NBER WORKING PAPER SERIES

BANK FAILURES IN THEORY AND HISTORY: THE GREAT DEPRESSION AND OTHER "CONTAGIOUS" EVENTS

Charles W. Calomiris

Working Paper 13597
http://www.nber.org/papers/w13597

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2007

This paper was prepared for the Oxford Handbook of Banking, edited by Allen Berger, Phil Molyneux, and John Wilson. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Bank Failures in Theory and History: The Great Depression and Other "Contagious" Events
Charles W. Calomiris
NBER Working Paper No. 13597
November 2007
JEL No. E5,G2,N2

ABSTRACT

Bank failures during banking crises, in theory, can result either from unwarranted depositor withdrawals during events characterized by contagion or panic, or as the result of fundamental bank insolvency. Various views of contagion are described and compared to historical evidence from banking crises, with special emphasis on the U.S. experience during and prior to the Great Depression. Panics or "contagion" played a small role in bank failure, during or before the Great Depression-era distress. Ironically, the government safety net, which was designed to forestall the (overestimated) risks of contagion, seems to have become the primary source of systemic instability in banking in the current era.

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

"Contagion" vs. Fundamentals as Causes of Bank Failures

Concerns about the susceptibility of banks to unwarranted withdrawals of deposits during panics, the possibility of bank failures and contractions of bank credit resulting from unwarranted withdrawals of deposits (which is sometimes described as the result of "contagious" weakness among banks), and the attendant adverse macroeconomic consequences of bank disappearance or bank balance sheet contraction have motivated much of the public policies toward banks. Those policies include assistance mechanisms intended to protect banks from unwarranted withdrawals of deposits (central bank lending during crises, deposit insurance, and government-sponsored bank bailouts), and a host of prudential regulatory policies (intended to promote banking system stability, and especially to prevent banks from taking advantage of government protection by increasing their riskiness – the so-called "moral-hazard" problem of protection).

Theoretical models have been devised in which banking crises result from systemic "contagion," when banks that are intrinsically solvent are subjected to large unwarranted withdrawals, and may fail as a consequence of this withdrawal pressure. Advocates of the view that banking systems are inherently vulnerable to such contagion often emphasize that the structure of banks – the financing of illiquid assets with demandable debts, and the "sequential service constraint" (which mandates that depositors who are first in line receive all of their deposits) – tends to aggravate the tendency for unwarranted withdrawals (see Douglas W. Diamond and Phillip H. Dybvig 1983, Franklin Allen and Douglas Gale 2000, Douglas W. Diamond and Raghuram Rajan 2002).

Unwarranted withdrawals (that is, those unrelated to the solvency of the bank) can occur, in theory, for a number of reasons. Diamond and Dybvig (1983) develop a banking model with multiple equilibria, where one of the equilibria is a systemic bank run, which occurs simply because depositors believe that others will run. More generally, observers of historical panics sometimes document depositors imitating each other's withdrawal behavior; depositors may line up to withdraw their funds simply because others are doing so, particularly in light of the incentives implied by the sequential service constraint. It is important to recognize, however, that evidence about mimetic withdrawals does not generally confirm the all-or-nothing runs by all depositors imagined by some theoretical models; rather, mimesis may be partial and gradual (see O'Grada and White 2003, and Bruner and Carr 2007).

A second possibility, which is particularly relevant for understanding pre-World War I banking panics in the U.S. (e.g., the nationwide U.S. Panics of 1857, 1873, 1884, 1890, 1893, 1907, and some events during the Great Depression, including the Chicago banking panic of June 1932), is that a signal is received by depositors, which contains noisy information about the health of the various banks. Depositors have reason to believe that a loss has occurred that might cause a bank to become insolvent, but they cannot observe which bank has suffered the loss. In that circumstance, depositors may withdraw large amounts of funds from all banks, including those that are (unobservably) solvent, simply because they would rather not risk leaving their money in a bank that turns out to be insolvent.

Third, exogenous shocks to depositors' liquidity preferences, or to the supply of reserves in the banking system, unrelated to banks' asset condition, may cause an

excess demand for cash on the part of depositors relative to existing reserves, which can lead banks to a scramble for reserves, which can produce systemic runs (a banking version of the game “musical chairs”). Liquidity demand and supply shocks may be related to government policies affecting the reserve market, or to foreign exchange risks that lead depositors to want to convert to cash. This mechanism may have had a role in some banking system crises (notably, the nationwide U.S. Panics of 1837 and 1933).

