Depression-Era Bank Failures: The Great Contagion or the Great Shakeout?

John R. Walter

Deposit insurance was created, at least in part, to prevent unfounded bank failures caused by contagion. The legislation that created the Federal Deposit Insurance Corporation (FDIC) was driven by the widespread bank failures of the Great Depression. In the years immediately before the 1934, when the FDIC began insuring bank deposits, over one-third of all extant banks failed. Many observers argue that these failures occurred because the banking industry is inherently fragile since it is subject to contagion-induced runs. Fragility arises because banks gather a large portion of their funding through the issuance of liabilities that are redeemable on demand at par, while investing in illiquid assets. Specifically, loans, which on average account for 56 percent of bank assets, tend to be made based on information that is costly to convey to outsiders. As a result, if a significant segment of bank customers run, that is, quickly require the repayment of their deposits, the bank is unlikely to be able to sell its assets except at a steep discount. Bank failure can result.

But do Depression-era bank failures imply the need for government-provided deposit insurance, or is there another explanation of the failures other than contagion and inherent fragility? Some observers question the view that banks are inherently fragile. They argue instead that the banking industry developed various market-based means of addressing runs such that the danger of failure was reduced. They also argue that the banks that failed

The author benefited greatly from comments from Tom Humphrey, Ned Prescott, John Weinberg, and Alex Wolman. Able research assistance was provided by Fan Ding. The views expressed herein are not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.
in response to runs were weak and likely to fail regardless of runs (Calomiris and Mason 1997; Benston and Kaufman 1995).

If not fragility, what might explain the widespread failures before 1934? One possible explanation is that the banking industry was experiencing a shakeout, not unusual in industries that have previously enjoyed significant growth. The number of banks had grown briskly from the mid-1880s until 1921. Beginning in 1921, bank failures increased significantly, such that the number of banks began a precipitous decline that continued until 1934. There are reasons to think that the industry had become overbuilt and that macroeconomic shocks, in conjunction with overbuilding, produced a retrenchment in the industry that lasted for the next 12 years. Indeed, many authors point to the relationship between bank failures and weakening economic conditions. This article suggests that overbuilding could have made the banking industry all the more sensitive to macroeconomic shocks.

A number of other industries provide examples of growth followed by shakeouts, the most recent of which is the telecom industry. If a large portion of Depression-era banking failures were the result of a shakeout rather than contagion, an important argument for deposit insurance is undercut.

Though the termination of bank failures and the creation of the FDIC in 1934 occurred simultaneously, implying that contagion must have been at work, other explanations are just as credible. First, deposit insurance augmented the profits of risky banks, protecting them from failure. Second, the creation of deposit insurance undercut a market process that caused supervisors to close troubled banks quickly.

1. GROWTH IN THE NUMBER OF BANKS

The number of U.S. banks grew rapidly from 1887 until 1921 (Figure 1). Much of the increase coincided with improving economic conditions. Yet, commentators also claim that a good portion of the increase resulted from a statutory change that lowered the minimum capital required to form a new bank as well as careless application of entry standards by regulators. Many of the new banks were viewed by commentators as being ill-prepared for the business of banking. In other words, too many banks were formed without adequate financial or managerial resources. The banking market was overbanked.

As shown in Figure 1, the number of banks began growing rapidly in the late 1880s. The initial run-up in the number of banks followed an economic recovery occurring in 1885 and 1886. The increase in the number of banks was rapid enough, and the size of new banks small enough, to drive down the U.S. average bank size fairly significantly. The average size bank shrank from

---

1 Temin (1976, 83–95) discusses banking failures that resulted from macroeconomic weakness during the Depression.
$1.04 million in 1886 to a low of $660,000 in 1896 and did not return to its 1886 level until 1916. The growth in number of banks was much faster than the pace of economic growth, so that the increase in the number of banks is still quite apparent even when the number of banks is deflated by the level of real GDP (Figure 2).

