

THE PAST MIRROR: NOTES, SURVEYS, DEBATES

Banking crises yesterday and today¹

CHARLES W. CALOMIRIS

Columbia Business School and NBER

cc374@columbia.edu

Pundits, policy makers and macroeconomists often remind us that banking crises are nothing new, an observation sometimes used to argue that crises are inherent to the business cycle, or perhaps to human nature itself. Charles Kindleberger and Hyman Minsky were prominent and powerful advocates of the view that banking crises are part and parcel of the business cycle, and result from the propensities of market participants for irrational reactions and myopic foresight.² Some banking theorists, starting with Diamond and Dybvig, have argued in a somewhat parallel vein that the structure of bank balance sheets is itself to blame for the existence of panics; in their canonical model, banks structure themselves to provide liquidity services to the market and thus create large liquidity risks for themselves, and also make themselves vulnerable to self-fulfilling market concerns about the adequacy of bank liquidity.³ The theoretical modelling of banking theorists, like the myopia theory of Minsky, is meant to explain prevalent banking fragility – a phenomenon that any blogger can now trace at least as far back as AD 33, when Tacitus (Book vi) tells us that the Roman Empire suffered a major banking panic, which was quelled by a large three-year interest-free loan to the banking system by Emperor Tiberius.⁴

Are these historical presumptions correct? This article examines the long-term record of banking crises in and outside the US, and places the recent crisis in that historical context.

¹ This article summarises a longer paper entitled, 'Banking crises and the rules of the game', NBER Working Paper no. 15403 (October 2009). The author gratefully acknowledges support from the Pew Trusts project on financial reform.

² C. P. Kindleberger, *Manias, Panics, and Crashes: A History of Financial Crises* (New York, 1978); H. P. Minsky, *John Maynard Keynes* (New York, 1975).

³ D. Diamond and P. Dybvig, 'Bank runs, deposit insurance, and liquidity', *Journal of Political Economy*, 91 (1983), pp. 401–19.

⁴ Cornelius Tacitus, *The Annals of Imperial Rome* (New York, 1996).

I

When and why do banking crises occur? To answer that question requires a definition of banking crises. Banking crises properly defined consist either of panics or severe waves of bank failures. Banking panics are moments of temporary confusion about the unobservable incidence across the banking system of observable aggregate shocks that are severe enough to give rise to collective action by bankers.⁵ Severe waves of bank failures are defined as those resulting in aggregate negative net worth of failed banks in excess of 1 per cent of GDP.⁶

Banking crises are a distinct subset within the broader set of phenomena known as financial crises. Financial crises broadly defined, which include asset price bubbles, exchange rate collapses, and a host of other phenomena, as well as banking crises, do appear to be a common and fairly constant feature of the economic cycle. Is the same true of banking crises?

Four basic facts about banking crises provide a starting point for understanding their origins, and show the importance of distinguishing banking crises from other financial crises.⁷

First, the record of banking crises (whether defined as panics or waves of severe failures) reveals that they are not random events. Banking crises, like other kinds of financial crises, tend to occur around the time of cyclical downturns. They are closely associated with prior rises in the liabilities of failed businesses and declines in asset prices. Not surprisingly, waves of bank failures are clearly traceable to large declines in the values of bank loans, which reflect declines in the fortunes of borrowers.

Second, unlike financial crises broadly defined, banking crises were relatively rare historically, despite the fact that the government policy interventions designed to stabilise the banking system (modern central bank lending, government-backed insurance of deposits, and additional forms of government assistance to distressed banks) are much more prevalent in the current financial system than they were in the past.

⁵ C. W. Calomiris and G. Gorton, 'The origins of banking panics: models, facts, and bank regulation', in R. G. Hubbard (ed.), *Financial Markets and Financial Crises* (Chicago, 1991), pp. 107–73.

⁶ G. Caprio and D. Klingebiel, 'Bank insolvencies: cross country experience', World Bank Working Paper no. 1620 (1996).

