in eighteen of those countries were still controlled; the number fell to nine by 1987. By the turn of the twentieth century, interest rate regulation had disappeared from the OECD (Bröker (1989)). Government-imposed constraints on banking were not limited to interest rate regulation, but covered a great deal of the banking business, including the products banks were allowed to offer and how they could use their funds. As a consequence, banking did not change very much in the three decades following the outbreak of World War II. According to one observer, “In most countries the business of banks did not change significantly from the late 1940s to the 1970s. The products offered, the management and sales techniques applied showed great consistency” (Steinherr (1992: 1)).

The banking regulations imposed in the aftermath of the Great Depression had been aimed at reducing financial instability. In this, they were largely effective. From the beginning of the Napoleonic Wars through World War I, not one decade had passed without a substantial banking crisis somewhere in the developed world. By contrast, after the end of World War II, nearly 30 years would pass before the industrialized world was again subject to a serious episode of banking instability. Thus, the period between the end of World War II and the onset of the 1973 oil crisis constitutes the longest sustained period of banking stability this industrialized world has ever know.

This stability came with a price, however. Many of the constraints imposed on banks reduced the scope for—and the need to—compete for business. Deposit rate ceilings, for example, meant that banks did not need to offer competitive rates to attract depositors. Implicit or explicit government guarantees, such as deposit insurance in the United States, eliminated the need for bank customers to monitor and, if necessary, discipline poorly performing banks. This reduced innovation in banking. According to Arthur Burns, a former chairman of the US Federal Reserve: “The law marked out a protected domain in which banks could profitably operate, and the banks tended to stay in that domain” (Burns (1988: 6)). Thus, banks were profitable, but the development of the banking business was retarded.

7. Deregulation and the Reemergence of Crises
Both despite and because of the restrictive nature of banking regulation during the decades immediately following World War II, banks grew and prospered. The total assets held by commercial banks in 16 industrialized countries amounted to less than one half of GDP in 1953; by 1973, it exceeded three quarters. And bank profitability during the 1960s and early 1970s exceeded that during the 1980s and 1990s (Grossman (2010: 257-258)). The impressive growth of banking during this period can be attributed in part to strong economic growth: per capita GDP grew more than twice as rapidly during 1950-72 as during 1870-1913 and 1.5 times as rapidly as the period 1973-2001. Robust economic growth and demand for banking services, combined with the anti-competitive nature of regulation, made the decades after World War II profitable ones for banks.

The restrictive banking regulation enacted during the Great Depression and World War II began to be loosened in the late 1960s. Liberalization included the elimination of interest rate controls on bank lending and deposit rates, the growth of market-oriented mechanisms for the allocation of capital (e.g., the development of money markets and auction techniques for the issue of government debt securities), the elimination of direct controls on bank lending, and allowing branching and foreign entry into domestic banking markets. The impetus to deregulate, discussed in more detail below, came primarily from inflation-induced high and volatile interest rates that accompanied the end of the Bretton Woods system of fixed exchange rates and the OPEC oil shocks of 1973 and 1979. The result of these developments was a substantial depreciation of the dollar, heightened exchange rate volatility, an increase in the volume of international capital flows, and an end to the post-World War II “golden era” of rapid economic growth. When Federal Reserve Chairman Paul Volker engineered a reversal of America’s previously expansionary monetary policy, short-term interest rates hit record highs.

Banks had been profitable as long as economic growth was robust and interest rates were low and stable. Consistently moderate interest rates were important, because banks profited from the spread between government-mandated low deposit rates and higher lending rates. Further, because interest rates were low and stable, the interest rate ceilings were not binding and their effects were muted. As interest rates increased, a gap grew between market interest rates and

---

10 Bordo and Eichengreen (1993: 465-66), Grossman (2010: 260 ff.). Although the oil shocks and the collapse of Bretton Woods were important factors in the heightened inflation and interest rates of the 1970s, U.S. rates had been rising during the rapid economic growth of the 1950s and 1960s, and had been given a push from increased government spending on the Vietnam War and the Great Society programs.
those that ban could pay their depositors. This led depositors to look for alternatives to low-interest-rate bank deposits. The subsequent disintermediation, particularly during the U.S. credit crunches of 1966 and 1969, weakened banks (Wojnilower (1980)). Regulators responded to the disintermediation, in part, by deregulation. Interest rate caps were loosened, and eventually scrapped, and the scope of permitted bank activities was substantially broadened. An unintended consequence of this newfound freedom was greater instability.