Most commentators focus on the increase in the number of banks, especially of very small banks, after the beginning of the 20th century. Figure 1 shows that the growth in the number of banks was indeed rapid from 1900 until 1921. An important explanation for the growth in the number of banks during these two decades was the reduction in the minimum capital required to form a bank (Mengle 1990; Wheelock 1993). Specifically, the Currency Act of 1900 lowered from $50,000 to $25,000 the minimum capital needed from investors to start a national bank. In turn, over the next ten years, two-thirds of newly formed banks were quite small, averaging capital of only slightly more than the minimum $25,000 (Mengle 1990, 6).

Beyond this reduction in minimum capital, regulatory laxity was also thought to have contributed to the rapid increase in the number of banks. For example, Federal Reserve analysts concluded that during the first two decades of the 20th century “insufficient attention was paid to the qualifications of those
to whom charters were granted” (Federal Reserve Board 1933, 63–65). These observers saw the banking industry as overbuilt by 1920 (Federal Reserve Board 1933, 67).

Most of these new small banks were formed in small towns and rural communities—especially in the corn and cotton belts of the country. Rising prices of farm commodities along with rising farm real estate values may have played a significant role in the attractiveness of rural banking to new investors (Federal Reserve Board 1933, 65). More generally, economic growth was strong between 1887 and 1920, with the annual rate of growth of GDP averaging over 6 percent for the period. GDP growth was especially strong during 1916, and in the two years of the U.S. participation in World War I, 1917 and 1918.

The formation of state deposit insurance systems in a number of states may also have contributed to a perception of safety and allowed the rapid growth of new small banks. Following the banking panic of 1907, eight states adopted such programs for state-chartered banks (Wheelock 1993, 860). Using regression analysis, Wheelock (1993, 865) finds that the presence of deposit insurance systems encouraged bank formation.
An important trend associated with rapid entry was gradually declining banking profitability. Contemporary writers blamed declining returns on increased competition from new banks. For example, a 1933 Federal Reserve study claimed that “the overbanked condition, which reached its peak shortly after 1920, caused units struggling for existence to increase services to their clients, thus adding to expenses. It served to introduce into many banks high risk and marginal business” (Federal Reserve Board 1933, 67). Net profits relative to assets for the industry fell fairly consistently from 1900 through 1920, from about 2.55 percent to about 1.70 percent (Federal Reserve Board 1933, 67).\(^2\) Bank regulators also noted declining loan standards as rapid entry occurred (Federal Reserve Board 1933, 4).

By several measures the banking industry appears to have become overbuilt. As noted earlier, the numbers grew faster than did overall economic growth. Additionally, as shown in Figure 3, the number of banks increased much more rapidly than did U.S. population. The expansion in the number of competitors may also have been driving down bank profitability and loan

\(^2\)Earnings figures are for national banks only. Until the formation of the Federal Deposit Insurance Corporation, figures on state bank earnings were not collected in any uniform manner.
standards beginning as early as 1900. This declining profitability implies that even without the agricultural shocks of the 1920s, and the macroeconomic shocks of the Depression, the banking industry would have experienced some retrenchment.

2. 1921–1933 BANKING FAILURES

Banking failures, which began growing in number in the early 1920s, coincide with regional agricultural problems, and later with broader economic problems of the Great Depression. Analysts argue that contagion and branching restrictions account for a significant portion of the failures.

The number of U.S. banks peaked in 1921 at about 31,000 banks. The year also produced the beginning of a period of rapid annual rates of bank failure. In 1921 there were 505 failures. As shown in Figure 4, during most of the next eight years failures remained between 500 and 1,000 per year. From 1930 through 1933 failure rates were in the thousands.

The shift in banking failures in 1921 was precipitated by widespread crop failures of that year. Farm problems were evident in falling farm real estate values in corn- and cotton-producing regions. Real estate values fell each year...
J. R. Walter: Depression-Era Bank Failures

from 1921 through 1930 (White 1984, 126). The bank failures of the 1920s were heaviest in states with the most rapid growth prior to the 1920s (Wicker 1996, 7). Difficulties suffered by farmers in the Midwest seem to have driven much of the failure.

Between 1921 and 1930, half of all small banks in agricultural regions failed. Larger banks, however, suffered much less. For example, on average between 1926 and 1930, 74 percent of the smallest banks, those with assets less than $150,000, had weak profits. Here, weak profits are defined as return on equity of less than 6 percent. In contrast, only 21 percent of those banks in the largest size category, with assets greater than $50 million, produced profit rates averaging below 6 percent between 1926 and 1930.