⁷ Banking crises are also distinct from other financial crises because of their especially large social costs. Asset price collapses that are not accompanied by banking crises – such as those in the US in 1987 and 2000 – did not have the severe macroeconomic consequences of the financial crises that are accompanied by banking crises. See B. S. Bernanke, 'Nonmonetary effects of the financial crisis in the propagation of the Great Depression', *American Economic Review*, 73 (1983), pp. 257–76; C. W. Calomiris and R. G. Hubbard, 'Price flexibility, credit availability, and economic fluctuations: evidence from the US, 1894–1909', *Quarterly Journal of Economics*, 104 (1989), pp. 429–52; and C. W. Calomiris and J. R. Mason, 'Fundamentals, panics and bank distress during the depression', *American Economic Review*, 93 (2003), pp. 1615–47. Indeed, banking distress manifested in significant deposit shrinkage and loan losses, even when not associated with a banking crisis, typically poses substantial costs for the economy because of the contraction of money and loan supply.

Third, an historical analysis of the two banking crises phenomena (panics and waves of failures) reveals that they do not always coincide, although they sometimes do. Bank panics can happen without a significant increase in failed banks (the panic of 1907, for example); while at other times, many bank failures occur without any systemic banking panic (as during the wave of US agricultural bank failures in the 1920s). This suggests that somewhat different phenomena underlie the two types of crises. Confusion about small losses can cause banking panics without a severe wave of failures; and large losses whose incidence within the banking system is easy to discern can cause many severe failures without a panic.

Fourth, perhaps most interestingly, banking crises of both types vary in their frequency across countries and across time, and the differences in the propensities for crises are dramatic. The US banking system experienced an unusually high propensity for both panics and waves of bank failures historically. Nationwide banking panics occurred in 1819, 1837–9, 1857, 1873, 1884, 1890, 1893, 1896, 1907 and 1933.⁸ In the four decades prior to World War I the US was unique in its propensity for panics. The US also experienced an unusually high frequency of severe waves of bank failures: in the 1830s, 1920s and 1930s. Great Britain also suffered an unusually high propensity for banking panics in the first half of the nineteenth century, but experienced a dramatic change in its propensity for panics in the middle of the nineteenth century; banking panics occurred in 1819, 1825, 1836, 1847, 1857 and 1866, and then (with the exception of a crisis induced by the onset of World War I) there were none for more than a century. Only four countries experienced severe waves of bank insolvency worldwide in the years 1875–1913; in the period 1978–2009, in contrast, roughly 140 such episodes have occurred, more than 20 of which are more severe than any of the pre-World War I episodes in terms of negative net worth of failed banks relative to GDP.⁹

Thus, banking crises cannot be seen as an inevitable result of human nature or the liquidity-transforming structure of bank balance sheets, and adverse macroeconomic circumstances alone are not sufficient to produce banking crises.

⁸ Recent research (e.g. Calomiris and Mason, 'Fundamentals, panics and bank distress') has shown that the large number of bank failures in the US during the Great Depression, a phenomenon that was largely confined to small banks, primarily reflected the combination of extremely large fundamental macroeconomic shocks and the vulnerable nature of the country's unit banking system. Panic was not a significant contributor to banking distress on a nationwide basis until near the trough of the Depression, at the end of 1932. For these reasons, the Great Depression bank failure experience has more in common with the bank failures of the 1920s than the panics of the pre-World War I era.

⁹ This record for the pre-World War I period is one of impressive banking stability, especially considering the high volatility of the macroeconomic environment during that period. The roughly 140 episodes in which banking systems experienced losses in excess of 1% of GDP include more than 20 episodes of negative net worth in excess of 10% of GDP, more than half of which resulted in losses in excess of 20% of GDP (these extreme cases include, for example, roughly 25–30% of GDP losses in Chile in 1982–3, Mexico in 1994–95, Korea in 1997, and Thailand in 1997, and a greater than 50% loss in Indonesia in 1997).

II

What accounts for the variation across time and across countries in the frequency and severity of panics and waves of bank failures? A survey of the history of banking crises traces unusual bank fragility to risk-inviting microeconomic rules of the banking game established by governments. Those rules of the game have been the key necessary condition for producing banking distress, whether in the form of a high propensity for banking panics or a high propensity for waves of bank failures.

