

NBER WORKING PAPER SERIES

BANKING CRISES AND THE RULES OF THE GAME

Charles Calomiris

Working Paper 15403

<http://www.nber.org/papers/w15403>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

October 2009

This paper will appear as a chapter in a Festschrift in honor of Forrest Capie. I thank participants in the Forrest Capie Festschrift conference at City University London, as well as attendees at the Plenary Session of the September 2009 Economic History Association Meetings, and seminar participants at the Graduate Center of the City University of New York, for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2009 by Charles Calomiris. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Banking Crises and the Rules of the Game
Charles Calomiris
NBER Working Paper No. 15403
October 2009
JEL No. E5,E58,G2,N2

ABSTRACT

When and why do banking crises occur? Banking crises properly defined consist either of panics or waves of costly bank failures. These phenomena were rare historically compared to the present. A historical analysis of the two phenomena (panics and waves of failures) reveals that they do not always coincide, are not random events, cannot be seen as the inevitable result of human nature or the liquidity transforming structure of bank balance sheets, and do not typically accompany business cycles or monetary policy errors. Rather, risk-inviting microeconomic rules of the banking game that are established by government have always been the key additional necessary condition to producing a propensity for 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 U.S., Argentina's government guarantees for mortgages in the 1880s, Australia's government subsidization of real estate development prior to 1893, the Bank of England's discounting of paper at low interest rates prior to 1858, and the expansion of government-sponsored deposit insurance and other bank safety net programs throughout the world in the past three decades, including the generous government subsidization of subprime mortgage risk taking in the U.S. leading up to the recent crisis.

Other risk-inviting rules historically have involved government-imposed structural constraints on banks, which include entry restrictions like unit banking laws that constrain competition, prevent diversification of risk, and limit the ability to deal with shocks. Another destabilizing 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.

Regulatory policy often responds to banking crises, but not always wisely. The British response to the Panic of 1857 is an example of effective learning, which put an end to the subsidization of risk through reforms to Bank of England policies in the bills market. Counterproductive responses to crises include the decision in the U.S. not to retain its early central banks, which reflected misunderstandings about their contributions to financial instability in 1819 and 1825, and the adoption of deposit insurance in 1933, which reflected the political capture of regulatory reform.

Charles Calomiris
Graduate School of Business
Columbia University
3022 Broadway Street, Uris Hall
New York, NY 10027
and NBER
cc374@columbia.edu

I. Introduction

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 (1973) and Hyman Minsky (1975) 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.

Some banking theorists, starting with Diamond and Dybvig (1983), 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. The theoretical modeling 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 33 AD, 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.¹

¹ “The destruction of private wealth precipitated the fall of rank and reputation, until at last the emperor interposed his aid by distributing throughout the banks a hundred million sesterces, and allowing freedom to borrow without interest for three years, provided the borrower gave security to the State in land to double the amount. Credit was thus restored, and gradually private lenders were found.” This account by Tacitus traces the crisis to government enforcement of a long-neglected usury law. Unfortunately, confusion has arisen about the origins of the Roman Panic of 33 AD due to an apparent attempt at humor by an early 20th century American, William Stearns Davis. Davis, who was known both for historical writings and historical fiction, seems to have invented a colorful account that was informed by his knowledge of events during the US Panic of 1907. He presented it as a factual account in his 1910 treatise, *The Influence of Wealth in Imperial Rome*. Calomiris (1989)

There is, however, at least one obvious thing wrong with all these arguments that purport to show how myopia, business cycles, and inherent bank liquidity transformation can explain the historical constancy of banking crises: in fact, the propensity for banking crises has not been at all constant over time or across countries. Banking crises have not regularly and consistently accompanied business cycles. In fact, banking crises have been much more frequent in some eras than in others and much more frequent in some countries than in others. The differences across countries and across time are dramatic, as this paper will demonstrate.

This is, in fact, a central lesson of the history of banking crises, which economic historians should be emphasizing in their discussions with macroeconomists, theorists, and policy makers in the wake of the current global banking crisis: banking crises are not an historical constant, and therefore, the propensity for banking crises cannot possibly be said to be the result of factors that have been constant over time and across countries for hundreds of years, including business cycles, human nature, or the liquidity transformation inherent in bank balance sheets.

A second, related lesson of the history of banking crises, and the main point of this paper, is that the structure of the rules governing the banking system within a country – defined by the rules that govern the location, powers, and the operations of each of the banks, including government subsidies or special rights granted to favored participants in the banking system and the incentive consequences of those subsidies and rights – has been at the center

took this account seriously, as have many other scholars and journalists. Davis provided no references or footnotes; a review of Roman sources indicates no factual basis in any known source for Davis's account, and various humorous aspects of the portrayal (for example, problems in the market for ostrich feathers) add to the likelihood that the account is fictional.

of the explanation of the propensity for banking crises for the past two centuries. In times and places where politically determined microeconomic rules of the banking game have encouraged risky practices or prevented effective private measures to limit banking crisis risk, the risk of banking crises is high; conversely, the absence of such adverse political rules of the game have resulted in stable banking systems.

Some of this evidence is visible in the history of particular countries, not just in cross-country comparisons. When the political equilibrium governing the rules of the banking game changed for the better (worse) in a particular country, previously unstable (stable) banking systems became stable (unstable). The primary purpose of this paper is to review these experiences and consider their lessons for current policy reactions to the global banking crisis of 2007-2009. Specifically, Sections II, III, and IV: (1) review the experience of the United States in the 19th and early 20th centuries in this regard, (2) compare and contrast the banking rules of the game of the highly stable pre-World War I period with those of the last thirty years' highly unstable banking experience, and (3) explain how Great Britain changed from suffering a highly unstable banking system during the first six decades of the 19th century to becoming a paragon of stability in the pre-World War I era.

This emphasis on the microeconomic rules of the banking game, and the political economy that gives rise to those rules, should not be interpreted as an argument for the irrelevance of macroeconomic considerations (monetary policy, the phases of business cycles, etc.) in understanding banking crises. Monetary policy and other macroeconomic considerations have indeed been an important source of financial crises, which include asset price bubbles, exchange rate collapses, and a host of other phenomena, as well as banking

crises. It is also true that financial crises, broadly defined to include asset pricing booms and busts have been a common feature of business cycles throughout time. But although monetary policy errors have often set the stage for banking crises (see Bordo 2007, Bordo and Wheelock 2007, 2009, Calomiris 2009a), monetary policy errors and business cycle swings more generally have not proven to be sufficient conditions for *banking* crises. Destabilizing monetary policy, or other macroeconomic considerations, only tend to produce banking crises alongside cyclical contractions when the microeconomic rules of the banking game are poorly designed.

Indeed, a third lesson from the history of banking instability is that the ability to derive useful lessons about banking crises depends on defining banking crises properly. Banking crises must be distinguished from the broader category of “financial crises,” which include a variety of other phenomena (i.e., sovereign debt defaults, exchange rate depreciations, land price declines, and stock market declines), which may or may not be associated with banking distress. And banking problems, including significant declines in deposits for the system as a whole, or the failure of one or two banks, do not equate to a banking crisis.² What makes a banking problem into a banking crisis?

When defining banking crises it is important to distinguish between two different aspects of banking crises – waves of bank insolvency (episodes in which bank losses result in many failed banks), and banking panics (moments in which the banking system as a whole

² 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 U.S. in 1987 and 2000 – did not have the severe macroeconomic consequences of the financial crises that are accompanied by banking crises (see Bernanke 1983, Calomiris and Hubbard 1989, Calomiris and Mason 2003b). 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. For a cross-country analysis, see Bordo and Eichengreen (2003), who study the effects of historical banking distress (broadly defined) on business cycle severity.

suffers from sudden, large withdrawals of deposits).³ Sometimes these two aspects have coincided (as during some episodes in the U.S. in the 1830s and the 1930s, and many recent episodes), but often they have not coincided. The 1920s in the U.S. witnessed a severe wave of bank failures, but not panics. The U.S. experience between the Civil War and World War I witnessed several banking panics but no significant waves of bank failures. It is useful to recognize panic and insolvency as separate aspects of banking crises because these different aspects reflect separate causal influences. In my review of the history of banking crises, therefore, I take account of both panics and episodes of high insolvency. Nevertheless, the key insight of this paper – that politically driven rules of the banking game drive the presence or absence of crisis risk – applies to both the panic and insolvency aspects of banking crises; that is, poorly designed structures and incentives for the banking system explain both the propensity for banking panics and the propensity for severe waves of bank failures.

This review offers important insights for policy makers. The crisis of 2007-2009 has sharpened or redefined many public policy questions of central importance to prudential financial regulation (a means of preventing crises) and the proper role of government assistance policy (a means of mitigating the costs of crises). What do we learn from the policy responses to banking crises in the past about the appropriate reforms we should undertake in reaction to recent events? How should the past guide our current policy adaptations?

³ It is important to define banking panics carefully. For my purposes, panics are moments of confusion about the incidence of losses in banks that are sufficiently severe as to create systemic withdrawal pressure on a large number of banks that is sufficient to elicit collective action by the banks and/or the government (e.g., joint issuance of liabilities, like clearing house certificates, or the undertaking of joint action, such as suspension of convertibility, or other explicit attempts to coordinate behavior to alleviate the effects of panic). This definition creates an objective standard that distinguishes true panics from less severe moments of stress that are not truly systemic in scope.

A fourth lesson from the history of banking crises concerns the circumstances that tend to produce effective learning in the policy responses to crises. In previous banking crises, reforms have often followed in the wake of crises, but the record of reform is uneven. One successful historical reform described in this paper – the mid-19th century British reform of the Bank of England, reviewed in Section IV, which successfully eliminated lending rules that gave rise to the frequent panics that plagued Britain in the early 19th century – shows that meaningful structural reforms that reduce incentives to take on excessive risk can stabilize banking systems. On the other hand, policy responses sometimes make matters worse: the failure to re-charter a central bank in the U.S. in the 1830s reflected, in part, mistaken views about the Second Bank of the United States during the crises of 1819 and 1825; and the bank regulatory changes in the U.S. in 1933 reflected political deal making rather than a proper response to the root causes of banking instability.

After considering these historical perspectives on the origins of banking distress and the policy solutions to address it – which consist of a detailed analysis of U.S. banking crises in Section II, a broader review of the global history of bank insolvency in Section III, and the history of British panics in Section IV – Section V reviews the causes of the financial crisis of 2007-2009 and consider prospects for reform today. Section VI concludes.

II. The Microeconomic Foundations of U.S. Banking Crises: 1790-1933

The Peculiar Fragility of U.S. Banks in the Pre-Depression Era

As many scholars have recognized for many years, U.S. banks were unusually vulnerable to systemic banking crises compared to banks in other countries (for reviews, see Bordo 1985,

and Calomiris 2000). The U.S. was uniquely vulnerable to panics in the years between the Civil War and World War I. Sprague (1910) and Calomiris and Gorton (1991) identify six episodes of particularly severe banking panics in the United States between the Civil War and World War I. Prior to the Civil War, there were other nationwide banking crises in 1819, 1837, and 1857, in which both systemwide panic and many bank failures occurred. In the 1920s, the U.S. experienced waves of bank failures in agricultural states, which have always been identified with fundamental shocks to banks, and which did not give rise to national or regional panics.

Other countries, including the U.S.'s northern neighbor, Canada, did not suffer banking crises during these episodes of systemic U.S. banking distress. The key difference between the U.S. and other countries historically lay in the structure of the U.S. banking system. The U.S. 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 sometimes 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 organize appropriate collective action to stem financial crises.⁴ Other countries did not choose the fragmented U.S. approach to banking, and no other country experienced the

⁴ 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.

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

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 (Calomiris 2000, Chapter 1).

The United States was unable to mimic this behavior on a national or regional scale, although during the antebellum period, a few southern branch banking states, and three Midwestern states that formed mutual guarantee systems with small numbers of members, were able to implement successful, stabilizing coalitions of banks at the state level for purposes of mutual protection during banking crises (Calomiris 1989, 1990, 2000, Calomiris and Schweikart 1991). But these were short-lived and isolated exceptions; U.S. law prohibited nationwide branching, and most states prohibited or limited within-state branching. U.S. banks were numerous (e.g., numbering more than 29,000 in 1920), undiversified, insulated from competition, and geographically isolated from one another, and thus were unable to diversify adequately or to coordinate their response to panics (U.S. banks did establish clearing houses in

cities, which facilitated local responses to panics beginning in the 1850s, as emphasized by Timberlake 1984 and Gorton 1985).

The fragmented structure of U.S. banking explains why the United States uniquely suffered banking panics in the years between the Civil War and World War I despite the fact that the vast majority of banks were healthy throughout this period, and were consistently able to avoid ultimate failure.⁵ Empirical studies show that the major U.S. banking panics of 1873, 1884, 1890, 1893, 1896, and 1907 were moments of heightened asymmetric information about bank risk, but not times when bank failure risk was large for the country as a whole (Calomiris and Gorton 1991, Bruner and Carr 2007).

Banking necessarily entails the delegation of decision making to bankers, who specialize in screening and monitoring borrowers and making non-transparent investments. Bankers consequently have private information about the attendant risks. During normal times, the risk premium banks pay in capital markets and money markets contains a small “opacity” premium – part of the risk depositors and bank stockholders face and charge for comes from not being able to observe the value of bank assets moment to moment – that is, not being able to mark bank portfolios to market. During the U.S. panics, the normally small opacity premium became very large, as people became aware that risks had increased and as they also were aware of what they *didn't* know, namely the incidence among banks of the probable losses that accompanied the observable increased risk.