Withdrawal pressures, whether they are associated with warranted or unwarranted withdrawals, can accumulate over time or can take the extreme form of a “bank run” (when depositors decide en masse to remove deposits). Some financial historians (notably Milton Friedman and Anna J. Schwartz 1963) have pointed to the Great Depression of the 1930s as a time when unwarranted depositor withdrawals, and sometimes “runs” or “panics,” led to large numbers of bank failures, and rapid declines in deposits of solvent and insolvent banks alike. Bank distress is associated not only with bank failures, but with general macroeconomic consequences resulting from the reduced supply of loans and deposits, which can amplify business cycle downturns and spread panic-induced financial distress from banks to the whole economy (Ben Bernanke 1983, Charles Calomiris and Joseph Mason 2003b). Other episodes of banking panics outside the Great Depression have also been identified as possible episodes of unwarranted bank failures, especially in the United States during the nineteenth and early twentieth centuries, with similar inferences drawn by some about the contagious causes and costly consequences of bank distress.

Another view of banking distress (which I will label the "fundamentalist," as opposed to the "panic," approach), stresses a different direction of causality: a chain of causation from non-panic-related, observable, exogenous adverse changes in the economic conditions of banks, to intrinsic weakening of bank condition, ultimately leading to bank failure. According to this view, fundamental losses to bank borrowers cause losses to banks, which may bankrupt some banks and lead other weakened banks to curtail the supplies of loans and deposits as part of a rebalancing of portfolios to limit default risk in a disciplined market (Calomiris and Wilson 2004). Endogenous contractions of deposits and loans, just like unwarranted contractions, will limit the supply of money and credit, and thus they will exacerbate the macroeconomic decline that caused them. Thus, according to the fundamentalist view, banking distress can magnify economic downturns even if banks are not the originators of shocks; banks will tend to magnify macroeconomic shocks through their prudential decisions to curtail the supplies of loans and deposits in response to adverse shocks, even if banks are passive responders to shocks and even if depositors avoid engaging in unwarranted runs or panics.

Differences in opinion about the sources of shocks that cause bank failures have important implications for policy. While both the panic and fundamentalist views can be used to motivate public policy to protect banks (since both views see banks as important magnifiers of macroeconomic disturbance), the panic view provides special motives for public policies to protect banks from withdrawal risk. The fundamentalist view, in contrast, sees banks as inherently stable – that is, neither victims of unwarranted withdrawals, nor a major source of macroeconomic shocks.

According to the fundamentalist view, market discipline of banks is not random, and indeed, helps preserve efficiency in the banking system. It may be desirable to limit or even avoid government protection of banks to preserve market discipline in banking (making banks more vulnerable to the risk of depositor withdrawal). Preserving market discipline encourages good risk management by banks, even though bank deposit and credit contractions attendant to adverse economic shocks to bank borrowers may aggravate business cycles. Indeed, some empirical studies have argued that policies that insulate banks from market discipline tend to produce worse magnifications of downturns, due to excessive bank risk taking in response to protection (for example, John Boyd, Pedro Gomis, Sungkyu Kwak and Bruce Smith 2000, and James Barth, Gerard Caprio, and Ross Levine 2006).

These two views of the sources of bank distress (the panic view that banks are fragile and highly subject to panic, or alternatively, the fundamentalist view that banks are stable and generally not subject to unwarranted large-scale withdrawals) do not define the universe of possibilities. One or the other extreme view may do a better job explaining different historical crises, and both fundamentals and unwarranted withdrawals may play a role during some banking crises. The recent empirical literature on banking crises has tried to come to grips with the causes and effects of systemic bank failures in different places and times, to ascertain the dominant causal connections relating banking distress and macroeconomic decline, and to try to draw inferences about the appropriate public policy posture toward banks. The remainder of this chapter selectively reviews the empirical literature on the causes of bank failures during systemic banking crises. This review begins with a

lengthy discussion of the Great Depression in the United States, which is followed by a discussion of U.S. bank distress prior to the Depression, historical bank distress outsides the United States, and contemporary banking system distress (which is discussed more fully in Chapter 26 of this volume, by Gerard Caprio and Patrick Honohan).