Some risk-inviting rules took the form of visible subsidies for risk taking, as in the historical state-level deposit insurance systems in the US that failed disastrously in the 1920s, Argentina's government guarantees for risky mortgages in the 1880s, Italy's pre-1893 guarantees for the liabilities of the Banca di Roma, which financed the Roman real estate boom of that period, and Australia's government subsidisation of real estate development prior to 1893.¹⁰

In the US in both the 1830s and the 1920s, some states suffered more than others from waves of bank distress. In the 1830s, states that had an active role in directing the credit of their banks fared particularly badly.¹¹ In the 1830s and the 1920s, states that had enacted systems of bank liability insurance in which neither entry nor risk taking was effectively constrained experienced far worse banking system failure rates and insolvency severity of failed banks than did other states.¹² Indeed, the basis for the substantial opposition to federal deposit insurance in the 1930s – an opposition that included President Franklin D. Roosevelt, his Treasury Secretary, and the Federal Reserve – was the disastrous experimentation with insurance in several US states during the early twentieth century, which resulted in banking collapses in all the states that adopted insurance, and especially severe collapses in states that made deposit insurance compulsory.

¹⁰ During the pre-World War I era, Argentina in 1890 and Australia in 1893 were the exceptional cases; they each suffered banking system losses of roughly 10% of GDP in the wake of real estate market collapses in those countries. The negative net worth of failed banks in Norway in 1900 was roughly 3% and in Italy in 1893 roughly 1% of GDP, but with the possible exception of Brazil (for which data do not exist to measure losses), there seem to be no other cases in 1875–1913 in which banking losses in a country exceeded 1% of GDP. See C.W. Calomiris, 'Victorian perspectives on the banking distress of the late 20th century', working paper, 2007.

¹¹ L. Schweikart, *Banking in the American South from the Age of Jackson to Reconstruction* (Baton Rouge, 1987).

¹² The states of Indiana, Ohio and Iowa during the antebellum period were the exceptions to this rule, as their mutual guarantee systems were limited to a small number of banks which bore unlimited mutual liability for one another, and which also had broad enforcement powers to limit abuse of that protection. See C. W. Calomiris, 'Deposit insurance: lessons from the record', *Economic Perspectives*, Federal Reserve Bank of Chicago, 1989 (May/June), pp. 10–30; C. W. Calomiris, 'Is deposit insurance necessary? A historical perspective', *Journal of Economic History*, 50 (1990), pp. 283–95; and C. W. Calomiris, 'Do vulnerable economies need deposit insurance? Lessons from US agriculture in the 1920s', in P. L. Brock (ed.), *If Texas Were Chile: A Primer on Bank Regulation* (San Francisco, 1992), pp. 237–349, 450–8.

One of the most interesting examples of risk-inviting policy was the Bank of England's unlimited discounting of paper at low interest rates prior to 1858, which drove the boom and bust cycle in Britain that caused banking panics roughly every decade from 1819 to 1866. For decades the Bank of England (which operated as a for-profit institution) was effectively required by Parliament to provide an unlimited put option on banker's bills in the London market as a quid pro quo for maintaining its monopoly privileges, a requirement that had been openly sought in the political arena by bankers and borrowers seeking protection from loss.

Recent research that investigates the determinants of banking fragility across different countries in the current era reaches a similar conclusion: the expansion of government-sponsored deposit insurance and other bank safety net programmes throughout the world in the past three decades accounts very well for the increasing frequency and severity of banking crises in the current era. Empirical studies of this era of unprecedented frequency and severity of banking system losses has concluded uniformly that deposit insurance and other policies that protect banks from market discipline, intended as a cure for instability, have instead become the single greatest source of banking instability.¹³

Other risk-inviting rules historically have involved government-imposed structural constraints on banks, which include entry restrictions like unit banking laws that limit competition, prevent diversification of risk, and hamper the ability of the banking system to deal with shocks. The key difference between the US and other countries historically lay in the structure of its banking system. The US system was mainly based on unit banking – geographically isolated single-office banks. Unit banking meant that banks could not enjoy diversification economies by pooling loan risks from different regions. Unit banking, which resulted in thousands, and in some periods, tens of thousands of banks, also limited the ability of banks to pursue collective action by pooling resources during periods of adverse shocks. A system with tens of thousands of geographically distant banks simply could not organise appropriate collective action to stem financial crises.¹⁴ Other countries did not imitate the fragmented US approach to banking, and no other country experienced the US pattern of