⁵ The absence of a lender of last resort, as discussed below, was also an important contributor to bank instability, but the structure of unit banking appears to be the more important influence; Canada, which operated on a branching basis, avoided panics during this era, although it did not charter a central bank until 1935.

Calomiris and Gorton (1991) show that banking panics were uniquely predictable events that happened at business cycle peaks. In the pre- World War I period (1875-1913), every quarter in which the liabilities of failed businesses rose by more than 50% (seasonally adjusted) and the stock market fell by more than 8%, a panic happened in the following quarter. This happened five times, and the Panic of 1907 was the last of those times. Significant national panics (i.e., events that gave rise to a collective response by the New York Clearing House) never happened otherwise during this period.

Bank failure rates in the years between the Civil War and World War I, even during these panic episodes, were low, and the losses to depositors associated with them were also small. In 1893, the panic with the highest failure rate and highest depositor loss rate, depositor losses were less than 0.1% of GDP (Calomiris 2007). *Expected* depositor losses during the panics also appear to have been small. Sprague (1910, pp. 57-8, 423-24) reports that the discount applied to bankers' cashier checks of New York City banks at the height of the Panic of 1873 did not exceed 3.5%, and with the exception of an initial 10-day period, that discount remained below 1%. A similar pattern was visible in the Panic of 1893. A 1% premium would be consistent with depositors in a New York City bank estimating a 10% chance of a bank's failing with a 10% depositor loss if it failed. Clearly, banking panics during this era were traceable to real shocks, but those shocks had small consequences for bank failures in the aggregate, and even at the height of the crisis those consequences were expected to be small. Historical U.S. panics teach us that even a small expected loss can lead depositors to demand their funds, so that they can sit on the sidelines until the incidence of loss within the banking system has been revealed (usually a process that took a matter of weeks).

Bank failure rates in the 1830s, 1850s and 1920s were higher than those of the other pre-Depression systemic U.S. banking crisis episodes. The 1830s, in particular, saw a major macroeconomic contraction that caused many banks to fail, which historians trace to large fundamental problems that had their sources in government-induced shocks to the money supply (Rousseau 2002), unprofitable bank-financed infrastructure investments that went sour (Schweikart 1987), and international balance of payments shocks (Temin 1969).

The 1920s agricultural bank failures were also closely linked to fundamental problems, in this case, the collapses of agricultural prices at the end of World War I, which were manifested in local bank failures because of the lack of regional or national loan portfolio diversification (Calomiris 1992, Alston, Grove and Wheelock 1994).

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 (Schweikart 1987). Prior to both the bank failure waves of the 1830s and the 1920s, some states had enacted systems of deposit insurance in which neither entry nor risk taking was effectively constrained. These states experienced far worse banking system failure rates and insolvency severity of failed banks than did other states (Calomiris 1989, 1990, 1992).⁶ 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 U.S. states

⁶ 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.

during the early 20th century, which resulted in banking collapses in all the states that adopted insurance, and especially severe collapses in states that made deposit insurance compulsory.

In the 1920s, state-chartered banks that participated in deposit insurance fared much worse than either national banks in those states or state-chartered banks in neighboring states. The disastrous experience of those banks reflected a combination of moral hazard and adverse selection. Moral hazard was reflected in the higher loan-to-asset ratios and lower capital-to-asset ratios of state-chartered banks in insured states. Furthermore, states that passed deposit insurance experienced substantial entry into banking by small operators in rural areas, who apparently overestimated the potential for agricultural prices (temporary boosted by World War I) to remain high.

In contrast, in the 1920s, states that had enacted laws permitting branch banking tended to outperform unit banking states, both with respect to failure rates and failure severity (Calomiris 1990, 1992). The evidence of the stabilizing effects of even limited branch banking in the U.S. (note that branching was not permitted across states, and in many cases was constrained even when it was allowed within states) helped to produce significant relaxations of branch banking restrictions in many states and a merger wave of banks during the 1920s. From 1921 through 1931, more than five thousand banks were absorbed by acquirers. In 1910, for the U.S. as a whole, there were 292 branching banks operating 548 branches, with total loans and investments of \$1.3 billion, and in 1920, there were 530 branching banks operating 1,281 branches, with total loans and investments of \$6.9 billion; by 1931, there were 723 branching banks operating 3,467 branches, with total loans and investments of \$20.7 billion (Calomiris 2000, 57).

U.S. Bank Distress during the Great Depression

The legacy of branch banking restrictions continued to destabilize banks during the Depression. Mitchener (2005) finds that states that prohibited branching had higher rates of bank failure, *ceteris paribus*. Despite these trends and evidence, the stabilizing trend toward bank consolidation and greater structural stability in the U.S. was derailed by the global macroeconomic policy disaster of the Great Depression, and its adverse political consequences for continuing bank consolidation. Most importantly, Congressman Henry Steagall of Alabama lobbied successfully on behalf of his state's unit bankers for federal deposit insurance, which was embraced by unit bankers as a political tool to prevent competition and continuing pressure for consolidation (Calomiris and White 1994). Initially deposit insurance was passed as a temporary emergency measure limited to only cover small deposits (effectively a subsidy for small banks, for whom such deposits comprised a large fraction of their liabilities). Despite the opposition of Senator Carter Glass, the Federal Reserve System, the Treasury Department, and President Roosevelt – all of whom were aware of the disastrous consequences of deposit insurance in the states that had experimented with it in the early 20th century – Steagall managed to succeed in passing deposit insurance, which was soon transformed from a temporary to a permanent measure, and which now covers virtually all U.S. deposits.

Beginning in the 1880s there had been 150 attempts to introduce federal deposit insurance legislation in Congress (Calomiris and White 1994). Opponents understood and espoused the theoretical arguments against deposit insurance that are familiar today – that deposit insurance removes depositors' incentives to monitor and discipline banks, that it frees bankers to take imprudent risks (especially when they have little or no remaining equity at

stake, and see an advantage in “resurrection risk taking”); and that the absence of discipline promotes banker incompetence, which leads to unwitting risk taking. Deposit insurance won the day as legislation in 1933 for political, not ideological reasons, and ironically (given Roosevelt’s opposition) remains the main surviving legacy of the banking legislation of the New Deal – a stark reminder of the power of crises to change the course of banking regulation.⁷

Deposit insurance, which was very limited in coverage, and became effective only in 1934, after the banking crises of 1930-1933 had passed, had little role in stabilizing banks during the Depression of 1929 to 1933. Bank failures and losses were high in the early 1930s by historical standards. Recent research on the Depression has investigated the extent to which those failures reflected extremely adverse macroeconomic shocks and their consequences for bank borrowers, as opposed to excessive, panicked responses to those shocks by depositors that may have forced many solvent banks into financial distress. Recent research shows that much if not all of the bank distress of the 1930s resulted from fundamental shocks to bank assets, much like the shocks that had buffeted agricultural banks in the 1920s.

The list of fundamental shocks that weakened banks during the Great Depression is a long and varied one. It includes declines in the value of bank loan portfolios produced by waves of rising default risk in the wake of regional, sectoral, or national macroeconomic shocks to bank borrowers, as well as monetary policy-induced declines in the prices of the bonds held by banks.

⁷ The other two principal measures – Regulation Q and the separation of commercial and investment banking – were essentially done away with in the last two decades of the 20th century, although some remnants of Regulation Q remain.

Friedman and Schwartz (1963) argued that many bank failures resulted from unwarranted “panic” and that failing banks were in large measure illiquid rather than insolvent. Friedman and Schwartz’s emphasis on contagion posited that bank failures mainly reflected a problem of illiquidity rather than insolvency. Illiquid but solvent financial institutions, in their view, failed purely as the result of withdrawal demands by depositors, particularly during sudden moments of panic. In contrast, an insolvent institution fails to repay depositors as the result of fundamental losses in asset value, rather than the suddenness of depositor withdrawals.⁸

Using a narrative approach similar to that of Friedman and Schwartz, but relying on data disaggregated at the level of Federal Reserve districts, Wicker (1996) argues that it is

⁸ Friedman and Schwartz attach great importance to the banking crisis of late 1930, which they attribute to a “contagion of fear” that resulted from the failure of a large New York bank, the Bank of United States, which they regard as itself a victim of panic. They also identify two other banking crises in 1931 – from March to August 1931, and from Britain’s departure from the gold standard (September 21, 1931) through the end of the year. The fourth and final banking crisis they identify occurred at the end of 1932 and the beginning of 1933, culminating in the nationwide suspension of banks in March. The 1933 crisis and suspension was the beginning of the end of the Depression, but the 1930 and 1931 crises (because they did *not* result in suspension) were, in Friedman and Schwartz’s judgment, important sources of shock to the real economy that turned a recession in 1929 into the Great Depression of 1929-1933.

The Friedman and Schwartz argument is based upon the suddenness of banking distress and the absence of collapses in relevant macroeconomic time series prior to those banking crises (see Charts 27-30 in Friedman and Schwartz 1963, p. 309). But there are reasons to question Friedman and Schwartz’s view of the exogenous origins of the banking crises of the Depression. As Temin (1976) and others have noted, the bank failures during the Depression marked a continuation of the severe banking distress that had gripped agricultural regions throughout the 1920s. Of the nearly 15,000 bank disappearances between 1920 and 1933, roughly half predate 1930. And massive numbers of bank failures occurred during the Depression era outside the crisis windows identified by Friedman and Schwartz (notably, in 1932). Wicker (1996, p. 1) estimates that “[b]etween 1930 and 1932 of the more than 5,000 banks that closed only 38 percent suspended during the first three banking crisis episodes.” Recent studies of the condition of the Bank of United States indicate that it too may have been insolvent, not just illiquid, in December 1930 (Joseph Lucia 1985, Wicker 1996). So there is some prima facie evidence that the banking distress of the Depression era was more than a problem of panic-inspired depositor flight.

Friedman and Schwartz omitted important aggregate measures of the state of the economy relevant for bank solvency, for example, measures of commercial distress and construction activity may be useful indicators of fundamental shocks. Second, aggregation of fundamentals masks important sectoral, local, and regional shocks that buffeted banks with particular credit or market risks. The empirical relevance of these factors has been demonstrated in the work of Wicker (1980, 1996) and Calomiris and Mason (1997, 2003a).

incorrect to identify the banking crisis of 1930 and the first banking crisis of 1931 as national panics comparable to those of the pre-Fed era. According to Wicker, the proper way to understand the process of banking failure during the Depression is to disaggregate, both by region and by bank, because heterogeneity was very important in determining the incidence of bank failures.⁹

Microeconomic studies of banking distress have provided some useful evidence on the reactions of individual banks to economic distress. White (1984) shows that the failures of banks in 1930 are best explained as a continuation of the agricultural distress of the 1920s, and are traceable to fundamental disturbances in agricultural markets.

⁹ Once one disaggregates, Wicker argues, it becomes apparent that at least the first two of the three banking crises of 1930-1931 identified by Friedman and Schwartz were largely regional affairs. Wicker (1980, 1996) argues that the failures of November 1930 reflected regional shocks and the specific risk exposures of a small subset of banks, linked to Nashville-based Caldwell & Co., the largest investment bank in the South at the time of its failure. Temin (1989, p. 50) reaches a similar conclusion. He argues that the “panic” of 1930 was not really a panic, and that the failure of Caldwell & Co. and the Bank of United States reflected fundamental weakness in those institutions.

Wicker’s analysis of the third banking crisis (beginning September 1931) also shows that bank suspensions were concentrated in a very few locales, although he regards the nationwide increase in the tendency to convert deposits into cash as evidence of a possible nationwide banking crisis in September and October 1931. Wicker agrees with Friedman and Schwartz that the final banking crisis (of 1933), which resulted in universal suspension of bank operations, was nationwide in scope. The banking crisis that culminated in the bank holidays of February-March 1933 resulted in the suspension of at least some bank operations (bank “holidays”) for nearly all banks in the country by March 6.

From the regionally disaggregated perspective of Wicker’s findings, the inability to explain the timing of bank failures using aggregate time series data (which underlay the Friedman Schwartz view that banking failures were an unwarranted and autonomous source of shock) would not be surprising even if bank failures were entirely due to fundamental insolvency. Failures of banks were local phenomena in 1930 and 1931, and so may have had little to do with national shocks to income, the price level, interest rates, and asset prices. The unique industrial organization of the American banking industry plays a central role in both the Wicker view of the process of bank failure during the Depression, and in the ability to detect that process empirically. Because banks in the United States were smaller, regionally isolated institutions, large region-specific shocks might produce a sudden wave of bank failures in specific regions even though no evidence of a shock was visible in aggregate macroeconomic time series (see the cross-country evidence in Bernanke and James 1991, and Grossman 1994). The regional isolation of banks in the United States, due to prohibitions on nationwide branching or even statewide branching in most states, also makes it possible to identify regional shocks empirically through their observed effects on banks located exclusively in particular regions.

Calomiris and Mason (1997) study the Chicago banking panic of June 1932 (a locally isolated phenomenon). They find that the panic resulted in a temporary contraction of deposits that affected both solvent and insolvent banks. Fundamentals, however, determined which banks survived. Apparently, no solvent banks failed during that panic. Banks that failed during the panic were observably weaker ex ante, judging from their balance sheet and income statements, and from the default risk premia they paid on their debts. Furthermore, the rate of deposit contraction was not identical across banks; deposits declined more in failing banks than in surviving banks.