Deregulation was a liberating experience for banks, allowing them to broaden the range of activities in which they engaged, to increase flexibility in setting interest rates, and to compete for new and different types of business. The growth in international transactions due to the collapse of the Bretton Woods system and the improvement in the technology for making these transactions also provided outlets for bank expansion. This liberation came at a cost, however: banking, which had been constrained but safe, became freer but more vulnerable to crisis. The industrialized world’s banks became much more vulnerable after 1973. Bordo et al. (2001) look at the incidence, duration, and severity of banking and currency crises during 1880–1997. They find that the probability of a banking crisis occurring was similar during 1880–1913 and 1973–97, approximately double this rate during the turbulent interwar period of 1919–39, and essentially zero during 1945–71.

Although the banking crises of the last quarter of the twentieth century were, like their predecessors, caused primarily by boom-bust macroeconomic cycles, three additional powerful factors also contributed. First, the collapse of the Bretton Woods system led to enhanced opportunities for engaging in high-risk international transactions, and improvements in technology allowed larger risks to be undertaken more quickly and easily than ever before. Second, the increasing size of the government sector relative to GDP meant that peacetime changes in fiscal policy might contribute to a boom-bust cycle. Third, the rollback of Depression-era regulations gave banks more freedom to engage in other risky transactions. These causes are readily identifiable as being behind two of the earliest post-World War II banking crises, those associated with the 1974 failures of Bankhaus I. D. Herstatt and Franklin National Bank.

Herstatt, a Cologne private bank, lost an estimated $200 million in currency trading—a substantial amount, considering that the bank’s total assets had been estimated at $800 million just six months prior to its failure. Herstatt was closed by West German authorities around 3:30
pm local time, after foreign counterparties had made substantial irrevocable Deutsche mark payments to Herstatt but before those counterparties had received dollars in return. This was particularly problematic for U.S. banks, where the business day was not yet half over when Herstatt was closed (Galati (2002)).

On the same day it reported the failure of Herstatt, the New York Times noted that, “In a somewhat similar situation, the Franklin National Bank of New York had foreign exchange losses of $45.8 million in the first five months of 1974.” By October 8, 1974, Franklin had closed—at the time, the largest bank failure in American history (Kotsopoulos (1981: 322)). International exposure, combined with the Franklin’s acrobatics to avoid US regulations, were the main sources of its problems. Located on Long Island, New York, Franklin was chartered in 1926. It bank expanded rapidly during the 1950s, primarily through mergers and acquisitions: between 1950 and 1958, Franklin merged with or acquired 13 other banks in Nassau County. Franklin’s prosperity was aided by robust economic growth on Long Island during the 1950s and also because New York State banking law prohibited banks from branching across county lines, protecting Franklin from competition from banks in neighboring New York City. A 1960 law allowed banks to operate in contiguous counties and Franklin soon opened an office in New York City. Subsequent growth brought offices in the Bahamas and London, where Franklin entered the Eurodollar market, although not as a major player (Spero (1980: 41)). Franklin—and many other US banks—found expansion overseas enticing, because overseas branches of US banks were subject to neither US reserve requirements (Regulation D) nor deposit interest ceilings (Regulation Q).

The inauguration of tight monetary policy in October 1979, combined with regulatory restraints, had disastrous consequences for the US savings and loan (S&L) industry. An important supplier of home mortgage lending, S&Ls were subject to a classic interest rate squeeze when interest rates rose. S&L assets were legally restricted to long-term fixed rate mortgages, which were funded with deposits that were subject to interest rate ceilings. Higher market interest rates encouraged depositors to withdraw their funds from S&Ls. Because S&L loan portfolios consisted of relatively low-yielding mortgages, even if they had been free to pay higher deposit interest rates, it would have been too costly to do so.11

---

11 Even if a market for mortgages had existed at the time, high market interest rates would have meant that selling these low fixed-interest rate loans would have involved a substantial capital loss.
The legislative response to this squeeze was to broaden the powers of savings and loans, hoping that S&Ls could earn their way back to profitability. Laws passed in the early 1980s permitted S&Ls to issue demand and money market deposit accounts, to make personal and credit card loans, and loans for the acquisition and development of real estate. Laws regulating commercial banks were also liberalized, permitting them to undertake real estate lending and relaxing limits on the amount that could be loaned to one borrower. All of these changes encouraged the rapid expansion of banks and S&Ls into relatively risky types of lending and contributed to a series of real estate booms and busts in different parts of the country.