¹³ See, for example, Caprio and Klingebiel, 'Bank insolvencies'; A. Demirguc-Kunt and E. Detragiache, 'Does deposit insurance increase banking system stability?', IMF Working Paper no. 3 (2000); J. Barth, G. Caprio, Jr and R. Levine, *Rethinking Bank Regulation: Till Angels Govern* (Cambridge, 2006); A. Demirguc-Kunt, E. Kane and L. Laeven (eds.), *Deposit Insurance Around the World* (Cambridge, MA, 2009).

¹⁴ Bank clearing houses or informal alliances among banks to make markets in each other's deposits during crises required that members in these coalitions adhere to guidelines, and that they be able to monitor one another to ensure compliance. Not only did geography get in the way of such coordination, the sheer number of banks made collective action difficult. The benefits of one bank choosing to monitor another are shared, but the monitoring and enforcement costs are borne privately; coalitions with 30 members seemed able to motivate individual banks to bear the private costs of monitoring on behalf of the coalition, but coalitions of hundreds or thousands of banks unsurprisingly were not able to structure effective monitoring and enforcement.

periodic banking panics prior to World War I, or the waves of agricultural bank failures that gripped the US in the 1920s.

For example, Canada's early decision to permit branch banking throughout the country ensured that banks were geographically diversified and thus resilient to large sectoral shocks (like those to agriculture in the 1920s and 1930s), able to compete through the establishment of branches in rural areas (because of low overhead costs of establishing additional branches), and able to coordinate the banking system's response in moments of confusion to avoid depositor runs (the number of banks was small, and assets were highly concentrated in several nationwide institutions). Coordination among banks facilitated systemic stability by allowing banks to manage incipient panic episodes to prevent widespread bank runs. In Canada, the Bank of Montreal occasionally would coordinate actions by the large Canadian banks to stop crises before the public was even aware of a possible threat.¹⁵

Another destabilising rule of the banking game is the absence of a properly structured central bank to act as a lender of last resort to reduce liquidity risk without spurring moral hazard. Early experiments with limited central banking in the US resulted in the failure to recharter central banks twice in the early nineteenth century, which reflected, in part, a difficulty in reconciling the financial limitations of a private bank of limited means with the public pressures on that bank to 'pay for' its privileges by performing unprofitable services in the public interest. Although some observers accused the central bank, the Second Bank of the United States (SBUS), of contributing to financial instability through contractionary policies prior to and during both the panic of 1819 and the financial crisis of 1825–6, those accusations say more about unrealistic public expectations of the power of the SBUS to prevent systemic problems than they do about the desirability of rechartering the SBUS. Although neither the First nor Second Banks of the United States were equipped to act fully as lenders of last resort during crises, the SBUS succeeded in reducing systemic financial risk on average and over the seasonal cycle, foreshadowing the stabilising effect of the Fed after 1913. After the demise of the SBUS, the US functioned without a central bank until the founding of the Fed in 1913.

The key destabilising elements of the US system – a fragmented industrial structure, the absence of an effective lender of last resort, and the occasional presence of a destabilising deposit insurance regime – compounded one another. Canada, which avoided chartering a central bank until 1935, managed to avoid banking crises due to the stabilising role of its branch banking system, despite the absence of a central bank. In the US, the fragility of the banking structure made the absence of a central bank more harmful than it otherwise would have been; likewise, the absence of an effective central bank magnified the destabilising effects of unit banking.

History also teaches us that regulatory policy often responds to banking crises, but not always wisely. The British response to the string of panics culminating in the panic of 1857 is an example of effective learning, which put an end to the subsidisation of

¹⁵ C. W. Calomiris, *US Bank Deregulation in Historical Perspective* (Cambridge, 2000), ch. 1.

risk through reforms to Bank of England policies in the bills market. In March 1858, with the support of the Parliament, the Bank of England explicitly repealed its implicit commitment to provide a put option in the bills market. In 1866, that policy change was tested during the Overend, Gurney crisis. The Bank refused to bail out Overend, which established the credibility of its announced policy change and ushered in an era of unprecedented banking stability.