Calomiris and Wilson (2004) study the behavior of New York City banks during the interwar period, and in particular, analyze the contraction of their lending during the 1930s. They find that banking distress was an informed market response to observable weaknesses in particular banks, traceable to ex ante bank characteristics. It resulted in bank balance sheet contraction, but this varied greatly across banks; banks with higher default risk were disciplined more by the market (that is, experienced greater deposit withdrawals), which encouraged them to target a low-risk of default.

Calomiris and Mason (2003a) construct a survival duration model of Fed member banks throughout the country from 1929 to 1933. This model combines aggregate data at the national, state, and county level with bank-specific data on balance sheets and income statements to identify the key contributors to bank failure risk and to gauge the relative importance of fundamentals and panics as explanations of bank failure. Calomiris and Mason find that a fundamentals-based model can explain most of the failure experience of banks in the U.S. prior to 1933. They identify a significant, but small, national panic effect around

September of 1931, and some isolated regional effects that may have been panics, but prior to 1933, banking panics were not very important contributors to bank failures compared to fundamentals.

The fact that a consistent model based on fundamentals can explain the vast majority of U.S. bank failures prior to 1933 has interesting implications. First, it indicates that the influence of banking panics as an independent source of shock to the economy was not important early in the Depression. Only in 1933, at the trough of the Depression, did failure risk become importantly de-linked from local, regional, and national economic conditions and from fundamentals relating to individual bank structure and performance. Second, the timing of this observed rise in risk unrelated to indicators of credit risk is itself interesting. In late 1932 and early 1933, currency risk became increasingly important; depositors had reason to fear that President Roosevelt would leave the gold standard, which gave them a special reason to want to convert their deposits into (high-valued) dollars before devaluation of the dollar (Wigmore 1987).

As part of their bank-level analysis of survival duration, Calomiris and Mason (2003a) also consider whether, outside the windows of “panics” identified by Friedman and Schwartz, the occurrence of bank failures in close proximity to a bank affects the probability of survival of the bank, after taking into account the various fundamental determinants of failure. Calomiris and Mason consider this measure of “contagious failure” an upper bound, since in part it measures unobserved cross-sectional heterogeneity common to banks located in the same area, in addition to true contagion. They find small, but statistically significant, effects associated with this measure. The omission of this variable from the analysis raises forecasted

survival duration by an average of 0.2%. They also consider other regional dummy variables associated with Wicker's (1996) instances of identified regional panics, and again find effects on bank failure risk that are small in national importance.

The large number of bank failures in the U.S. 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.

Central Banking and Bank Instability in U.S. History

Part of the microeconomic rules of the game in any banking system relate to the operations of the central bank, which include its policies for purchasing assets or lending against them, how it funds itself, and the extent to which and the ways in which it competes with other banks. In explaining Great Britain's change from a crisis-prone financial system from 1800 through 1857 to a crisis-resistant system after 1866 – the subject of Section IV – the evolution of the Bank of England's lending policies, its financing, its competition with other banks, and the ways in which banking crises affected the evolution of these three aspects, will be central to the explanation of the stabilization of the system after 1857. But in the U.S., the role of central banking in the history of banking crises was more limited.

The first central bank, the Bank of the United States (BUS), founded in 1791 and chartered for twenty years, was the only nationally chartered institution in the country and the

only one to operate in more than one state. It operated as a for-profit banking enterprise (in which the government owned one-fifth of its \$10 million in initial capital stock), made loans, issued notes and accepted deposits. It was not conceived as a tool for regulating other banks, or acting as a lender of last resort to the financial system. The BUS's most important role in the economy was as a lender to the government and as a fiscal agent for the government, managing the financial flows relating to taxes and debts.

The first panic in U.S. history, in 1792, occurred just as the BUS was gearing up its operations. As Sylla, Wright and Cowen (2009) show, the early experience of the U.S. in dealing with the Panic of 1792 illustrates that central bankers can be sources of banking system risk as well as mitigators of those risks. The nascent BUS actually fueled the panic through an overexpansion of credit in its first months of operation. But the Treasury Secretary, Alexander Hamilton, acted as an ad hoc central banker, inventing and applying "Bagehot's (1873) rule" of lending freely on good collateral at a penalty rate eight decades before that rule would be penned.¹⁰ Thus, although the BUS itself was not a source of stability during the panic, Secretary Hamilton and the Treasury acted as an effective ad hoc lender of last resort.

Sylla, Wright and Cowen (2009) argue that Hamilton's success in undoing the negative effects of the BUS's destabilizing actions had significance beyond its immediate consequences for the financial system; the intervention avoided a political backlash against the Hamiltonian financial system, of which the BUS was a part. Such a backlash in response to failed financial

¹⁰ As Charles Goodhart has pointed out, Bagehot (1873) nowhere employs the phrase "penalty rate." His rule is best understood as lending at a rate in excess of the normal market rate (during non-crisis times), to prevent abuse of the option to borrow. As discussed in more detail below, a more accurate characterization of the evolution of central banking in the late 19th century would emphasize the development of incentive-compatible loss sharing arrangements, of which central bank lending on good collateral is one example.

policy innovations was more than a hypothetical possibility in 1792, given the experience of France and Britain decades earlier:

Earlier in the eighteenth century, John Law had attempted to modernize France's financial system, but his efforts backfired when he failed to prevent the collapse of the Mississippi Bubble in 1720. At the same time, across the Channel, the collapse of the related South Sea Bubble also led to financial crisis. The British financial system, however, was more developed than that of France, as Britain had begun the modernization process in 1688, whereas France did not do so until 1715. A wounded but robust British financial system survived the shock, although legislation passed during the crisis stunted the development of Britain's corporate sector for a century. (p. 63)

This discussion illustrates two broader points: (1) the actions of central banks are not always stabilizing for the banking system, and (2) central banks' privileges can be enacted and also withdrawn. Indeed, the charter of the BUS was allowed to lapse, largely as the result of Jeffersonian objections to the concentration of financial power in a national bank.

Problems in managing fiscal affairs during the War of 1812, along with a desire to reestablish specie convertibility of state bank notes after the suspension of convertibility that had attended the War, led to the establishment of the Second Bank of the United States (SBUS), in which the government subscribed for one-fifth of its \$35 million in capital stock. The SBUS was charged with assisting the government in its financial affairs, reestablishing specie convertibility of other banks' notes, and operating a general banking business. Like its predecessor it operated as the only nationally chartered bank until its charter's renewal was blocked by President Jackson in 1832. In 1836 the SBUS was granted a new charter by the State of Pennsylvania, and operated as a state bank after that date.

As Temin (1969, 46) points out, like its predecessor in 1792, the SBUS became overextended almost immediately after it was chartered. Its western and southern branches

did not coordinate their lending with the eastern branches. Notes issued by the SBUS branches in the periphery to fund loans were presented for payment in the East, where the SBUS branches accepted them at par, thereby encouraging further note issuance and lending by its branches in the West and South. The lack of discipline in 1817-1818 extended to other banks as well. The Treasury asked the SBUS to delay the collection of balances owed to it by state banks in 1817 and 1818, which removed the SBUS as a source of inter-regional discipline over other banks' issuance and promoted increased leverage in the banking system. When as an act of self-preservation the SBUS finally cracked down on its branches in the West and South and on other banks, by demanding that they support their own note issues and pay their outstanding debts, a contraction of credit resulted, which according to Catterall (1902) "precipitated the panic." Gouge (1833) famously quipped that "The Bank was saved and the people were ruined." Public hostility toward the SBUS because of its role in the causing the Panic of 1819 never disappeared (Temin 1969, 48).

The hostility toward the SBUS was further fueled by perceptions of its behavior during the financial crisis in 1825-1826, when once again it acted to limit its own credit and disciplined the state banks by demanding that they redeem their obligations. The public had expected the SBUS to prevent a financial contraction in 1825-1826. According to Hilt (2009), the SBUS was a stabilizing force during the 1825-1826 financial crisis. He argues that, despite widespread financial failures during the crisis, "there was no generalized banking crisis in the United States...in part because of the Second Bank of the United States worked assiduously with its New York branch to provide credit to the banking community there." Hilt (2009, footnote 36) also points to evidence that Nicholas Biddle viewed this as an important part of the mission of

the SBUS. Nevertheless, such lending was limited by the BUS's need to protect itself during the 1825-1826 contraction.

The public hostility toward the SBUS that resulted from its failure to prevent financial crises was largely misplaced. The government itself had encouraged the excessive expansion of credit in the periphery in 1817-1818, and had asked the SBUS to accommodate it. Furthermore, according to Hilt (2009), the SBUS had, in fact, been successful in preventing the financial collapse of 1825-1826 from turning into a bona fide banking crisis as the result of the assistance it provided New York banks. The SBUS's decisions to contract in 1819 and 1825-1826 were necessary to its own preservation; it was, after all, a privately owned bank, and therefore, responsible for its own survival and profitability. Most importantly, the upheavals of 1819 and 1825, like that of 1792, illustrated the limitations of the powers of the BUS and the SBUS. The BUS and SBUS lacked the full-fledged powers of a central bank to deal with crises. Indeed, some financial historians (Temin 1969, 45) have questioned whether the BUS or SBUS qualify to be called central banks. Unlike the Bank of England, the BUS and SBUS did not have the power to issue an unlimited supply of their own bank notes with the implied backing of the sovereign. It is hard to fault the SBUS for failing to use powers that it did not possess.

That is not to say that the SBUS was entirely powerless or unsuccessful in reducing systemic risk in the banking system, as the successful interventions in New York by the SBUS in 1825-1826 illustrate. Because of its special position as the only bank operating branches in various regions, the SBUS was large, had wide geographical reach, and played an especially important role in the bankers acceptance market for financing commerce (intermediation via its various branches the financing of trade flows, as described in Calomiris 2000, Chapter 1) and in

the market for transporting and redeeming the notes of other issuing banks. In addition to its limited powers to assist banks during crises, it could act, and did act, to stabilize the system and help avoid the risk of panics, in two ways: (1) as a source of discipline over other banks' note issuance, it limited the overextension of credit and bank leveraging during booms, and (2) as a unique interregional provider of trade credit, the SBUS reduced seasonal volatility in financial markets related to the planting and harvesting of crops. Bernstein, Hughson, and Weidenmier (2009a) find evidence in support of increased average risk, and greater seasonality of risk, after the failure to re-charter the SBUS. From 1816-1836 stock return volatility across the months of September and October (the harvesting season) averaged 2.45 percent, virtually identical to the 2.43 percent for the rest of the year. Following the SBUS's demise, from 1837-1860, stock return volatility rose to 6.30 percent in September and October versus 5.02 percent during the rest of the year. These volatilities rose even higher during the National Banking Period, where they were 7.30 and 5.80 percent, respectively.

Despite its stabilizing role in the financial system, the inability of the SBUS to prevent financial crises, along with various other political and ideological battles in which the SBUS and its leader, Nicholas Biddle, became embroiled, ultimately resulted in a fierce battle over the future of the bank, and Jackson's eventual veto of its re-chartering.¹¹

The history of the SBUS illustrates a broader theme in the early history of central banking. Central banks, including the BUS, the SBUS, the Bank of England, and others, were chartered as for-profit company's with special privileges (in the case of the BUS and the SBUS,

¹¹ For an excellent treatment of the ideological conflicts leading to the veto, and the gamesmanship on the part of the bank's advocates, led by Nicholas Biddle, and President Andrew Jackson during that struggle, see Schweikart (1988).

the special privileges had to do with their unique branching operations, and their unique relationship with the government) that also gave rise to the expectation that they would undertake special responsibilities to the public in addition to maximizing their economic value for their stockholders. That dual mandate of profitability and social responsibility implied that central banks not only had to satisfy their stockholders that they were achieving a good return and acting prudently, but also that they had to satisfy the public, through elected officials, that they were achieving their social mission. Because achieving social missions (like making markets in risky bank notes at par) tends to be costly, there has often been an inherent conflict between the private profitability of central banks and their public missions.

That conflict is precisely what makes 18th and 19th century central banks so interesting to financial historians (Goodhart 1995). Even if central bank management only cared about value maximization (a distinct possibility), bank managers had to navigate the political process that was the source of their valuable special privileges. In other words, even a selfish central banker had to find a way to optimally give services to the public – providing enough continuing service to society to prevent the central bank’s privileges from being revoked. Some central bankers, like the managers of the BUS and SBUS, did not succeed in satisfying the demands for public service, and thereby lost their charters. Other central bankers, like those at the Bank of England, successfully adapted to changing circumstances (especially financial crisis in the 19th century) to satisfy changing public demands, and thereby managed to keep their charters and their jobs, although the nature of their privileges and responsibilities changed significantly over time as the game between the central banks and the public evolved. (We will return to that

theme in Section IV's discussion of the evolving role of the Bank of England and its consequences for the stabilization of the British banking system in the 19th century.)

The U.S. banking system operated without a central bank from 1836 until 1913 when the Federal Reserve System was established. The establishment of the Fed was a direct reaction to the Panic of 1907, and the perception that private bankers acting as a coalition (organized to some extent through their local clearing houses, and through ad hoc efforts like those undertaken in 1907 by J.P. Morgan) had insufficient ability to preserve systemic stability. In the wake of the 1907 Panic, the National Monetary Commission was established, and it issued a voluminous and substantive report in 1910, which formed the factual and theoretical basis on which the Federal Reserve Act of 1913 was constructed, and to this day the National Monetary Commission report still contains some of the most valuable information about the operations of banks of that era throughout the world.