A regulatory shift may also account for the disappearance of some small banks. In the early 1920s the Comptroller of the Currency, the agency that regulates national banks, dropped its branching prohibitions (Mengle 1990, 6). In turn, the number of bank branches grew from 1,400 to 3,500 between 1921 and 1930 (Calomiris and White 2000, 170). These new branches would have brought fresh competition to banking markets, and since branch banks probably had advantages in diversity and scale over small unit banks, unit banks would have been imperiled. Therefore, this liberalization of branching restrictions acted as a shock to small bank profitability, occurring in the 1920s.

Some contemporary commentators claim that improving transportation technology accounts for the decline of small banks, which were once protected from competition by the costs their customers faced to travel to other towns and cities to conduct banking business. These small banks were suddenly faced with new competition once customers’ travel costs fell. The growing availability of the car opened the opportunity to purchase services, including banking services, in central cities (Wheelock 1993).

While the number of banking failures (and therefore the decline in the number of banks) grew rapidly in the 1920s, it grew even more rapidly after the onset of the Depression. Between 1930 and 1932, the number of failures per year averaged 1,700. In 1933, slightly more than 4,000 banks failed.

Contagion is often cited to explain the rapid pace of failures between 1930 and 1933. Contagion could work as follows. A prominent bank fails, and because there is no federal deposit insurance protection, depositors of the failed bank suffer losses. Customers of other banks learn of the failure, believe that their bank might suffer the same fate, and run their banks—i.e., demand cash repayment of their deposits. Since bank assets are typically tied up in loans and securities, to meet these demands for cash, banks must liquidate these assets. If many banks attempt to sell their securities, prices will fall, and banks will suffer losses on the sales. Further, because outsiders have

---

3 Between 1907 and 1917, eight states created deposit insurance systems (White 1983, 207–18). By 1928, all of these state systems had failed.
difficulty determining the worth of bank loans, they will sell at a steep discount, i.e., at firesale prices. Therefore, runs of otherwise healthy banks could cause such banks to suffer losses large enough that they would be unable to meet all depositor demands, creating failures of the banks experiencing the runs. The process would become a cycle, spreading widely.

Friedman and Schwartz (1963) identify three banking crises during the Depression involving widespread runs. During these crises, they, along with others, argue that waves of widespread runs created by a “contagion of fear” produced bank illiquidity and the failure of otherwise healthy banks. They hold that much of the bank failure was the result of such contagion and that healthy banks failed as a result.

Branching restrictions are also viewed as an important explanation of bank failures (Mengle 1990, 7–8). While in the 1920s, branching restrictions were liberalized for national banks, most states placed severe restrictions on branching, or banned it altogether. As of 1929, 22 states prohibited branching and another ten states restricted it somewhat (Mengle 1990, 6). Branching restrictions prevented banks from diversifying their lending, forcing them to concentrate their lending in one geographic area. The lack of diversity made banks more susceptible to failure caused by localized economic weaknesses. Therefore, oddly enough, both branching restrictions and the branching liberalization, discussed earlier, could have contributed to bank failures.

3. CONTAGION IS AN INCOMPLETE EXPLANATION OF FAILURES

Contagion is an incomplete explanation of banking failures for two reasons. First, contagion was not a factor in the failures that occurred in the 1920s. The bank failures of the 1920s were not caused by and did not create banking panics, that is, banking runs or heavy withdrawals of currency. There was no increase in currency in circulation during the period (Wicker 1996, 1). If factors other than contagion were important in the 1920s, it seems likely that these same factors would also be at work in the 1930s as well. Therefore, contagion is unlikely to account for at least some of the failures in the Depression years of the 1930s.