U.S. Bank Distress during the Great Depression

The list of fundamental shocks that may have 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. There is no doubt that adverse fundamental shocks relevant to bank solvency were contributors to bank distress; the controversy is over the size of these fundamental shocks – that is, whether banks experiencing distress were truly insolvent or simply illiquid.

Friedman and Schwartz (1963) are the most prominent advocates of the view that many bank failures resulted from unwarranted "panic" and that failing banks were in large measure illiquid rather than insolvent. 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 during the panics that they identify, 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 Peter Temin (1976) and many others have noted, the bank failures during the Depression marked a continuation of the severe banking sector distress that had gripped agricultural regions throughout the 1920s. Of the nearly 15,000 bank disappearances that occurred 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). Elmus 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.

But how can one attribute bank failures during the Depression mainly to fundamentals when Friedman and Schwartz's time series evidence indicates no prior changes in macroeconomic fundamentals? 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).

Using a narrative approach similar to that of Friedman and Schwartz, but relying on data disaggregated to the level of the Federal Reserve districts and on local newspaper accounts of banking distress, Wicker argues that it is 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.

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. Banks in the United States (unlike banks in other countries) did not operate throughout the country. They were smaller, regionally isolated institutions. In the United States, therefore, 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 Ben S. Bernanke and Harold James 1991, and Richard S. 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.

Microeconomic studies of banking distress have provided some useful evidence on the reactions of individual banks to economic distress. Eugene N. 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.

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, and in that sense, unwarranted deposit contraction did occur. 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 weak banks than in surviving banks.

Calomiris and Berry 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 (Barry Wigmore 1987). Currency risk, of course, is also a fundamental.

It is also interesting to connect this account of bank distress during the Depression – which emphasizes fundamental shocks, rather than simply illiquidity, as the source of bank distress – with the history of lender of last resort assistance to banks during the Depression. Many commentators have faulted the Federal Reserve for failing to prevent bank failures 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

problem for most banks in the 1930s that were experiencing 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). Commentators at the time noted that, because the collateralized RFC and Fed loans were senior to deposits, and because depositor withdrawals from weak banks reflected real concerns about bank insolvency, loans from the Fed and the RFC to banks experiencing withdrawals did nothing to help, and actually often did harm to 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 preferred stock of financial institutions (which was junior to depositors), RFC assistance to troubled banks was effective in reducing the risk of failure (Mason 2001).

Microeconomic Studies of Local Contagion

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 recognize that this measure of "contagious failure" is 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.

Cormac O'Grada and Eugene White (2003) provide a detailed account of depositor behavior based on individual account data during the 1850s for a single bank, the Emigrant Savings Bank of New York, which offers a unique perspective on depositor contagion during banking panics. In 1854, Emigrant experienced an unwarranted run that can be traced to mimetic behavior among inexperienced, uninformed depositors. This run, however, was easily handled by the bank, which was able to pay off depositors and restore confidence. In contrast, the run in 1857 was an imitative response to the behavior of informed, sophisticated depositors who were running for a reason, and that run resulted in suspension of convertibility. Furthermore, in both of these episodes, mimesis was not sudden: "In neither 1854 nor 1857 did depositors respond to a single signal that led them to crowd into banks all at once. Instead, panics lasted a few weeks, building and sometimes ebbing in intensity, and only a fraction of all accounts were closed" (p. 215). O'Grada and White show that contagion can be a real contributor to bank distress, but they also show that runs based on random beliefs tend to dissipate with little effect, while runs based on legitimate signals tend to grow in importance over time. The fact that runs

are not sudden, and that many depositors do not participate in them at all, is important, since it implies the ability of events to unfold over time; that is, for a form of collective learning among depositors to take place during panics.

A similar account of mimetic withdrawals based on a random rumor can be found in an article by Henry Nicholas in *Moody's Magazine* in 1907. A bank in Tarpen Springs, Florida experienced an unwarranted outflow of deposits based on a false rumor that was spread through the local Greek-American community, which included many of the bank's depositors. The bank quickly wired to have cash sent from its correspondent bank, which arrived in time to prevent any suspension of convertibility, and brought the run to an end. Nicholas noted that, if the bank had really been in trouble, not only would the correspondent not have provided the funds, but it and other banks would have probably withdrawn any funds it had on deposit at the bank long before the public was aware of the problem (a so-called "silent run;" see the related discussions in Halac and Schmukler 2004, and Stern and Feldman 2003).