The Federal Reserve System, like all central banks, was a creature of a political process and compromise that balanced various competing interests, and that compromise evolved over time. The structure of the system (12 regional Reserve Banks and a Board in Washington, with member bank ownership of the Reserve Banks) evolved into a system effectively owned by the taxpayers but still managed by a process of power sharing that gave weight to local bank and business interests (who control the Reserve Banks' boards and executive appointments), political leaders in Washington (who appoint Federal Reserve Board members and to whom the Board reports), and rural interests (who received special favors in the structuring of clearing arrangements and in the use of agriculture-related loans as collateral).

The philosophical foundations of the Fed are rather amorphous, and much of the logic that was embodied in its initial rules has been discredited by monetary economists, notably the “real bills doctrine” that was supposed to govern its lending operations. Suffice it to say for our current purposes that the Fed obtained broad powers to lend to member banks against good collateral (initially construed as high-quality commercial bills, and later, also government securities) and to engage in open market operations to control the supply of reserve holdings by member banks at the central bank.

Importantly, the Fed’s charter and its powers did not envision it as a crisis manager for failing banks or as a bailout agency, and the Fed’s role in causing or averting banking crises primarily revolved around the way its policies affected *market* prices and flows, rather than the affairs of particular banks. The Fed was designed to use lending and other actions to regulate the aggregate supply of reserves, money, and credit in a way that would reduce seasonal and cyclical volatility of interest rates and increase the seasonal and cyclical elasticity of reserves and loans, (initially, it was expected to accomplish those seasonal and cyclical objectives while remaining on the gold standard, a goal that was temporarily put aside several times, and permanently abandoned in 1973).

The record of the Fed as a source of stability for the banking system is mixed. On the one hand, the Fed was sometimes a source of great instability in the system because the policy rules it followed for targeting monetary policy were often ill-conceived. Friedman and Schwartz (1963), and many others since, have shown that Fed monetary policy errors produced the monetary collapse that caused the economic and banking crises of the 1930s, and Calomiris and

Mason (2003a) document measurable connections between the deteriorating macroeconomic and local economic environments in which banks operated and the resulting bank failures.

Wicker (1966), Brunner and Meltzer (1968), and Wheelock (1991) trace the Fed's policy errors in the 1930s and at other times to the misuse of interest rates and borrowed reserves as short-term monetary policy instruments. Wheelock (1991), in particular, argues that it was a consistently faulty monetary policy methodology, rather than a lack of leadership at the Fed following Benjamin Strong's death (which Friedman and Schwartz 1963 posit to explain Fed failures after 1929), that explains the policy errors that gave rise to the Great Depression. The monetary policy errors that caused the Great Depression show that vesting authority in a central bank can be risky; although the central bank may intend to stabilize the system, it may, in fact have the opposite effect.

On the other hand, there is substantial evidence (Miron 1986, Richardson and Troost 2006, and Bernstein, Hughson, and Weidenmier 2009b) that the founding of the Fed reduced liquidity risk in the banking system, which in turn reduced the propensity for bank panics. Miron (1986) showed that the founding of the Fed was associated with reduced seasonal variability of interest rates and increased seasonal variability of lending. Miron, however, did not explain how the Fed achieved this result. Why, exactly, did Fed lending practices make the loan supply function more elastic?

Miron's (1986) findings can be explained by a variant of the deposit risk targeting model in Calomiris and Wilson (2004). In that model, the riskiness of deposits is a function of bank asset risk and bank leverage. Because total bank capital and total cash assets in the economy do not vary much over the year, a seasonal increase in bank lending (especially to finance crop

harvesting and transport in the fall, which Davis, Hanes and Rhode 2007 show was largely driven by the cotton cycle) implies a commensurate increase in bank asset risk and in bank leverage, which unambiguously means an increase in the riskiness of deposits (the actuarially fair default risk premium). This is a source of seasonal variation in the risk of deposit withdrawals, since market discipline makes the risk of withdrawal in the deposit market sensitive to increases in default risk (i.e., some depositors are intolerant of risk, and will withdraw when risk increases). A bank that increases its lending, *ceteris paribus*, faces increased deposit withdrawal risk, particularly if an adverse cyclical shock hits during a seasonal lending spike. All six of the major banking panics of the pre-World War I era happened at cyclical peaks; they were clearly responses to adverse economic shocks to banks' balance sheets (Calomiris and Gorton 1991). Furthermore, these panics all occurred either during the spring planting season or the fall harvest, at times when lending (and bank liquidity risk) was at a seasonal peak.

From the perspective of this model, the founding of the Fed provided a means of reducing liquidity risk to banks by giving them a source of liquidity to stem deposit withdrawals (making them less vulnerable to withdrawal risk at times when seasonal lending peaks coincided with cyclical downturns). The founding of the Fed thus flattened the bank loan supply function, making loans vary more over the cycle, and interest rates vary less.

Bernstein, Hughson, and Weidenmier (2009b) provide additional evidence consistent with that interpretation. They compare the standard deviations of stock returns and short-term interest rates over time in the months of September and October (the two months of the year when markets were most vulnerable to a crash because of financial stringency from the harvest season) with the rest of the year before and after the establishment of the Fed. Stock volatility

in those two months fell more than 40 percent, and interest rate volatility more than 70 percent, after the founding of the Fed. Like the SBUS before it (discussed above), the Fed succeeded in reducing seasonal variations in liquidity. They also show that this result is driven by years in which business cycles peaked. In other words, the main risk that the Fed eliminated was associated with combined cyclical peaks in economic activity and seasonal peaks in lending.

Many commentators have faulted the Federal Reserve for failing to prevent bank failures during the Great Depression with more aggressive discount window lending. While it is certainly true that expansionary monetary policy, particularly in 1929-31, could have made an enormous difference in preventing bank distress (through its effects on macroeconomic fundamentals), that is not the same as saying that more generous terms at the discount window (holding constant the overall monetary policy stance) would have made much of a difference. Discount window lending only helps preserve banks that are suffering from illiquidity, which was not the primary problem underlying large depositor withdrawals.

Indeed, in 1932, President Hoover created the Reconstruction Finance Corporation (RFC), to enlarge the potential availability of liquidity, but this additional source of liquidity assistance made no difference in helping borrowing banks avoid failure (Mason 2001). As commentators at the time noted, because collateralized RFC and Fed loans were senior to deposits, and because deposit withdrawals from weak banks reflected real concerns about bank insolvency, loans from the Fed and the RFC to banks experiencing withdrawals did not help much, and actually could harm banks, since those senior loans from the Fed and the RFC reduced the amount of high quality assets available to back deposits, which actually increased the riskiness of deposits and created new incentives for deposit withdrawals. In 1933,

however, once the RFC was permitted to purchase banks' preferred stock (which was junior to deposits), RFC assistance to troubled banks was effective in reducing the risk of failure (Mason 2001).

Despite the limitations inherent in the ability of collateralized lending to prevent bank failure, there is some evidence that greater Fed assistance to banks early in the Depression could have been helpful in avoiding some bank failures. Richardson and Troost (2006) show that, despite the limited ability of Fed discount window lending to absorb credit risk, Fed provision of liquidity to member banks mitigated bank failure risk associated with illiquidity somewhat in 1930, and could have played a greater role in stemming illiquidity-induced failures if the Fed had been more willing to relax lending standards to member banks. They study the failure propensities of Mississippi banks. The Federal Reserve Act of 1913 divided Mississippi between the 6th (Atlanta) and 8th (St. Louis) Federal Reserve Districts. The Atlanta Fed championed a more activist role in providing loans to member banks experiencing troubles, while the St. Louis Fed rigidly adhered to the real bills doctrine and eschewed the extension of credit to troubled banks. Mississippi banks in the 6th District failed at lower rates than in the 8th District, particularly during the banking panic in the fall of 1930, suggesting that more aggressive discount window lending reduced failure rates during periods of panic.

Summary of U.S. Historical Experience

The unusually unstable U.S. historical experience of frequent nationwide banking panics (1819, 1837, 1857, 1873, 1884, 1890, 1893, 1896, 1907, and 1933) and a propensity for unusually severe and widespread waves of bank failures (the 1830s, the 1920s, and the 1930s) reflected a unique feature of the microeconomic structure of U.S. banking – namely the

fragmented banking structure of unit banking – which made it harder to diversify lending risk ex ante and coordinate the management of banking system risk ex post.

Comparisons across regions and across states within the U.S. also reveal important cross-sectional differences in banking stability that are similarly traceable to structural features. The presence of branch banking, clearing houses, or other local institutional arrangements for collective action were stabilizing forces, but these stabilizing mechanisms were only permitted on a local or statewide basis. The presence of deposit insurance, which was advocated by unit bankers as a means of protecting them from debt market discipline, resulted in adverse selection in bank entry and moral hazard in bank risk taking, and was a destabilizing force that produced the worst localized bank failure experiences of the 1830s and the 1920s.

Early experiments with limited central banking in the U.S. resulted in the failure to re-charter central banks twice in the early 19th 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 SBUS of contributing to financial instability through contractionary policies prior to and during both the Panic of 1819 and the financial crisis of 1825-1826, 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 BUS or SBUS were equipped to act as lenders of last resort during crises, the SBUS succeeded in reducing systemic financial risk on average and over the seasonal cycle, foreshadowing the stabilizing effect of the Fed after 1913.

After the demise of the SBUS, the U.S. functioned without a central bank until the founding of the Fed in 1913. The record of the Fed vis a vis banking crises is mixed. On the one hand, the Fed (like the BUS in its first year of operation) could be a source of substantial risk to the system, resulting from inappropriate policy responses. The mistaken use of borrowed reserves and interest rates as monetary instruments created false impressions in 1929-1932 that encouraged monetary contraction, which precipitated the Great Depression, the real effects of which produced massive bank failures in the 1930s. On the other hand, the existence of the discount window substantially reduced systemic liquidity risk, especially the risk that banks would be caught in an illiquid position at times of seasonal peaks in lending that coincided with cyclical peaks in economic activity. Although the ability to employ the discount window to stem bank failures during the Depression was limited – since shocks buffeting banks were primarily related to solvency rather than illiquidity – there is evidence that relatively aggressive discount window lending by the Atlanta Fed during 1930 did help to prevent some bank failures.

In summary, the microeconomic rules of the banking game – the unit banking structure of the industry, the occasional reliance on destabilizing deposit insurance, and the lack of an effective lender of last resort for the pre-World War I era – all contributed to the peculiar historical instability of the U.S. banking system. The key destabilizing elements of the U.S. system – a fragmented industrial structure, the absence of an effective lender of last resort, and the occasional presence of a destabilizing deposit insurance regime – compounded one another. Canada, which avoided chartering a central bank until 1935, managed to avoid banking crises due to the stabilizing role of its branch banking system, despite the absence of a

central bank. In the U.S., 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 destabilizing effects of unit banking.

III. A Worldwide Tale of Two Banking Eras: 1875-1913 and 1978-2009

Although the U.S. was unique in its high propensity for panics (reflecting its peculiar banking structure), and it occasionally experienced high rates of banking loss, other countries sometimes experienced loss rates that exceeded that of the U.S. In the pre-World War I period (1875-1913), the highest nationwide banking system loss rate (i.e., the negative net worth of failed banks relative to GDP) for the U.S. was roughly 0.1%, which was the loss rate for bank failures in the panic of 1893. Other countries generally experienced even lower bank failure rates, but there were a handful of episodes in the world (between 4 and 7) during this period in which the negative net worth of failed banks exceeded 1% of GDP (a minimal severity standard used by Caprio and Klingebiel to gauge banking crises today).¹²

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

¹² Two caveats are in order. First, due to the lack of existing data, there is uncertainty about whether one or more of Brazil's various pre-World War I financial crises may have produced losses in excess of 1% of GDP, and this accounts for the claim that the number is between 4 and 7. Second, the number of countries is much fewer in the historical sample than in the post-1978 sample. Nevertheless, the thrust of the comparison is still valid; the frequency and severity of bank insolvency events has increased dramatically.

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 (Calomiris 2007).

By recent standards, 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. In contrast, over the past thirty years roughly 140 episodes have been documented in which banking systems experienced losses in excess of 1% of GDP, and more than 20 episodes resulted in losses 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 1982-82, Mexico in 1994-95, Korea in 1997, and Thailand in 1997, and a greater than 50% loss in Indonesia in 1997).¹³

Loss rates in the pre-World War I period tended to be low because banks structured themselves to limit their risk of loss by maintaining adequate equity-to-assets ratios, sufficiently low asset risk, and adequate liquidity. Market discipline (the potential for depositors fearful of bank default to withdraw their funds) provided incentives for banks to behave prudently (for a theoretical framework, see Calomiris and Kahn 1991). The picture of small depositors lining up around the block to withdraw funds has received much attention by journalists and banking theorists, but perhaps the more important source of market discipline was the threat of an informed (“silent”) run by large depositors (often other banks). Banks maintained relationships with each other through interbank deposits and the clearing of deposits, notes, and bankers’

¹³ Data are from Caprio and Klingebiel (1996), updated in private correspondence with these authors, and by subsequent additions.

bills. Banks often belonged to clearing houses that set regulations and monitored members' behavior. A bank that lost the trust of its fellow bankers could not long survive.