Second, in their influential work on U.S. banking and monetary history, Friedman and Schwartz (1963) maintain that the bank failures of the 1930s were contagion-induced. Later reviews, however, cast doubt on the contagion explanation. Using regression analysis to identify factors that accounted for individual bank failures during the 1930 crisis, which occurred in November and December of that year, White (1984) finds that these failures were no different from those occurring in the 1920s. In other words, he finds no evidence that these failures were driven by contagion-sparked illiquidity. Instead, these
failures were the result of bad loans. Friedman and Schwartz (1963) give special emphasis to the failure of the Bank of the United States in December 1930 as a cause of contagion-induced bank failures. Wicker (1996, 36–38) questions this view, concluding that the bank’s failure did not lead to major deposit withdrawals and could not have accounted for the failures of other banks.

Further, of the bank failures that took place between 1930 and 1932, more than 60 percent occurred outside of panic months (Wicker 1996, 1). While some of these failures might have been the residual effect of runs, it seems likely that most of the failed banks had financial problems unrelated to runs. If an otherwise healthy bank survived a panic, it would have survived the aftermath. Last, Calomiris and Mason (1997) examined the financial condition of Chicago banks during the 1932 banking panic in that city. They found that “while depositors did confuse panic survivors with panic failures, the failure of solvent banks did not result from that confusion.” Calomiris and Mason note that the Chicago situation may have been special since Chicago banks met the panic by joining forces to protect banks the group viewed as solvent. Such combined efforts would have been more difficult for the many far-flung banks outside of major cities. Nevertheless, through correspondent relationships, support of banks could have transpired even in a sprawling banking system.

Still, one might wonder: if contagion was not the cause of bank failures, then why did bank failures largely come to a halt with the creation of federal deposit insurance in 1934 (as seen in Figure 4)? If, as I argue, a portion of the bank failures of the 1920s and 1930s were the result of a typical shakeout, there is no reason to expect the shakeout to have ended with the creation of the FDIC. One would only expect failures to be ended by the FDIC’s creation if they were caused by contagion, a problem overcome by the creation of federal deposit insurance.

Part of the reason for the cessation of failures, and one that is not dependent on the existence of contagion, is the shifting economic backdrop. Those inherently low-profit banks that were created during the rapid growth of the banking industry would have been the first to be driven to insolvency by the extreme macroeconomic problems of the early 1930s. In other words, economic difficulties concentrated the shakeout in the early 1930s. As a result, low-profit banks created during the growth of the industry—banks which might have continued to fail over a number of years after 1933—were quickly expunged by the severe economic times. The severity of the macroeconomic problems is shown by the declines in economic output, which fell by 8.61 percent in 1930, and by 6.42 percent, 13.00 percent, and 1.27 percent in 1931, 1932, and 1933, respectively. But then in 1934 economic output began growing rapidly, increasing by 10.81 percent, and averaged 7.08 percent from 1934 through 1939 so that banks were strong enough to weather the early 1930s the following years.
An additional reason exists—not dependent on contagion—as to why failures ended at the same time the FDIC was formed: Prior to the FDIC’s creation, depositors had strong incentive to monitor their banks’ health and, in the case of signs of weakness, withdraw deposits. When depositors withdrew funds, government supervisors were forced to review the state of the bank and either close it or allow it to reopen if shown healthy. Once deposits gained FDIC protection, this disciplining mechanism was removed. Without the mechanism, the opportunity for supervisors to forbear arose. Certainly, in the years immediately after the widespread bank failures of the 1920s and early 1930s, it seems that supervisors would tend to err on the side of stability and be reticent to close banks unless the evidence of a bank’s problems was quite strong. Consequently, that failures largely ceased at the same time the FDIC was created is not necessarily evidence that contagion was the cause of most bank failures.

Further, FDIC insurance provided a significant profit-boosting subsidy to the riskiest banks. With the government backing their deposits, in 1934 all banks could suddenly gather deposits at the risk-free interest rate. Those banks that might have failed in the near future—the riskiest banks—benefited most. Unless depositors were completely oblivious to their bank’s health, in the absence of deposit insurance, such banks would have had to pay depositors high interest rates. Had the FDIC charged risk-adjusted deposit insurance premia, the riskiest banks would have enjoyed no benefit. But premia were calculated as a simple percentage of deposits. Therefore, troubled banks enjoyed a sudden boost in their profits due to the introduction of FDIC insurance. Failures would naturally be minimized, not because contagion was halted, but instead because FDIC insurance provided a subsidy to those banks most likely to fail.