Not all policy reactions to banking crises have been wise. One counterproductive response was the decision in the US in the 1830s not to reinstate the charter of the SBUS, which had been stabilising the banking system prior to its demise. That decision reflected misunderstandings about the Second Bank's contributions to financial instability in 1819 and 1825.

The decision in the US in 1933 to end bank consolidation and adopt federal deposit insurance instead was a mistake of a different kind; that policy was understood to be contrary to the stabilisation of banking and the pursuit of the public interest and was opposed by President Roosevelt, the Federal Reserve, the Treasury and the leading bank reformer in the Senate, Carter Glass. Nonetheless, Congressman Henry Steagall succeeded in pushing through deposit insurance as part of a political compromise, and thereby captured the regulatory process on behalf of his unit banking constituents in Alabama.

III

As I have discussed in detail in a recent analysis,¹⁶ the subprime crisis, like the episodes of historical banking crises described above, was not just a bad accident. On an ex ante basis, subprime default risk was excessive and substantially underestimated during 2003–7. Reasonable, forward-looking estimates of risk were ignored, and compensation for asset managers created incentives to undertake underestimated risks. Those risk-taking errors reflected a policy environment that strongly encouraged financial managers to underestimate risk in the subprime mortgage market. Among the causes of the crisis discussed were policies specifically designed to encourage risk taking in the mortgage market; these are especially deserving of emphasis.

Numerous housing policies promoted subprime risk taking by financial institutions by subsidising the inexpensive use of leveraged finance in housing. Those policies included:

- Political pressures from Congress on the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac, to promote 'affordable housing' by investing in high-risk subprime mortgages
- Lending subsidies for housing finance via the Federal Home Loan Bank System to its member institutions

¹⁶ C. W. Calomiris, 'The subprime turmoil: what's old, what's new, and what's next', *Journal of Structured Finance*, 15 (2009), pp. 6–52.

- Federal Housing Administration (FHA) subsidisation of extremely high mortgage leverage and risk
- Government and GSE mortgage foreclosure mitigation protocols that were developed in the late 1990s and early 2000s to reduce the costs to borrowers of failing to meet debt service requirements on mortgages, which further promoted risky mortgages
- 2006 legislation enacted to encourage ratings agencies to relax standards for subprime securitisations

All these policies encouraged the underestimation of subprime risk, but the behaviour of members of Congress toward Fannie Mae and Freddie Mac, in the name of affordable housing, was arguably the single most destructive influence leading up to the crisis.¹⁷

The mid-nineteenth-century British discussions of financial reform that led to the successful removal of destabilising subsidies for risk taking share important features with the current debates over prudential regulatory and housing finance policy reforms in the US. Many aspects of the current debate would seem familiar to nineteenth-century British observers. Public resentment over the abuse of special privileges by mortgage monopolists, Fannie Mae and Freddie Mac, who fuelled the subprime bubble, and whose internal emails show that they did so largely to preserve the special privileges conferred upon them by the government,¹⁸ is reminiscent of the discussion of the moral hazard produced by the Bank of England. The liquidity risk that arose from the heavy dependence on repo financing by US investment banks in recent years parallels the growth of the discount brokers in London who built up huge liquidity risk in the banking system, which was the primary means of inflating bubbles during the first half of the nineteenth century in Britain. Just as the debate over