Recent research attempting to explain the unprecedented systemic bank failures worldwide over the past three decades has emphasized the destabilizing effects of bank safety nets. This has been informed by the experience of the U.S. Savings and Loan industry debacle of the 1980s, the banking collapses in Japan and Scandinavia during the 1990s, and similar banking system debacles throughout the world. 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 (see, for example, Caprio and Klingebiel 1996, Boyd et al. 2000, Demirguc-Kunt and Detragiache 2000, Barth, Caprio, and Levine 2006; Demirguc-Kunt, Kane, and Laeven 2009).

It is also significant that the four countries that suffered the most severe bank failure episodes of the pre-World War I era – Argentina, Australia, Norway, and Italy – had two things in common: (1) All of them suffered real estate boom and busts that exposed their financial systems to large losses; and (2) prior to these crises all of them had employed unusually large government subsidies for real estate risk taking that were designed to thwart market discipline (Calomiris 2007). In Argentina, that subsidy took the form of special mortgage guarantees issued by the government, which guaranteed holders of the mortgages repayment. Banks were licensed to originate these guaranteed mortgages, and then resold them as guaranteed liabilities in the London market, where they were traded as Argentine sovereign debts. This is akin to deposit insurance in that it makes the financing cost of the mortgage invariant to its

risk, which entails the same moral hazard as deposit insurance: the guarantee makes the profitability of mortgage lending increasing in the riskiness of the mortgage portfolio, and thus encourages originators to lend to risky borrowers.

The Australian case was a bit different; financial market policies toward the private sector were not the primary means through which the government promoted the land boom that preceded the bust of 1893. The pre-1890 Australian economic expansion was largely an investment boom in which the government played a direct role in investing in land and financing farmers' investments. Government investments in railroads, telegraphs, irrigation, and farms were financed by government debt floated in the British capital market and by government-owned savings banks and postal savings banks (M. Butlin 1987, N. Butlin 1964, S. Butlin 1961, Davis and Gallman 2001).

The less dramatic banking system losses during the Norwegian and Italian land busts reflected less aggressive, more regionally-focused government policies promoting land development. In Norway, that was achieved through government-sponsored lending and accommodative monetary policy; in Italy, this was achieved through liability protection for the Banca di Roma, which financed a Roman land boom at the behest of the Pope, who had lobbied for protection of the bank's liabilities (Canovai 1911). The Norwegian banks' losses amounted to roughly three percent of GDP, and the Italian banks' losses (which largely reflected exposures to the Roman land market) were roughly one percent of GDP (Calomiris 2007).

The theory behind the problem of destabilizing subsidization of risk taking has been well-known for well over a century, and I have already noted that it was the basis for opposition to deposit insurance in the U.S. in 1933. Deposit insurance was seen by opponents as

undesirable special interest legislation designed to benefit small banks (Calomiris and White 1994). Roosevelt, Glass, and others acquiesced for practical political reasons, to get other legislation passed, not because they wanted deposit insurance, per se. Bad economics is sometimes good politics. Similarly, Argentine mortgage subsidies were transparently intended to benefit landowners in the pampas, just as the real estate risk subsidies in Australia, Rome, and Norway were conscious attempts to support constituencies that favored real estate development.

It is worth emphasizing that all of these risk subsidizing government interventions (mortgage guarantees, liability insurance, government lending on land) were intended to overcome market discipline that had been limiting risk taking. Whatever their merits, these interventions served powerful special interests by subsidizing real estate risk, destabilized their country's banking systems, and produced substantial losses. Bank insolvency crises in the pre-World War I era fundamentally were about imprudent government policies.

Research on the banking collapses of the last three decades offers a similar message. Empirical findings uniformly show that the greater the role of government in directing credit or in providing protection to private banks through the government safety net (e.g., deposit insurance), the greater the risk of a banking collapse (Caprio and Klingebiel 1996, Boyd et al. 2000, Demirguc-Kunt and Detragiache 2000, Barth, Caprio, and Levine 2006; Demirguc-Kunt, Kane, and Laeven 2009). Some of this research has identified the political economy of subsidizing risk taking as a core problem, and one that would-be reformers and financial regulators have had a difficult time overcoming. Empirical research on prudential bank regulation emphasizes the inefficacy of government regulations in preventing risk taking (since

they are subject to the same political forces that purposely subsidize risk), and the importance of subjecting some bank liabilities to the risk of loss to promote discipline of risk taking as the primary means of reining in excessive risk taking (Board of Governors 1999, Shadow Financial Regulatory Committee 2000, Mishkin 2001, Calomiris and Powell 2001, Barth, Caprio and Levine 2006) – in other words, finding a means to use markets to constrain risk taking.

These studies of recent experience echo the conclusions of the studies of historical deposit insurance discussed above (Calomiris 1990, 1992), and the historical experiences of Argentina, Australia, Norway and Italy in the pre-World War I era. The difference is that what used to be the exception – moral hazard and adverse selection resulting from government protection that give rise to excessive risk taking – has become the rule.¹⁴

This evidence stands in sharp contrast to the approaches of Minsky (1975), Kindleberger (1973), and Diamond and Dybvig (1983) for explaining bank fragility. Rather than seeing market behavior, human nature, and the market-determined structure of bank balance sheets as the root cause of banking crises, this new literature argues that the solution to banking crises lies in empowering markets to rein in the risk taking that is otherwise subsidized by the government.

IV. The Political Economy of Central Banking: The Stabilization of British Banking, 1800-1900

One of the most fascinating historical examples of a change from banking instability to stability occurred in Great Britain in the middle of the 19th century. As Capie (2009) notes,

¹⁴ The initial experience with federal deposit insurance in the U.S. was one of unusual stability. That experience, from World War II through the 1960s, reflected the unusual stability of interest rates, asset prices, and growth in that era, in contrast to the periods before or afterward, and even more importantly, the limited insurance of deposits in the early decades of deposit insurance (size limits were increased in phases over time; the system initially covered only a small fraction of banking system deposits, but has grown to now cover virtually all deposits).

Britain experienced major banking panics in 1825, 1836-39, 1847, 1857 and 1866, but then the propensity for panic ended for over a century. Scholars have traditionally credited changes in Bank of England policies with this transformation (Bagehot 1873, Andreades [1909] 1966, Hawtrey 1932, 1938, King 1936, Clapham 1944, Hughes 1960, Capie 2002, 2009). Prior to 1858, the Bank's policy was to accept a virtually unlimited amount of paper for discount at a uniform rate, both in the bubble phases leading up to financial crashes, and in the aftermath during the scramble for liquidity. In 1858 the Bank changed its discounting policies to make them significantly less generous, and over time, the Bank came to rely on variation in its discount rate and the use of open market operations as means of promoting monetary and financial stability (Hawtrey 1932, 1938). Furthermore, the decision not to assist Overend, Gurney & Co. during the crisis of 1866 was also significant, since it demonstrated that the newly announced 1858 policy change was credible. Together, policy actions in 1858 and 1866 substantially reduced problems of moral hazard in British banking.

Other changes in the microeconomic rules of the banking game, most notably greater freedom of entry into banking and expanded branching, were also complementary and helpful in stabilizing the system, partly because of how they affected the political equilibrium in which the Bank operated, and because greater banking system integration across regions reduced the demand for discounting bills.

All of these changes encouraged the evolution of the Bank from a destabilizing central bank with an uncertain future – one crippled by an inconsistent mixture of private and public missions that was forced to play a complex and uncertain political game to retain its privileges – into an institution with a relatively clear set of objectives and policies, which operated in a

relatively stable financial and political environment. The combination of a well-defined understanding of the public mission of the Bank, and its success in managing systemic risk meant that public expectations of the Bank were less likely to be disappointed. The important political transition was from an early unstable political cycle of accommodation, crisis, conflict, and political reaction (over the course of each decade's financial crisis) to a political equilibrium after 1866 with a well-understood delineation of the Bank's responsibilities, effective rules to achieve those goals, and an absence of crises with political backlashes.

In addition to developing its operational tools for normal times (using variation in the discount rate and open market operations) the Bank developed a philosophy for assisting the financial system which was embodied in two incentive-compatible mechanisms of central bank assistance: (1) Bagehot's (1873) famous rule for lending during panics on good collateral at a penalty rate, and (2) risk sharing between the Bank and the London clearing banks in the provision of credit guarantees to forestall the spread of crises, which was exemplified by the agreement among London banks to stand behind Barings in 1890. As discussed further below, the "meta-rule" that subsumes both of these forms of crisis management is that central bank interventions to prevent crises should place the central bank in a senior position relative to other banks with respect to the absorption of loan losses.

This section briefly lays out the evolution of these various institutional changes, and places them in the appropriate political context. The central themes of my account are as follows: (1) The early 19th century propensity for panics reflected the moral hazard associated with the Bank's operations in the discount market. (2) Those operations reflected the Bank's mixed status as a private institution with a public mission; and reflected its attempt to preserve

and take advantage of its special monopoly privileges and manage the political risk of a potential change in those privileges. (3) The recurring financial crises that this strategy entailed made this approach increasingly undesirable for the Bank, both because it implied large economic risks for the Bank, and because it undermined public support for the Bank. (4) The consequences of this political and economic disequilibrium produced by recurring banking crises required a combination of changes to restore economic and political equilibrium, including a diminution of the Bank's monopoly position, and changes in its discounting and lending practices. (5) That evolution restored stability to the banking system by removing the primary source of moral hazard (Bank discounting), fostering greater market competition and discipline, and encouraging the development of modern central bank practices, including Bagehot's rule, and incentive-compatible risk sharing between private London banks and the Bank in response to the threats of panics (like the Barings Crisis of 1890). The remainder of this section develops this argument in more detail.

(1) The early 19th century propensity for panics reflected the moral hazard associated with the Bank's operations in the discount market.

The British banking system developed a peculiar organizational structure in the early 19th century that encouraged financial bubbles to form. This was the result of a combination of a lack of effective prudential limits on bank leverage or the growth of risky lending, and pressures placed on the Bank of England to accommodate lending booms, which took the form of booms in the London discount market. The Bank of England was uniquely situated at the center of the financial system, with a monopoly of note issuance in London (banks outside of London could also issue a limited amount of notes, but they were not significant in size, and the

Bank's notes were uniquely a legal tender). It also enjoyed barriers to bank entry that limited competition, although these were relaxed substantially over time.

Under the Bank Act of 1844 (the brainchild of the so-called "currency school"), the Bank was required to maintain 100% specie reserves against its note issues. That was considered a means to make the equilibrating international specie-flow mechanism more effective by linking the total supply of currency to the balance of payments. Of course, notes and specie were not the total money supply; bank deposits were widely used and bills of exchange were "the principal means of payment" used in business transactions (King 1936, 32), and thus the 1844 Act did not effectively constrain credit growth, deposit growth, or bill of exchange growth. In fact, credit growth and specie flows often moved in opposite directions, as in the 1850s when specie was flowing out of London (having mainly to do with bimetallic arbitrage involving France) while credit was growing dramatically (Hughes 1960, 250-56). From 1852 to 1857, currency outstanding fell from 34 million to 25 million pounds, while the deposits of the five London joint stock banks rose from 17.7 million to 40 million pounds, and the average volume of bills of exchange in circulation (the asset intermediated in the London discount market) rose from 66 million pounds to an estimated 180-200 million pounds (Hughes 1960, p. 258). The Bank's accommodative lending policies, combined with the absence of any prudential limits other than the 100% note reserve requirement (e.g., the absence of capital or reserve requirements relating to deposits), promoted huge growth in discounts and deposits during booms.

The precise intermediation mechanics of this remarkable growth in deposits and bills during booms is the subject of King's (1936) masterful treatise on the London discount market,

which is also analyzed by Hughes (1960) in his lengthy Appendix 5. A detailed analysis is beyond our scope here.¹⁵ Suffice it to make two points: (1) the effects of the Bank's willingness to discount bills of exchange had ramifications throughout the financial system, and this was especially visible in panic years, and (2) the London discount market's importance in the financial system reflected the limitations on bank branching at the time – the discount market operated as the primary means of redistributing financial resources from locations of surplus deposits (relative to lending opportunities) to locations of surplus loan opportunities relative to deposits. As King (1936, xii-xiii) put it: "a localized banking system could function efficiently only through a sort of central pool by means of which the credit surpluses of one type of district (the agricultural areas) and the credit deficiencies of the other type of district (the industrial areas) could find their level." The London discount market was that central pool.

The Bank's accommodation of the discount market was relied upon to manage the adjustment to financial collapses by having the Bank stand ready to discount bills as needed, which it did on a widespread basis.¹⁶

The deposit runs on the London discount houses and their efforts to cover their deposits led to a surge of applications to the Bank for aid during the [1857] crisis. Sanderson, Sanderman & Co., the largest London discount house to stop payment, had 3.5 million of deposits; a run on which caused Sanderson's to suspend on 11 November....From 10 to 12 November Overend & Gurney alone were given 1.8 million. This inspired the

¹⁵ Hughes (1960, 306) provides a succinct analysis of the structure of intermediation detailed in King (1936): "the rise of the joint-stock banks and the transformation of the bill brokers into bill dealers who accepted money at call and discounted bills on their own accord had been complementary developments. After 1825 the London banks had given up the practice of rediscounting at the Bank and had begun employing their excess funds with the discount houses. After 1833 this practice had been given a great impetus by the bill market being granted discount facilities at the Bank. By the 1850s when the deposits of the joint-stock banks surged upwards, the practice of keeping their money at call with the discount houses had become the most important form of employing the deposits of the banks."