The deposit insurance system, created in the aftermath of the banking crises of the Great Depression, further contributed to US banking difficulties. Unlike most types of private insurance, deposit insurance premia did not distinguish between institutions that engaged in especially risky activities and those that did not. The ensuing moral hazard encouraged more risk-taking on the part of banks and thrifts. Moral hazard effects were enhanced by an increase in the deposit insurance limit to $100,000 from $40,000 in 1980 and by a federal hiring freeze in the late 1970s and early 1980s, which led to a decline of about 20% in the number of federal bank examinations, removing another potential constraint on greater risk-taking (Federal Deposit Insurance Corporation (1997: 57)). Moral hazard-induced problems were exacerbated by technological innovation. Because deposits were insured, banks and S&Ls in need of funds could raise virtually unlimited sums by offering above-market interest rates on insured certificates of deposit. The improved technology of money transfer made it possible for a national market in brokered deposits to develop, removing the need for banks and S&Ls to raise funds locally. These innovations encouraged troubled institutions to raise additional funds and to employ them in ever-riskier activities in, what Barth, Bartholomew, and Labich (1989: 25) aptly term “gambling for resurrection.”

Like the S&L crisis, the Nordic banking crises of 1987-1994 resulted from a series of boom-bust cycles which were amplified by banking deregulation (Koskenkylä (1994), Drees and Pazarbasioglu (1995), Jonung (2008)). From the end of World War II until the late 1970s, the Nordic countries tightly regulated their banking, credit, capital, and foreign exchange markets. Although these countries were industrially advanced, their stock and bond markets were not well developed and banks were the primary intermediators. Banks, however, were tightly constrained
by regulations designed to channel credit to favored sectors of the economy. Swedish financial institutions were subject to lending ceilings, which limited the rate of growth lending to low priority sectors, and liquidity ratios, which required banks to hold a substantial portion of their assets in bonds issued by the government and mortgage institutions (Englund (1999)). In Norway, a number of government lending institutions were responsible for allocating funds to the housing, manufacturing, and agriculture and fisheries sectors (Moe, Solheim, and Vale (2004), Knutsen and Lie (2002), Steigum (2004)). Banks in all three countries were subject to interest rate ceilings and in Finland the central bank set ceilings on each bank’s average and top lending rates (Englund and Vihriälä (2003)).

Restrictions on credit allocation led to unfulfilled demand for funds in many sectors. Excess demand was exacerbated by high inflation, low interest rate ceilings, and the generous tax treatment afforded to interest payments, which was particularly important in light of the high tax rates in the Nordic countries. All three countries began to loosen financial restrictions in the 1980s by removing interest rate ceilings on loans and deposits, permitting the development of CD, bond, and money markets, easing foreign exchange controls and allowing foreign banks to operate domestically, and removing quantitative controls on bank lending. At the same time, each reorganized their financial supervisory apparatuses, which in practice meant that the new supervisory regimes were not completely prepared for financial liberalization.

The response to deregulation was dramatic. Household and corporate indebtedness grew substantially and bank lending expanded rapidly through the early 1990s. Abetted by accommodating fiscal and monetary policies, aggressive lending increased the extension of credit to the real estate, construction, and services sectors, and the prices of stocks and real estate rose dramatically in all three countries. These booms collapsed in the late 1980s (Norway) and early 1990s (Finland and Sweden), leading to the failure or forced government take-over of a number of leading banks.

Like those of the Nordic countries and the United States, the financial system of Japan had been severely constrained in the period from the end of World War II until the mid-1970s (Hoshi and Kashyap (2000), Hoshi and Patrick (2000), Hoshi (2001)). Domestic stock and bond markets had been kept deliberately underdeveloped and foreign securities markets were off-limits, forcing both borrowers and lenders to rely upon the domestic banking system. Interest
rates were fixed and interest rate spreads were relatively high, guaranteeing banks high profits. Competition between banks was non-existent under the “convoy system” under which all institutions were allowed to grow at the same speed, and none was allowed to fail (Hoshi (2002)).