4. SHAKEOUTS IN OTHER INDUSTRIES

Rapid increase in the number of producers followed by an equally rapid decline is a pattern not only evident in banking, but also in a number of industries. Examples include automobiles, tires, televisions, penicillin, and most recently, telecommunications. While the banking industry failures are blamed on heavy reliance on liabilities that are redeemable upon demand, other factors must have been at work to produce rapid failures in these other industries since they are not reliant upon demand liabilities. Perhaps many of the failures in banking were simply the result of a typical shakeout, driven by factors other than contagion, as in these other industries.

The growth pattern in banking—a significant increase in the number of banks, followed by a rapid decline—is similar to the growth and shakeout that a number of industries experienced, industries for which there are no claims of inherent fragility or contagion-driven runs. Klepper (2002, 37) notes that “After their commercial introduction, the number of producers of automobiles,
tires, televisions, and penicillin initially grew and then experienced a sharp decline or shakeout.” For example, Figure 5 shows the pattern of growth and decline for automobiles.

More recently, the same pattern was displayed in the telecommunications sector. In the mid-1990s, three factors together implied great opportunities for growth in the industry. First, rapidly growing use of the Internet generated an expanded demand for data transmission services over phone, cable, and long-haul fiber networks. Observers claimed that Internet use was doubling every year, or even several times a year.

Second, an important regulatory change was brought to the telecommunications industry by the Telecommunications Act of 1996. The goal of the Act was to introduce new competition to local telephone service, which was largely monopolized by the regional Bells. It opened local telephone service to competition, for example, by requiring incumbent local service providers

---

to sell entrants access to the networks that reach homes and businesses. It was thought that entrants would provide local phone service and that selling access would avoid duplication of existing telephone lines to individual homes and businesses. The entrants, known as competitive local exchange carriers (CLECs), rapidly increased in number and earnings. From 30 CLECs in 1996, their number had grown to 711 in 2000. At the same time, revenues of the CLECs rose from $5 billion to $43 billion (Couper et al., 15).

Third, advances in fiber optic technology were occurring rapidly in the mid-1990s, lowering the cost of providing data and voice services to households and businesses. Existing firms, as well as founders of new firms, expected these factors to open huge new markets and responded by rapidly increasing investment in telecommunications equipment and communications lines. However, as it turned out, by 2000 the promise of the 1996 Act had not been borne out, long-haul fiber communications lines were significantly overbuilt, and the overall economy began to slow. The Act did not lead to nearly as much growth for competitors in local service. The Act’s pricing ambiguities meant that rules that would allow entrants to buy access into homes and businesses were slow to be developed by regulatory agencies and were held up by lawsuits. Overcapacity in fiber lines was the result of rapid improvement in the technology for transmitting data over these lines, meaning that fewer were needed than expected. Further, the growth in demand for data communications, while rapid, was not as rapid as expected (Couper et al., 13 and 19). Beginning in mid-2000, the telecom industry began a rapid retrenchment that involved huge numbers of failures of new firms that had only recently appeared to have bright futures. For example, the number of CLECs declined by about 80 percent from 1999 to 2001 (Grossi, 4).

5. EXPLANATION OF THE BANKING SHAKEOUT

Several important industries displayed a pattern of rapid growth followed by widespread failure, similar to that in the banking industry. Yet failures in these industries cannot be blamed on contagion resulting from a reliance on demandable liabilities, implying that something else may also have been at work during the similar boom and bust in banking. Still, these industries differed from banking: Automobiles, tires, televisions, and penicillin were all new industries, and the telecommunications industry was experiencing rapid technological change. Therefore a good bit of instability can be expected as the efficient industrial structure evolved. In contrast, the banking industry of the late 19th and early 20th centuries was not new.

While not new, the banking industry was undergoing rapid change. The industry was a primary source of capital to a nation undergoing rapid technological and industrial change. The expansion of railroads drove down transportation costs. The spread of electronic communication, the telegraph and
telephone, rapidly lowered communications costs. As noted earlier, the popularity of the automobile meant that communities were less isolated from one another, and commerce was better able to flow across local communities. Also, the growth of the corporation as a business structure and the issue of debt by corporations radically changed the financial structure of the business world. Further, branching and capital regulations in banking were being significantly modified.