¹⁶ The Bank was less generous in its treatment of Scottish banks, and two large Scottish banks failed in the 1857 crisis, although Hughes argues that the Bank was right to deny credit due to the insolvency of these institutions (Hughes 1960, 311-31).

remark after their failure in 1866 'Overends broke the Bank in 1866 because it went, and in 1857 because it was not let go.' (Hughes 1960, 305).

One aspect of the crisis of 1857 which has not been given proper emphasis by historians is that the Bank's discounts and advances were given to all sectors of the business community. Too much emphasis has been given to the needs of the discount houses alone. Of the 79.4 million of accommodation given in 1857 35.8 million were given in the last three months of the year. This represents for the most part the discounts of the panic period. Of this amount, half, or 17.8 million, was direct loans and discounts to merchants. Scottish banks received 1.3 million, country banks (including joint-stock banks in London and banks outside London) had 7.1 million, and the discount houses 9.5 million. Of the country banks, those outside London received almost 6 million, while those inside the city received only about 1 million. The heaviest demand for accommodation came from the merchants, followed by the discount houses and the country bankers. (Hughes 1960, 302-03)

How was the Bank able to accomplish this? Given the political imperative of promoting recovery during crises, the 1844 Act's provisions were suspended three times, during each of the three post-1844 crises, to allow the Bank of England to accommodate demand by issuing additional notes to use in discounting and lending, unconstrained by the availability of specie. Hughes (1960, 272) shows that the decision to relax the 1844 Act's provisions after November 12, 1857 resulted in an immediate easing in the market. As Bagehot (1873), Capie (2002) and others have emphasized, the Bank's note privileges were the key to its position as a provider of liquidity. Its ability to issue legal tender notes (after 1833), especially if unfettered by the 100% reserve requirement, and with the implicit backing of the state, allowed the Bank to provide substantial liquidity to the market in spite of the fact that, in theory, it was just another privately owned bank.

The fact that the government chose to amend the Bank's charter in 1833 to make its notes a legal tender indicates a conscious intention to enhance the capability of the Bank to provide credible support to the market during crises beyond the capacity of its own net worth.

The lead advocate in Parliament of making Bank of England notes a legal tender, Lord Althorp, explicitly pointed to the advantage of doing as freeing the bank from market discipline and potential redemptions of its notes during crises (Andreades [1909] 1966, 260). The capacity to issue legal tender notes without limit during crises set the Bank apart from many other central banks in its capacity to act as a lender of last resort, including the previously discussed examples of the BUS and SBUS.

Why did private market discipline fail to constrain growth in deposits and discounted bills, which so predictably resulted in crises? Hughes (1960, 264) remarks that “[t]his inflationary bill expansion represented at any time a potential demand for payment in gold or notes even though the ability of the banking system to provide payment in gold or notes seems to have entered little or not at all into the calculations of those whose expanding credit activities were responsible for the potential demand.” Why not? Hughes (1960, 264-5), citing Newmarch, provides a partial answer: market participants expected to be protected by the Bank. Indeed, some of the most rapid growth in bills intermediation occurred immediately after the onset of the liquidity squeeze, indicating that the banking system continued to expand systemic risk even after the onset of problems. Newmarch argued that this reflected the fact that the public knew that bills could be presented for discount (i.e., that the Bank of England had effectively guaranteed their liquidity).

Of course, there was no explicit guarantee to convert bills to cash at the Bank of England, but there was an implicit one. David Salomon, director of the London and Westminster Bank, saw this implicit guarantee as central to the problem of credit-fueled bubbles:

And do you think it is part of the functions of the Bank of England to discount a bill for anybody merely because the party holding the bill wishes to convert it to cash? As I said before, the Bank of England will have great difficulty in getting rid of that inconvenient idea which there is in the mind of the public, that the Bank of England is something more than an ordinary joint-stock bank. (Hughes 1960, 300)

The subsidization of risk by the Bank did not follow from any attempt to undercut the market rate. In fact, the Bank's rate was above market during the 1850s (Hughes 1960, 301).

The essence of the Bank's subsidy to the market was the put option inherent in the Bank's willingness to accommodate demand, which meant that whatever rate the bank charged, the market could comfortably discount at below that rate.¹⁷

(2) Those operations reflected the Bank's mixed status as a private institution with a public mission; and reflected its attempt to preserve and take advantage of its special monopoly privileges and manage the political risk of a potential change in those privileges.

The Bank of England did not want to provide risk-taking subsidies to the market, since doing so was not profitable, but the Bank was under intense pressure to do so, both in the boom periods leading up to crises and during the process of resolving the crises. King (1936, 71-72) recounts how London's merchants organized protests to pressure the Bank not to constrain its discounting policies as early as 1793 and 1795. The protesting merchants issued a joint statement threatening that if the Bank were unwilling to maintain sufficient discounting, "it will be requisite that some other Public Establishment should be created to supply the Deficiency; at the same time wishing that this assistance may be derived through the old and customary

¹⁷ Describing the discounting of bills as a put option is a bit of an overstatement. The Bank could, and did, raise the rate of discount during panics, so the price of the option was a moving target. Indeed, Hughes (1960, 371) remarks that the decision to raise the rate during the panic alarmed the market, which one could interpret as reflecting an adverse shock to market expectations. Also, the relaxation of the 1844 Act was not a certainty, so the ability to exercise the put option was subject to doubt.

Channel, the Bank of England.” In other words, if the Bank of England refuses to obey the needs of the public, the public should charter a new bank to meet those needs. Later, attempts by the Bank to establish a classification system for bills, and price bills of different quality accordingly, also met with merchant opposition and was abandoned (King 1936, 53).

The Bank’s reluctance to play the role of benefactor continued, especially during financial crises. In both 1847 and 1857, the Board of the Bank of England advised the government against suspending the 1844 Bank Act’s provisions (Hughes 1960, 320-21, Bagehot 1873, 65-66), but to no avail. Clearly, the government saw more advantage in clearing the way for the Bank to provide assistance to the market than the Bank did.

In principle, the Bank could have refused to increase its note issuing beyond the amount of its specie reserve in 1847 and 1857, and the Bank could have refused to accommodate bills, both before and during banking crises. But that would not have been possible in practice. As Schuster (1923, 8) notes:

...beyond strict compliance with the Act [of 1844], no special duty is, by law, imposed upon the Bank; yet that such duties exist through an unwritten law, that they have been recognized and are acted upon, is beyond doubt. They affect our commercial life so closely and are so indissolubly connected with the functions and duties which are properly those of the State that to look upon the Bank of England merely as a private trading institution, and not as virtually the State or Government bank, is an impossibility.

At the heart of the matter is the simple fact that the Bank relied on the government, and on public opinion, to maintain its special privileges, and “an institution so dependent on the Government of the day for the continuance of valuable rights was little able, as Mr. Ricardo observed, to withstand the cajoling of Ministers” (Schuster 1923, 11). In this regard it is important to recall that the Bank’s charter was subject to revocation by Parliament.

Interestingly, the political option to revoke the charter was established as a quid pro quo in 1833 (to take effect with a 12 year lag), in the same Act of Parliament that gave the Bank's notes legal tender status (Andreades [1909] 1966, 262).

The responsibility of the Bank to provide accommodation was not a matter of statute. The Bank Act of 1844 clearly explained the duties of the Issue Department of the Bank, but not the implicit duties of the Banking Department (namely, to accommodate unlimited bill discounting demand at a uniform rate); nonetheless, those duties were real, and market advocates made no bones about them. Schuster (1923, 21-22) summarizes the evidence from the Report of 1858 about the market's expectation of those duties:

[According to Mr. Arbuthnot] 'There can be little doubt...of the advantage which accrues to commerce from the employment of these funds, either directly or indirectly, in the discount of bills....they ...find their way into the money market and are applied to the purposes of trade; but when the demands for money are great, and the rate of interest consequently high, great advantages are afforded by the resource the Bank of England affords under the system of management now pursued; such houses are assured that the funds at the disposal of the Bank will always be available for the legitimate objects of trade at the current rate of interest. Whence ensues that confidence which is derived from uniformity of system. The Bank of England has then come to be regarded as the centre and mainstay of mercantile credit.' That the Bank was so regarded is clear from the tendency of the whole of the enquiry. Witnesses were specially asked whether every house which applied, and deserved assistance, received it. From the evidence it appears that the Directors of the Bank of England went into the country to examine the accounts of banks in difficulties, in order to render assistance if they appeared to be sound.

The Governor of the Bank of England was asked: 'You did not refuse accommodation to any person, even up to the time when the Act was suspended, who brought you good securities.' The answer was 'No.'

'I think you have admitted that you did not act during that time upon purely banking considerations, but that you had public considerations in view?' 'Yes.'

'You admit that the course which the relative position of the Bank took during that period is not one strictly in accordance with general banking rules?' 'Yes.'

(3) The recurring financial crises that this strategy entailed made this approach increasingly undesirable for the Bank, both because it implied large economic risks for the Bank, and because it undermined public support for the Bank. (4) The consequences of this political and economic disequilibrium produced by recurring banking crises required a combination of changes to restore economic and political equilibrium, including a diminution of the Bank's monopoly position, and changes in its discounting and lending practices.

The public debate that ensued after the Panic of 1857 tried to come to grips with the political realities that had produced moral hazard from the Bank's accommodation policy. The Report of (July) 1858 exposed the dangers of that moral hazard and constituted a public acknowledgment that the existing accommodation policy, itself a creature of public expectation rather than law, was not good public policy. *Bankers' Magazine* and *The Economist* also decried the destabilizing effects of the discounting policy of the Bank (King 1936, 202-03). All of this made it politically safe for the Bank to change policy, and its Court of Directors passed the following resolution of March 13, 1858:

That habitual advances by Discount or Loan to Bill Brokers, Discount Companies and Money Dealers being calculated to lead them to rely on the assistance of the Bank of England for their security in time of pressure; Advances to Bill Brokers, Discount Companies and Money Dealers shall be confined to Loans made at the period of the Quarterly advances, or to Loans made under special and urgent circumstances which shall be communicated by the Governors at the earliest opportunity to the Court for its approval (Hughes 1960, 305).

The put option was cancelled. But the true test of that cancellation came in the Overend, Gurney crisis of 1866, when the Bank proved in its response to the crisis that this new policy would be followed. The announced change in lending policy, and the willingness to allow Overend, Gurney to fail in 1866, credibly established the end of moral hazard.

Facilitating that change was a parallel change in the structure of the banking system – specifically, new bank entry and branching – which encouraged greater competition.

Competition reduced the implicit value of the privileges the Bank enjoyed, which made the Bank less beholden to the state. Also, bank entry and branching reduced the demand for bill discounting since the discount market had evolved largely to fill gaps in a banking system that was not integrated on a national basis. In 1836, the 61 registered joint-stock banks operated 472 banking facilities; by 1870, 111 joint stock banks operated 1,127 banking facilities, and for Britain as a whole, 378 banks operated 2,738 banking facilities (Capie and Webber 1985, 576).

(5) That evolution restored stability to the banking system by removing the primary source of moral hazard (Bank discounting), fostering greater market competition and discipline, and encouraging the development of modern central bank practices, including Bagehot's rule, and incentive-compatible risk sharing between private London banks and the Bank in response to the threats of panics (like the Barings Crisis of 1890).

Once the Bank was freed from politically induced moral hazard, it was able to develop modern central banking practices, employing the discount rate and open market operations as instruments of policy, and applying rules for central bank intervention during times of financial stringency, which it did successfully to prevent full-blown panics from occurring after 1866.

The Bank developed an innovative approach to assistance in the form of a loss sharing arrangement with the other London banks during the Barings Crisis of 1890.¹⁸ Barings'

¹⁸ To say that the Bank was innovative is not to say that it was uniquely innovative, or that it was the first to use this sort of technique. White (2009) shows that the Banque de France used a similar two-tiered risky sharing technique in its coordination of assistance (with French banks) for the Paris Bourse in 1882. In 1908, Jose Limantour, the finance minister of Mexico, used a guarantee approach to assist Mexican banks to float debt backed by bank loans in the wake of the Panic of 1907 (Conant 1910).

difficulties created an asymmetric-information problem for the London clearing banks, and attendant liquidity risks. Barings' losses in the Argentine bond market, and its uncertain financial links to individual clearing banks in London, created unknown risks (to depositors) in the London clearing banks, and those unknown risks created liquidity risk for all the London banks, who faced the prospect of deposit withdrawals resulting from their credit risk uncertainty. The ingenious solution devised by the Bank and the clearing banks was an insurance arrangement provided by the London banks and underwritten by the Bank of England. The London banks guaranteed Barings, and the Bank of England effectively underwrote the group's guarantee. So long as the participating banks' guarantee were adequate to meet any losses, the Bank of England would face no exposure to loss.

Given that it was apparent to depositors that Barings' losses almost certainly would not create losses in excess of the fund created by the group of banks, this arrangement resolved the asymmetric information problems facing individual depositors in their banks, and removed their incentive to run. The Bank of England's role was to provide a belt on top of the suspenders of mutual insurance within the group, and it did so with little possibility of loss, particularly as the group members paid in 17 million pounds to establish a guarantee fund in support of their endeavor (Clapham 1944, 326-39, Andreades 1966, 366-67).