U.S. Bank Distress in the Pre-Depression Era

As many scholars have recognized for many years, for structural reasons, U.S. banks were unusually vulnerable to systemic banking crises that saw large numbers of bank failures before the Depression, compared to banks in other countries (for reviews, see Michael Bordo 1985, and Calomiris 2000). Calomiris and Gary Gorton (1991) identify six episodes of particularly severe banking panics in the United States between the Civil War and World War I., and prior to the Civil

War, there were other nationwide banking crises in 1819, 1837, and 1857. In the 1920s, the U.S. experienced waves of bank failures in agricultural states, which have always been identified with fundamental shocks to banks, rather than national or regional panics. Other countries, including the U.S.'s northern neighbor, Canada, however, did not suffer banking crises during these episodes of systemic U.S. banking system distress. The key difference between the U.S. and other countries historically was the structure of the U.S. banking system. The U.S. system was mainly based on unit banking – geographically isolated single-office banks; no other country in the world imitated that approach to banking, and no other country experienced the 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.

The United States was unable to mimic this behavior on a national or regional scale (Calomiris 2000, Calomiris and Schweikart 1991). U.S. law prohibited nationwide branching, and most states prohibited or limited within-state branching. U.S. banks, in contrast to banks elsewhere, were numerous (e.g., numbering more than 29,000 in 1920), undiversified, insulated from competition, and geographically isolated from one another, 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 Gorton 1985).

The structure of U.S. banking explains why the United States uniquely suffered banking panics despite the fact that the vast majority of banks were healthy, and were able to avoid ultimate failure. Empirical studies show that the major U.S. banking panics of 1857, 1873, 1884, 1890, 1893, 1896, and 1907 were moments of heightened asymmetric information about bank risk. 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.

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, even during these panic episodes, were small, 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. Oliver 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 remained below 1%, and 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 and the 1920s were much higher than those of the other pre-Depression systemic U.S. banking crisis episodes. The 1830s 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 (Peter Rousseau 2002), unprofitable bank-financed infrastructure investments that went sour (Schweikart 1988), and international balance of payments shocks (Peter 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 in the absence of regional or national bank portfolio diversification (Calomiris 1992, Lee Alston, Wayne Grove and Davoid Wheelock 1994).

Other Historical Experiences with Bank Failures

Although the U.S. was unique in its propensity for panics, it was not the only economy to experience occasional waves of bank failures historically. Losses (i.e., the negative net worth of failed banks), however, were generally modest and bank failure rates were much lower outside the U.S. The most severe cases of banking distress during this era, Argentina in 1890 and Australia in 1893, were the exceptional cases; they 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 were 3% and in Italy in 1893 1% of GDP, but with the possible exception of Brazil (for which data do not exist to measure losses), there were no other cases in 1875-1913 in which banking losses in a country exceeded 1% of GDP (Calomiris 2007).

Loss rates 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 Charles 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' 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.

Bank Failures in the Late 20th Century

Recent research on systemic bank failures 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 occurring in 140

developing countries in the last two decades of the 20th century, all of which experienced banking system losses in excess of 1% of GDP, and more than 20 of which experienced losses in excess of 10% of GDP (data are from Caprio and Klingebiel 1996, updated in private correspondence with these authors). Empirical studies of these unprecedented losses concluded 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.

The theory behind the problem of destabilizing protection has been well-known for over a century, and was the basis for Franklin Roosevelt's opposition to deposit insurance in 1933 (an opposition shared by many). Ironically, federal deposit insurance is one of the major legacies of the Roosevelt presidency, despite the fact that President Roosevelt, the Federal Reserve, the Treasury, and Senator Carter Glass – the primary authorities on banking policy of the time – all were opposed to it on principle. Deposit insurance was seen by them and others as undesirable special interest legislation designed to benefit small banks. They acquiesced in its passage for practical reasons, to get other legislation passed, not because they wanted deposit insurance to pass per se. Numerous attempts, dating from the 1880s, to introduce federal deposit insurance legislation failed to attract support in the Congress (Calomiris and White 1994). Opponents understood the theoretical arguments against deposit insurance espoused today – that deposit insurance removes depositors' incentives to monitor and discipline banks, and 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 also promotes banker incompetence, which leads to unwitting risk taking.