¹⁷ For Fannie and Freddie to maintain lucrative implicit (now explicit) government guarantees on their debts they had to commit growing resources to risky subprime loans: see C.W. Calomiris, 'Statement before the Committee on Oversight and Government Reform, United States House of Representatives', 9 December 2008; and P.-J. Wallison and C. W. Calomiris, 'The last trillion-dollar commitment: the destruction of Fannie Mae and Freddie Mac', *Journal of Structured Finance*, 15 (2009), pp. 71–80. Due to political pressures, which were discussed openly in emails between management and risk managers in 2004, Fannie and Freddie purposely put aside their own risk managers' objections to making the market in no-docs subprime mortgages in 2004. The risk managers correctly predicted, based on their experience with no-docs in the 1980s, that their imprudent plunge into no-docs would produce adverse selection in mortgage origination, cause a boom in lending to low-quality borrowers, and harm their own stockholders and mortgage borrowers alike. In 2004, in the wake of Fannie and Freddie's decision to aggressively enter no-docs subprime lending, total subprime originations tripled. In late 2006 and early 2007, after many lenders had withdrawn from the subprime market in response to stalling home prices, Fannie and Freddie continued to accumulate subprime risk at peak levels. Fannie and Freddie ended up holding \$1.6 trillion in exposures to those toxic mortgages, half the total of non-FHA outstanding amounts of toxic mortgages: see E. J. Pinto, 'Statement before the Committee on Oversight and Government Reform, United States House of Representatives', 9 December 2008.

¹⁸ Calomiris, 'Statement before the Committee'.

financial regulation today grapples with the question of whether to impose prudential regulations on non-banks, Britain struggled with the problem of an ineffectual, narrow approach to defining prudential regulation, which was limited to the Bank Act of 1844's reserve requirement against Bank of England note issues, and did nothing to limit deposit growth or bill discounting by brokers. The concern about the 'Greenspan put' and the moral-hazard consequences of the 'too-big-to-fail' doctrine in the wake of the rescue of Bear Stearns, AIG, Citibank and other large financial institutions is reminiscent of the Bank of England's struggle to cancel its put option in the London market for bills and rein in other institutions' entitlements to unlimited accommodation during crises, a practice that was ended in 1858, and proven in 1866.

This is not the place to explore in detail how to apply the lessons of the successful reform of the British banking system in the nineteenth century to the current environment.¹⁹ The important point to emphasise here, as a consistent theme of the historical record, is that the ability to improve the financial system depends on the political environment.

The favourable outcome in Britain in the nineteenth century resulted from a political consensus in favour of reform that created strong political incentives to get reform right, in order to stop the boom and bust cycles that had plagued the economy for decades. The risk-inviting incentive problems that gave rise to the recent subprime crisis have much in common with prior experiences of unstable banking systems, and the principles for reform are similar. The key question is whether the political equilibrium will encourage favourable reforms in the wake of banking crisis, as it did in Britain in the nineteenth century, or unfavourable reforms as the result of populist misapprehension, as in the case of the disappearance of the SBUS, or the capture of financial reform by special interests, as was the case in the US in 1933.

IV

This brief survey of the history of banking crises traces unusual bank fragility to risk-inviting microeconomic rules of the banking game established by governments, the most important of which have been rules that subsidise risk. Other destabilising rules include limits on bank entry and the failure to establish a proper lender of last resort. The subprime crisis exemplifies the historical pattern all too well. Government subsidisation of risky mortgages in the US accelerated markedly in the years prior to the crisis. That along with prudential regulatory failures to prevent excessive risk taking allowed the mortgage risk binge of 2003–7 to produce a worldwide financial collapse. As the US gears up to respond to the

¹⁹ I have laid out my views on that reform agenda in Calomiris, 'The subprime turmoil'; 'Financial innovation, regulation, and reform', *Cato Journal*, 29 (2009), pp. 65–92; and 'Prudential bank regulation: what's broken and how to fix it', in T. L. Anderson and R. Soussa (eds.), *Reacting to the Spending Spree: Policy Changes We Can Afford* (Stanford, CA, 2009).

subprime crisis with regulatory reforms, history suggests important lessons. Regulatory policy often has responded to banking crises, but not always wisely. A favourable outcome in Britain in the nineteenth century, for example, resulted from a political consensus in favour of reform that created strong political incentives to get reform right, in order to stop the boom and bust cycles that had plagued the economy for decades. But counterproductive responses to crises are also a possibility, either due to misunderstanding about the sources of crises or political capture of regulatory reform. A consistent theme of the historical record is that the ability to improve regulation in reaction to banking crises depends crucially on the political environment.