Interestingly, like Bagehot's rule – which holds that loans from the Bank should be collateralized by good assets and provided at a higher than normal interest rate – the Barings intervention reflects the philosophy that the lender of last resort (or guarantor of last resort, in the Barings case), should take a senior position relative to other banks when assisting the financial system. By lending against good collateral, a central bank only suffers losses to the

extent that the borrowing institution fails and its collateral declines in value, which places the lender of last resort in a senior position. Similarly, by requiring the London clearing banks to establish a guarantee fund for Barings, the Bank assured that they would suffer the first tier of any losses that occurred during the crisis. These incentive-compatible arrangements minimized moral-hazard problems from central bank assistance to the banking system.

V. The 2007-2009 Crisis and Historical Lessons for Reform

As discussed in detail in Calomiris (2009a), 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 2003–2007. 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. Four categories of policy distortions were most important in producing that result.

1. Lax monetary policy, especially from 2002 through 2005, promoted easy credit and kept interest rates low for a protracted period. The history of postwar monetary policy has seen only two episodes in which the real federal funds rate remained negative for several consecutive years: the high-inflation episode of 1975–1978 (which was reversed by the rate hikes of 1979–1982) and the accommodative period of 2002–2005. The Fed deviated sharply from the “Taylor Rule” in setting interest rates during 2002–2005; the federal funds rates remained substantially and persistently below levels that would have been consistent with that

rule. Not only were short-term real rates held at persistent historic lows, but unusually high demand for longer term Treasuries related to global imbalances and Asian absorption of U.S. Treasuries flattened the Treasury yield curve during the 2002–2005 period, resulting in extremely low interest rates across the yield curve. Accommodative monetary policy and a flat yield curve meant that credit was excessively available to support expansion in the housing market at abnormally low interest rates, which encouraged the overpricing of houses and subprime mortgages.

2. Numerous housing policies promoted subprime risk taking by financial institutions by effectively subsidizing the inexpensive use of leveraged finance in housing (Calomiris 2009a, 2009b). Those policies included (a) 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, (b) lending subsidies for housing finance via the Federal Home Loan Bank System to its member institutions, (c) Federal Housing Administration (FHA) subsidization of extremely high mortgage leverage and risk, (d) 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, and—almost unbelievably—(e) 2006 legislation that encouraged ratings agencies to relax standards for subprime securitizations.

All these policies encouraged the underestimation of subprime risk, but the behavior of members of Congress toward Fannie Mae and Freddie Mac, which encouraged reckless lending by the GSEs in the name of affordable housing, were arguably the most damaging actions leading up to the crisis. 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 (Calomiris 2008, Wallison and Calomiris 2009). 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 (Pinto 2008).

3. Government regulations limiting the concentration of stock ownership and the identity of who can buy controlling interests in banks have made effective corporate governance within large banks extremely challenging. Lax corporate governance allowed *some* bank management (for example, at Citibank, UBS, Merrill, Lehman, and Bear, but *not* at Bank of America, JPMorgan Chase, Goldman, Morgan Stanley, and DeutscheBank) to pursue subprime investments aggressively, even though they were unprofitable for stockholders in the long run. When stockholder discipline is absent, managers can set up the management of risk to benefit themselves at the expense of stockholders. An asset bubble (like the subprime bubble of 2003–

2007) offers an ideal opportunity; if senior managers establish compensation systems that reward subordinates based on total assets managed or total revenues collected, without regard to risk or future potential loss, then subordinates have the incentive to expand portfolios rapidly during the bubble without regard to risk. Senior managers then reward themselves for having overseen “successful” expansion with large short-term bonuses and cash out their stock options quickly so that a large portion of their money is invested elsewhere when the bubble bursts.¹⁹

4. The prudential regulation of commercial banks and investment banks has proven to be ineffective. That failure reflects (a) fundamental problems in measuring bank risk resulting from regulation’s ill-considered reliance on inaccurate rules of thumb, credit rating agencies’ assessments, and internal bank models to measure risk, and (b) the too-big-to-fail problem (Stern and Feldman 2004), which makes it difficult to credibly enforce effective discipline on large, complex financial institutions (such as Citibank, Bear Stearns, AIG, and Lehman) even if regulators detect large losses or imprudently large risks.

The risk measurement problem has been the primary failure of banking regulation and a subject of constant academic criticism for more than two decades. Regulators use different means to assess risk, depending on the size of the bank. Under the simplest version of regulatory measurement of risk, subprime mortgages (like all mortgages) have a low asset risk weight (50 percent) relative to commercial loans, although they are riskier than those loans.

¹⁹ Although it is true that many bank CEOs lost huge amounts as their firms’ stock prices plummeted, that does not exculpate them from having purposely taken on excessive risk, for two reasons: first, they may have reasonably expected a less extreme collapse than the one that occurred, and second, the game was very profitable for them while it lasted, and may have been worth playing, ex ante, even if they had anticipated that it would eventually end badly.

More complex measurements of risk (applicable to larger U.S. banks) rely on the opinions of ratings agencies or the internal assessments of banks, neither of which is independent of bank management.

Rating agencies, after all, cater to buy-side market participants (i.e., banks, pensions, mutual funds, and insurance companies that maintained subprime-related asset exposures). When ratings are used for regulatory purposes, buy-side participants reward rating agencies for underestimating risk because that helps the buy-side clients reduce the costs associated with regulation. Many observers wrongly believe that the problem with rating agency inflation of securitized debts is that sellers (sponsors of securitizations) pay for the ratings; on the contrary, the problem is that the buyers of the debts want inflated ratings because of the regulatory benefits they receive from such ratings.

The too-big-to-fail problem involves the lack of credible regulatory discipline for large, complex banks. The prospect of their failing is considered so potentially disruptive that regulators have an incentive to avoid intervention. That ex post “forbearance” makes it hard to ensure compliance ex ante. The too-big-to-fail problem magnifies incentives to take excessive risks; banks that expect to be protected by deposit insurance, Fed lending, and Treasury-Fed bailouts and believe that they are beyond discipline will tend to take on excessive risk because taxpayers share the downside costs.

The too-big-to-fail problem was clearly visible in the behavior of large investment banks in 2008. After Bear Stearns was rescued in March, Lehman, Merrill Lynch, Morgan Stanley, and Goldman Sachs sat on their hands for six months awaiting further developments (i.e., either an improvement in the market environment or a handout from Uncle Sam). In particular, Lehman

did little to raise capital or shore up its position. But when conditions deteriorated and the anticipated bailout failed to materialize for Lehman in September 2008 (showing that there were limits to Treasury-Fed generosity), the other major investment banks immediately were either acquired or transformed themselves into bank holding companies to increase their access to government support.

19th Century Britain Redux?

The mid-19th century British discussions of financial reform share important features with the current debates over prudential reforms in the U.S. Many aspects of the current debate would seem familiar to 19th century British observers. Public resentment over the abuse of special privileges by mortgage monopolists, Fannie Mae and Freddie Mac, who fueled the subprime bubble, and whose internal emails (Calomiris 2008) show that they did so largely to preserve the special privileges conferred upon them by the government, 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 U.S. 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 19th century in Britain. Just as the debate over 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 19th century to the current environment (for perspectives on the reform agenda, see Calomiris 2009a, 2009b, and 2009c). The important point to emphasize here is a consistent theme of the historical record: the ability to improve the financial system depends on the political environment.

The favorable outcome in Britain in the 19th century resulted from a political consensus 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. Reform in reaction to crisis, however, is not always so successful. Despite the advantages of creating a properly constituted central bank with a predictable and well-defined role, and with the tools necessary to execute that role, the U.S. opted to cancel the charters of its first two central banks, the BUS and the SBUS, as the result of their lack of popularity, and in the case of the SBUS that was directly related to exaggerated public perceptions that it had acted badly during the crises of 1819 and 1825. In 1933, in the U.S., public anger over the Depression was channeled by politicians into undesirable and ineffective "reforms" of the banking system, including separation of investment and commercial banking, the limitation of interest payments on deposits, and the creation of deposit insurance (Calomiris 2000, 2009a). Those political decisions had long-lasting consequences; it took more than five decades of discourse, accumulation of historical evidence

and new experience to repeal Regulation Q and the separation of securities underwriting from commercial banking; deposit insurance – the last vestige of New Deal policy, and one that had no intellectual support at the time of its passage – seems destined to remain forever.

VI. Conclusion

Banking crises properly defined consist either of panics or waves of costly bank failures. These phenomena were rare historically compared to the present. A historical analysis of the two phenomena (panics and waves of failures) reveals that they do not always coincide, are not random events, cannot be seen as the inevitable result of human nature or the liquidity transforming structure of bank balance sheets, and do not typically accompany business cycles or monetary policy errors. Rather, risk-inviting microeconomic rules of the banking game that are established by government have always been the key additional necessary condition to producing a propensity for 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 U.S., Argentina's government guarantees for mortgages in the 1880s, Australia's government subsidization of real estate development prior to 1893, the Bank of England's discounting of paper at low interest rates prior to 1858, and the expansion of government-sponsored deposit insurance and other bank safety net programs throughout the world in the past three decades, including the generous government subsidization of subprime mortgage risk taking in the U.S. leading up to the recent crisis.

Other risk-inviting rules historically have involved government-imposed structural constraints on banks, which include entry restrictions like unit banking laws that constrain competition, prevent diversification of risk, and limit the ability to deal with shocks. The most important example of these structural constraints was the U.S. historical system of unit banking, which limited competition and diversification of loan risk by preventing branching, and by effectively preventing collective action by banks in the management of crises once adverse shocks had hit.

Finally, another destabilizing 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. That absence contributed to instability in the U.S. prior to 1913.

Panics, whether associated with waves of bank failure or not, have been times of temporary confusion (due to asymmetric information) about the incidence of shocks within the banking system. This asymmetric-information problem was particularly acute in the U.S. Indeed, in the late nineteenth and early twentieth centuries; system-wide banking panics like those that the U.S. experienced in that period generally did not occur elsewhere. The uniquely panic-ridden experience of the U.S., particularly during the pre-World War I era, reflected a combination of the unit banking structure of the U.S. system and the absence of a proper lender of last resort. Panics were generally avoided by other countries in the pre-World War I era because their banking systems were composed of a much smaller number of banks operating on a national basis (who consequently enjoyed greater diversification, and a greater

ability to coordinate their actions to stem panics ex post), and because they had developed incentive-compatible principles for central bank lending.²⁰

The U.S. and other countries also experienced waves of bank failures unrelated to panics (most notably in the U.S. in the 1920s), which reflected the vulnerability of banks to sector-specific shocks (e.g., agricultural price declines) in an undiversified banking system. In the U.S. , waves of bank failures in the 1920s were aggravated not only by the absence of branch banking but by the presence of deposit insurance in various states, which promoted moral hazard in lending and adverse selection in bank entry and which resulted in particularly severe failure experiences. In the U.S. during the antebellum period, particularly in the South, even worse failure outcomes were related to government-directed credit policies implemented through political control of banks (Schweikart 1987); Australia in 1893 suffered a similar fate.

More recent banking system experience worldwide indicates a dramatic upward shift in the costs of banking system distress – an unprecedented high frequency of banking crises, many bank failures during crises, and large losses by failing banks, sometimes with disastrous consequences for taxpayers who end up footing the bill of bank loss. This pandemic of bank failures has been traced empirically to the expanded role of the government safety net, as well as government involvement in directed credit. Government protection of banks and government direction of credit flows has encouraged excessive risk taking by banks and created greater tolerance for incompetent risk management (as distinct from purposeful increases in

²⁰ Of course, it was possible to have a stable banking system without a central bank, as was the case in the Canadian branch banking system, which did not establish a central bank to act as a lender of last resort until 1935.

risk). The government safety net, which was designed to forestall the (overestimated) risks of contagion, ironically has become the primary source of systemic instability in banking.

The desirable path for reform is best illustrated by the successful adaptation that occurred in the British banking system during the second half of the 19th century, when it overcame its legacy of moral hazard and its high propensity for banking crises by dramatically changing the rules of the banking game to eliminate the put option in the London bill market that had been provided by the Bank of England. This transformed the Bank into a instrument of systemic stability, and encouraged greater competition in banking. Those reforms set the stage for the development by the Bank of improved mechanisms for limiting liquidity risk in the system, including Bagehot's rule and other means of incentive-compatible sharing of risk between the Bank and the coalition of private London banks.

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 favorable reforms, as it did in Britain in the 19th century, or unfavorable reforms as the result of populist misapprehension, as in the case of the disappearance of the BUS and SBUS, or the capture of financial reform by special interests, as was the case in the U.S. in 1933.