Research on the banking collapses of the last two decades of the twentieth century have produced new empirical findings indicating that the greater the protection offered by a country’s bank safety net, the greater the risk of a banking collapse (see, for example, Caprio and Klingebiel 1996, Boyd et al. 2000, Demirguc-Kunt Detragiache 2000, and Barth et al. 2006). Empirical research on prudential bank regulation similarly emphasizes the importance of subjecting some bank liabilities to the risk of loss to promote discipline and limit risk taking (Shadow Financial Regulatory Committee 2000, Mishkin 2001, Barth et al. 2006).

Studies of historical deposit insurance reinforce these conclusions (Calomiris 1990). The basis for the opposition to deposit insurance in the 1930s was the disastrous experimentation with insurance in several U.S. states during the early 20th century, which resulted in banking collapses in all the states that adopted insurance. Government protection of banks had played a similarly destabilizing role in Argentina in the 1880s (leading to the 1890 collapse) and in Italy (leading to its 1893 crisis). In retrospect, the successful period of U.S. deposit insurance, from 1933 through the 1960s, was an aberration, reflecting limited insurance during those years (insurance limits were subsequently increased), and the unusual macroeconomic stability of the era.

Conclusion

Banking failures, in theory, can be a consequence either of fundamental, exogenous shocks to banks, or alternatively, unwarranted withdrawals by depositors associated with contagions of fear, or panics. Interestingly, although many economists associate contagions of fear with the banking distress of the Great Depression, empirical research indicates that panics played a small role in Depression-era distress, which was mainly confined to regional episodes (e.g., June 1932 in Chicago) or to the banking collapse of 1933.

More importantly, empirical research on banking distress clearly shows that panics are neither random events nor inherent to the function of banks or the structure of bank balance sheets. Panics in the U.S. were generally not associated with massive bank failures, but rather were times of temporary confusion about the incidence of shocks within the banking system. This asymmetric-information problem was particularly severe in the U.S. For the late-nineteenth and early twentieth centuries, system-wide banking panics like those that the U.S. experienced in that period did not occur elsewhere. The uniquely panic-ridden experience of the U.S., particularly during the pre-World War I era, reflected the unit banking structure of the U.S. system. 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 operated on a national basis, who consequently enjoyed greater portfolio diversification ex ante, and a greater ability to coordinate their actions to stem panics ex post. The U.S. also experienced waves of bank failures unrelated to panics (most notably in the 1920s), which reflected the vulnerability to

sector-specific shocks (e.g., agricultural price declines) in an undiversified banking system.

More recent banking system experience worldwide indicates unprecedented costs of banking system distress – an unprecedented high frequency of banking crises, many bank failures, and large losses by failing banks, sometimes with disastrous costs to taxpayers who end up footing the bill of bank loss. This new phenomenon has been traced empirically to the expanded role of the government safety net. Government protection removes the effect of market discipline. It thereby encourages excessive risk taking by banks, and also creates greater tolerance for incompetent risk management (as distinct from purposeful increases in risk). Ironically, the government safety net, which was designed to forestall the (overestimated) risks of contagion seems to have become the primary source of systemic instability in banking.

References

Allen, Franklin, and Douglas Gale (2004). "Financial Fragility, Liquidity, and Asset Prices," *Journal of the European Economic Association* 2, 1015-48.

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.

Barth, Caprio, and 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, 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.

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.

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.

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., 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 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 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.

Caprio, Gerard, and Patrick Honohan (2007). "Banking Crises," this volume, chapter 26.

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

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

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

Diamond, Douglas W. and Rajan, Raghuram (2002). "Liquidity Shortage and Banking Crises," National Bureau of Economic Research Working Paper No. 8937, May.

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

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

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

Halac, Marina, and Sergio Schmukler (2004). "Distributional Effects of Crises: The Financial Channel," *Economia* 5, 1-67.

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

Mason, Joseph R. (2003). "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.

Mishkin, Frederic S., (Editor) (2001). *Prudential Supervision: What Works and What Doesn't,* University of Chicago Press.

Nicholas, Henry C. (1907). "Runs on Banks," *Moody's Magazine,* December.

O'Grada, Cormac, and Eugene N. White (2003). "The Panics of 1854 and 1857: A View from the Emigrant Industrial Savings Bank," *Journal of Economic History*, 63, 213-40.

Rousseau, Peter (2002). "Jacksonian Monetary Policy, Specie Flows, and the Panic of 1837," *Journal of Economic History* 62, 457-88.

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 (2003). *Too Big to Fail: The Hazards of Bank Bailouts*, Brookings Institution.

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.

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

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.