Deregulation began in the mid-1970s, and included liberalization of the money, debt, equity, and foreign exchange markets, each of which had been under formal or informal control. The motivation for liberalization was in part fiscal: the oil shock-induced fiscal deficits of the 1970s heightened the government’s desire to sell more bonds. Liberalization, particularly of the bond market, opened up alternative funding sources for corporations; deregulation aimed at expanding the options of depositors was slower to develop. Thus, banks that still had substantial deposit bases began to lose their large corporate customers to the bond market, which led banks to expand their lending to small- and medium-sized business, particularly to real estate developers. The resulting boom in real estate and stock prices was further supported by expansionary monetary policy undertaken by the Bank of Japan during the second half of the 1980s. A distinguishing feature of the Japanese banking crisis was the length of time it took for the authorities to abandon the convoy system and allow banks to fail (Hartcher (1998), Amyx (2004)). This delay in confronting Japan’s banking problems made what would have been a severe crisis even worse, leading to a decade characterized by slow economic growth, high unemployment, and a long, costly clean-up of its financial system.

8. The Subprime and European Sovereign Debt Crises

The 1990s saw a gradual reduction in the frequency and severity of crises. With the resolution of the Scandinavian and Mexican crises in the mid-1990s, and the Asian financial crisis and the collapse of the US dot-com bubble at the end of the 1990s, the first decade of the 2000s was shaping up to be one of the calmer eras in financial history. Economists had argued that the 1980s and 1990s has seen a gradual reduction in macroeconomic volatility (Blanchard and Simon (2001), Stock and Watson (2002)). And there was some sense that improvements in international rules and coordination (e.g., the Basel Accords, World Trade Organization) had diminished the likelihood of serious crises and substantially reduced the dangers of contagion (Kapstein (1996), James (2001)).
The US subprime crisis was, like many of its predecessors, a boom-bust crisis. Following the collapse of the dot-com bubble in 2000, the US embarked on a long and vigorous economic expansion. Economic growth was promoted by expansionary fiscal policy under President George W. Bush, who signed three tax cuts into law during his first three years in office and embarked on wars in Afghanistan and Iraq following the 9/11 terrorist attacks. This fiscal stimulus was compounded by expansionary monetary policy undertaken by the Federal Reserve, which lowered interest rates following the collapse of the dot-com bubble and 9/11 and kept them low through 2004. The fiscal and monetary policy-induced expansion led to a boom in the US housing market, which was supported by the development of new techniques for real estate finance (e.g., mortgage backed securities) which regulators allowed to grow essentially unchecked. As opportunities among credit-worthy borrowers began to be exhausted, lenders found new ways of extending credit to less credit-worthy borrowers (i.e., subprime mortgages). The collapse of the real estate market during the second half of the decade of 2000 led to widespread defaults among holders of low-quality mortgage debt.

The subprime crisis quickly spread to Europe. Although Europe was not the prime instigator of the crisis, it was more than just an innocent bystander. European financial institutions had purchased substantial amounts of mortgage-backed and related derivative securities and took heavy losses when the prices of these plummeted. Further, European banks had borrowed relatively more than American banks, so when these losses were realized the financial distress in Europe was at least as severe as in the United States (Eichengreen (2009)). And although much has been made of the run-up in US housing prices, housing prices in some parts of Europe rose quite dramatically, such as Spain, Ireland, Belgium, the Netherlands, and France. The collapse of the booms in these high-flying real estate markets led to further problems, both for financial institutions and, importantly, for the national governments that bailed them out (Gros (2006), Hilbers et al. (2008)). The commitment to the euro further complicated the crisis in the eurozone, by removing the option of any affected country to address the crisis with monetary policy.

Banking crises have been a resilient phenomenon during the past two centuries. Although many of the details and circumstances differ across crises, the fundamental boom-bust pattern identified as early as 150 years ago by Evans (1859 [1969]), and analyzed by Fisher
(1932), Minsky (1982), Kindleberger and Aliber (2011) stands as a useful model for analyzing financial crises today.
References