References

- Alston, Lee J., Wayne A. Grove, and David C. Wheelock (1994). "Why Do Banks Fail? Evidence from the 1920s," *Explorations in Economic History* 30, 409-31.
- Andreades, A. (1966). *History of the Bank of England, 1640 to 1903*, New York: Augustus Kelley.
- Bagehot, Walter (1873). *Lombard Street: A Description of the Money Market*, Reprinted by Richard D. Irwin, Homewood, Illinois, 1962.
- Barth, James, Gerard Caprio, Jr., and Ross Levine (2006), *Rethinking Bank Regulation: Till Angels Govern*, Cambridge University Press.
- Bernanke, Ben S. (1983). "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review* 73, June, 257-76.
- Bernanke, Ben S. and James, Harold (1991). "The Gold Standard, Deflation, and Financial Crisis in the Great Depression: An International Comparison." In R. Glenn Hubbard, ed., *Financial Markets and Financial Crises*. University of Chicago Press, 33-68.
- Bernstein, Asaf, Eric Hughson, and Marc D. Weidenmier (2009a). "Quasi-Central Banking and Liquidity Risk: Lessons from the Second Bank of the US," Working Paper.
- Bernstein, Asaf, Eric Hughson, and Marc D. Weidenmier (2009b). "Identifying the Effects of a Lender of Last Resort on Financial Markets: Lessons from the Founding of the Fed," *Journal of Financial Economics*, forthcoming.
- Board of Governors of the Federal Reserve System (1999). "Using Subordinated Debt as an Instrument of Market Discipline," Staff Study 172, December.
- Bordo, Michael (1985). "The Impact and International Transmission of Financial Crises: Some Historical Evidence, 1870-1933," *Revista di Storia Economica*, 2d ser., v. 2, 41-78.
- Bordo, Michael D. (2007). "The Crisis of 2007: The Same Old Story, Only the Players Have Changed," Working Paper, Rutgers University.
- Bordo, Michael D., and Barry Eichengreen (2003). "Crises Now and Then: What Lessons from the Last Era of Financial Globalization?" in *Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart*, Volume 2, Edward Elgar, 53-84.

Bordo, Michael D., and David Wheelock (2007). "Stock Market Booms and Monetary Policy in the Twentieth Century," *Review*, Federal Reserve Bank of St. Louis, 89 (March/April).

Bordo, Michael D., and David Wheelock (2009). "When Do Stock Market Booms Occur: The Macroeconomic and Policy Environments of 20th Century Booms," in Jeremy Atack and Larry Neal, editors, *The Origins and Development of Financial Markets and Institutions*, Cambridge: Cambridge University Press, 416-49.

Boyd, John, Pedro Gomis, Sungkyu Kwak, and Bruce Smith (2000). "A User's Guide to Banking Crises." Conference Paper, The World Bank.

Bruner, Robert F., and Sean D. Carr (2007). *The Panic of 1907: Lessons Learned from the Market's Perfect Storm*, Wiley.

Brunner, Karl, and Allan H. Meltzer (1968). "What Did We Learn from the Monetary Experience of the United States in the Great Depression?" *Canadian Journal of Economics* 1, May, 334-48.

Butlin, Matthew W. (1987). "Capital Markets," in *The Australian Economy in the Long Run*, edited by Rodney Maddock and Ian W. Mclean, Cambridge University Press.

Butlin, Noel G. (1964). *Investment in Australian Economic Development, 1861-1914*, Cambridge: Cambridge University Press.

Butlin, Sydney J. (1961). *Australia and New Zealand Bank: The Bank of Australasia and the Union Bank of Australia Limited, 1828-1951*, London: Longmans.

Calomiris, Charles W. (1989). "Deposit Insurance: Lessons from the Record," *Economic Perspectives*, Federal Reserve Bank of Chicago, May/June, 10-30.

Calomiris, Charles W. (1990). "Is Deposit Insurance Necessary? A Historical Perspective," *Journal of Economic History*, 50, 283-95.

Calomiris, Charles W. (1992). "Do Vulnerable Economies Need Deposit Insurance? Lessons from U.S. Agriculture in the 1920s." In Philip L. Brock, ed., *If Texas Were Chile: A Primer on Bank Regulation*. San Francisco: The Sequoia Institute, 237-349, 450-458.

Calomiris, Charles W. (2000). *U.S. Bank Deregulation in Historical Perspective*, Cambridge University Press.

Calomiris, Charles W. (2007). "Victorian Perspectives on the Banking Distress of the Late 20th Century," Working paper.

Calomiris, Charles W. (2008). "Statement before the Committee on Oversight and Government Reform, United States House of Representatives," December 9.

Calomiris, Charles W. (2009a). "The Subprime Turmoil: What's Old, What's New, and What's Next," *Journal of Structured Finance* 15, Spring, 6-52.

Calomiris, Charles W. (2009b). "Financial Innovation, Regulation, and Reform," *Cato Journal* 29, Winter, 65-92.

Calomiris, Charles W. (2009c). "Prudential Bank Regulation: What's Broke and How to Fix It." In Terry L. Anderson and Richard Soussa, eds., *Reacting to the Spending Spree: Policy Changes We Can Afford*. Hoover Institution Press.

Calomiris, Charles W., and Gary Gorton (1991). "The Origins of Banking Panics: Models, Facts, and Bank Regulation," in R. Glenn Hubbard, ed., *Financial Markets and Financial Crises*, University of Chicago, 107-73.

Calomiris, Charles W., and R. Glenn Hubbard (1989). "Price Flexibility, Credit Availability, and Economic Fluctuations: Evidence from the U.S., 1894-1909," *Quarterly Journal of Economics*, August, 429-52.

Calomiris, Charles W., and Charles M. Kahn (1991). "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review* 81, 497-513.

Calomiris, Charles W., and Joseph R. Mason (1997). "Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic," *American Economic Review* 87, 863-83.

Calomiris, Charles W., and Joseph R. Mason (2003a). "Fundamentals, Panics and Bank Distress During the Depression," (with Joseph Mason) *American Economic Review* 93, 1615-47.

Calomiris, Charles W., and Joseph R. Mason (2003b). "Consequences of Bank Distress During the Great Depression," *American Economic Review* 93, 937-47.

Calomiris, Charles W., and Andrew Powell (2001). "Can Emerging Market Bank Regulators Establish Credible Discipline: The Case of Argentina, 1992-99," in *Prudential Supervision: What Works and What Doesn't*, edited by Frederic S. Mishkin, 147-96.

Calomiris, Charles W., and Larry Schweikart (1991). "The Panic of 1857: Origins, Transmission, and Containment," *Journal of Economic History*, 51, 807-34.

Calomiris, Charles W., and Eugene N. White (1994). "The Origins of Federal Deposit Insurance," in Claudia Goldin and Gary Libecap, eds., *The Regulated Economy: A Historical Approach to Political Economy*, University of Chicago, 145-88.

Calomiris, Charles W., and Berry Wilson (2004). "Bank Capital and Portfolio Management: The 1930s 'Capital Crunch' and Scramble to Shed Risk," *Journal of Business* 77, 421-55.

Canovai, Tito (1911). *The Banks of Issue in Italy*, National Monetary Commission, U.S. Senate, 61st Congress, 2d Session, Document No. 575.

Capie, Forrest (2002). "The Emergence of the Bank of England as a Mature Central Bank. In *The Political Economy of British Historical Experience, 1688-1914*, Donald Winch and Patrick O'Brien, eds. Oxford: Oxford University Press.

Capie, Forrest (2009). "Financial Crises in England in the Nineteenth and Twentieth Centuries," Working Paper, June.

Capie, Forrest, and Alan Webber (1985). *A Monetary History of the United Kingdom, 1870-1982*, London: George Allen and Unwin.

Caprio, Gerard, and Daniela Klingebiel (1996). "Bank Insolvencies: Cross Country Experience." Working Paper No. 1620, The World Bank.

Catterall, Ralph C. H. (1902). *The Second Bank of the United States*, Chicago: University of Chicago Press.

Clapham, Sir John (1944). *The Bank of England*. Cambridge: Cambridge University Press.

Conant, Charles Arthur (1910). *The Banking System of Mexico*, Washington D.C.: National Monetary Commission, 61st Congress, 2nd Session.

Davis, Joseph, Christopher Hanes, and Paul Rhode (2007). "Harvests and Business Cycles in Nineteenth Century America," Working Paper.

Davis, Lance E., and Robert E. Gallman (2001). *Evolving Financial Markets and International Capital Flows: Britain, the Americas, and Australia, 1865-1914*, Cambridge: Cambridge University Press.

Davis, William Stearns (1910). *The Influence of Wealth in Imperial Rome*, New York: Macmillan.

Demirguc-Kunt, Asli, and Enrica Detragiache (2000). "Does Deposit Insurance Increase Banking System Stability?" Conference Paper, The World Bank.

Demirguc-Kunt, Asli, Edward Kane, and Luc Laeven, eds.(2009). *Deposit Insurance Around the World* (Cambridge, Mass.: MIT Press).

Diamond, Douglas, and Philip Dybvig (1983). "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* 91, 401-19.

Friedman, Milton, and Anna J. Schwartz (1963). *A Monetary History of the United States, 1867-1960*, Princeton University Press.

Goodhart, Charles A.E. (1995). *The Evolution of Central Banks*, Cambridge MA: MIT Press.

Gorton, Gary (1985). "Clearing Houses and the Origin of Central Banking in the United States," *Journal of Economic History* 45, 277-83.

Gouge, William A. (1833). *An Inquiry into the Principles of the American Banking System*, Philadelphia.

Grossman, Richard S. (1994). "The Shoe That Didn't Drop: Explaining Banking Stability During the Great Depression." *Journal of Economic History*, 54, 654-82.

Hawtrey, R.G. (1932). *The Art of Central Banking*, London: Longmans, Green.

Hawtrey, R.G. (1938). *A Century of Bank Rate*, London: Longmans, Green.

Hilt, Eric (2009). "Wall Street's First Corporate Governance Crisis: The Panic of 1826," NBER Working Paper No. 14892, April.

Hughes, Jonathan R. T. (1960). *Fluctuations in Trade, Industry and Finance: A Study of British Economic Development, 1850-1860*. Oxford: Oxford University Press.

Kindleberger, Charles P. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*, New York: Basic Books.

King, W. T. C. (1936). *History of the London Discount Market*. London: George Routledge & Sons.

Lucia, Joseph L. (1985). "The Failure of the Bank of United States: A Reappraisal," *Explorations in Economic History* 22, 402-16.

Mason, Joseph R. (2001). "Do Lender of Last Resort Policies Matter? The Effects of Reconstruction Finance Corporation Assistance to Banks During the Great Depression," *Journal of Financial Services Research*, September 20, 77-95.

- Minsky, Hyman P. (1975). *John Maynard Keynes*, Columbia University Press.
- Miron, Jeffrey (1986). "Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the Fed," *American Economic Review* 76, 125-140.
- Mishkin, Frederic S., (Editor) (2001). *Prudential Supervision: What Works and What Doesn't*, University of Chicago Press.
- Mitchener, Kris James (2005). "Bank Supervision, Regulation, and Instability During the Great Depression," *Journal of Economic History* 65, 152-85.
- Pinto, Edward J. (2008). "Statement before the Committee on Oversight and Government Reform, United States House of Representatives," December 9.
- Richardson, Gary, and William Troost (2006). "Monetary Intervention Mitigated Banking Panics During the Great Depression? Quasi-Experimental Evidence from the Federal Reserve District Border in Mississippi, 1929 to 1933," NBER Working Paper No. 12591.
- Rousseau, Peter (2002). "Jacksonian Monetary Policy, Specie Flows, and the Panic of 1837," *Journal of Economic History* 62, 457-88.
- Sayers, R.S. (1976). *The Bank of England, 1891-1944*, Cambridge: Cambridge University Press.
- Schuster, Sir Felix (1923). *The Bank of England and the State*, London: University of Manchester Press.
- Schweikart, Larry (1987). *Banking in the American South from the Age of Jackson to Reconstruction*, Baton Rouge: Louisiana State University Press.
- Schweikart, Larry (1988). "Jacksonian Ideology, Currency Control, and Central Banking: A Reappraisal," *The Historian* (November), 78-102.
- Shadow Financial Regulatory Committee (2000). *Reforming Bank Capital Regulation*, American Enterprise Institute, 2000.
- Sprague, Oliver M. W. (1910). *History of Crises under the National Banking System*, National Monetary Commission.
- Stern, Gary H., and Ron J. Feldman (2004). *Too Big to Fail: The Hazards of Bank Bailouts*, Brookings Institution.

Sylla, Richard, Robert E. Wright, and David J. Cowen (2009). "Alexander Hamilton, Central Banker: Crisis Management during the U.S. Financial Panic of 1792," *Business History Review* 83, Spring, 61-86.

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

Temin, Peter (1969). *The Jacksonian Economy*, W.W. Norton.

Temin, Peter (1976). *Did Monetary Forces Cause the Great Depression?* W.W. Norton.

Temin, Peter (1989). *Lessons from the Great Depression*, MIT Press.

Timberlake, Richard H. (1984). "The Central Banking Role of Clearinghouse Associations," *Journal of Money, Credit and Banking* 16 (February), 1-15.

Wallison, Peter J., and Charles W. Calomiris, (2009). "The Last Trillion-Dollar Commitment: The Destruction of Fannie Mae and Freddie Mac," *Journal of Structured Finance* 15, Spring, 71-80.

Wheelock, David C. (1991). *The Strategy and Consistency of Federal Reserve Monetary Policy, 1924-1933*, Cambridge: Cambridge University Press.

White, Eugene N. (1984). "A Reinterpretation of the Banking Crisis of 1930," *Journal of Economic History*, 44, 119-38.

White, Eugene N. (2009). "The Banque de France as a Lender of Last Resort in the Nineteenth Century," Working Paper, Rutgers University.

Wicker, Elmus (1966). *Federal Reserve Monetary Policy, 1917-1933*, New York: Random House.

Wicker, Elmus (1980). "A Reconsideration of the Causes of the Banking Panic of 1930." *Journal of Economic History*, 40, 571-83.

Wicker, Elmus (1996). *The Banking Panics of the Great Depression*. Cambridge University Press.

Wigmore, Barrie A. (1987). "Was the Bank Holiday of 1933 a Run on the Dollar Rather than the Banks? *Journal of Economic History* 47, 739-56.