In an environment of wide-ranging industrial, technological, and regulatory change, it should not be surprising that the banking industry would struggle in ways similar to a new industry. Bank organizers perceived profit opportunities in the changing setting and rapidly formed banks to take advantage of such opportunities. In doing so, they drove down bank size and profitability, as noted earlier. Ultimately they caused the industry to become overbuilt, and failures ensued.

Were bank organizers acting irrationally? Not according to theories economists have advanced to explain the fairly common phenomenon of industry shakeout. Such theories argue that firms can often be expected to expand new investment too far. One such theoretical explanation argues that shakeouts are a result of investments in research and development (R&D) (Klepper 2002, 38). Under this explanation, firms invest in R&D to lower their costs, but the cost advantage grows with size, such that large firms benefit more from R&D. Further, some firms are better at R&D than others.

Firms enter and make R&D investments to acquire the high margins that are expected to accrue from such actions. Instead, rapid entry drives down profit margins. Eventually entry stops, and smaller firms, which have higher per-unit R&D costs, fail. Firms that make less productive use of R&D expenditures also fail, since their costs will be relatively high. In this environment, entry seems to progress too far, since many of the firms that were initially profitable fail when profit margins are driven down. Unanticipated increases in costs or decreases in earnings can also cause failure.

More broadly, overinvestment works as follows. A technological or business process innovation occurs, and firms make investments to take advantage of the new technology. Firms are uncertain concerning how much their output might increase as they implement the new technology but must invest before knowing. Alternatively, firms are uncertain of the extent of demand for their new product, but must invest before knowing. Ultimately firms often overshoot, profits decline, and the least efficient firms fail.

Was such overshooting likely in banking in the late 19th and early 20th centuries? Evolving telecommunications and transportation technology as well as shifting financial arrangements were influencing both the technology

---

5 Alternative explanations of boom and shakeouts are offered by Jovanovic and MacDonald (1994) and Barbarino and Jovanovic (2004).
of banking and of the firms to which banks make loans. In this environment, existing banks and potential investors in new banks faced a great deal of uncertainty about the proper scale of the industry, and entry was rapid, relative to GDP growth and population growth. Average profits were driven down in the banking industry, and after 1921, while larger banks remained strong, small bank profits fell significantly. Clearly the demand shocks of the 1920s and the Depression played a major role in the shakeout in banks, but overbuilding appears to have been of significant importance.

6. SUMMARY

As the U.S. economy grew and evolved in the late 19th and early 20th centuries, the banking industry grew even more rapidly, especially in raw numbers of banks, as well as in assets relative to GDP and numbers relative to population. Contemporary analysts maintained that the growth produced an overbuilt industry. Ultimately, failures shrank the number of banks. Failures first became significant in the early 1920s, continued throughout the 1920s, and became even more numerous during the Depression. They largely ended in 1934, at the time of the formation of the FDIC.

A long-held explanation for bank failures during the Depression is contagion, whereby the initial failure of one bank leads to widespread runs on other banks and their failure. According to this explanation, many of the Depression-era failures were inappropriate, meaning that the failed banks were solvent and would have survived without contagion-induced runs. The solution to contagion was deposit insurance provided by the federal government, which put a stop to failures.

Still, given that the cycle of failures began in the early 1920s, long before contagion was evident, one must question contagion as the overriding cause. Instead the banking industry appears to have been experiencing a shakeout, exacerbated by weakened economic conditions during the Depression. While at first blush, the fact that failures stopped virtually simultaneously with the formation of the FDIC implies that contagion must have been at work, other explanations for this simultaneity are just as credible. For example, risky bank profits were certainly boosted by the provision of deposit insurance at premiums that did not reflect bank risk, protecting these banks from failure. Further, the process that drove supervisors to quickly close troubled banks was undercut once deposit insurance was established.
REFERENCES


Federal Reserve Board. 1933. *Banking Profits, 1890–1931*. Federal Reserve Committee on Branch, Group, and Chain Banking.


