



Categories and causes of bank distress during the great depression, 1929–1933: The illiquidity versus insolvency debate revisited

Gary Richardson *

*Department of Economics, National Bureau of Economic Research, University of California,
3151, Social Science Plaza, Irvine, CA 92617-5100, USA*

Received 9 September 2006
Available online 29 May 2007

Abstract

During the contraction from 1929 to 1933, the Federal Reserve System tracked changes in the status of all banks operating in the United States and determined the cause of each bank suspension. This essay analyzes chronological patterns in aggregate series constructed from that data. The analysis demonstrates both illiquidity and insolvency were substantial sources of bank distress. Periods of heightened distress were correlated with periods of increased illiquidity. Contagion via correspondent networks and bank runs propagated the initial banking panics. As the depression deepened and asset values declined, insolvency loomed as the principal threat to depository institutions.

© 2007 Published by Elsevier Inc.

Keywords: Great Depression; Bank failure; Insolvency; Illiquidity; Federal Reserve; National Archives

1. Introduction

During the Great Contraction, depository institutions exited from the banking industry more rapidly than at any other time in United States history. Of the roughly 24,000 institutions in operation in January 1929, only about 14,000 remained when the banking holiday began in March 1933. Banks failures were clustered during events that contemporaries

* Fax: +1 949 824 2182.

E-mail address: garyr@uci.edu

described as panics, but banks closed their doors in large numbers at other times. What caused the banking system to contract so suddenly and severely?

Economists, politicians, and pundits have debated this question since the early 1930s. Innumerable explanations have been offered. The scholarly debate revolves around two competing theories. One school maintains that the underlying causes were withdrawals of deposits, illiquidity of assets, and the Federal Reserve's reluctance to act. A contagion of fear, a flight to cash holdings, and withdrawals *en masse* drained deposits from banks and pushed financial institutions towards collapse. Federal Reserve mistakes exacerbated the credit crunch (Friedman and Anna, 1971; Wicker, 1996). A contending school concludes that banks failed because the economy contracted. Asset prices fell. Loan default rates rose. Banks became insolvent. These fundamental forces accentuated a process of bank liquidation and that began during the 1920s (Temin, 1976; White, 1984; and Calomiris and Joseph, 2003).

Both points of view have been supported with extensive evidence. The evidence corroborating the traditional Friedman-Schwartz illiquidity hypothesis consists of data on bank suspensions aggregated at the national or regional level. These data reveal bank failures clustered in time and space, often coinciding with turning points in macroeconomic time-series such as indices of industrial production, the money supply, the money multiplier, interest rates, and the deflation rate. Narrative sources from the 1930s characterize many of these clusters as banking panics (Friedman and Anna, 1971; Wicker, 1996).

The evidence supporting the Temin-White insolvency hypothesis consists of data at lower levels of aggregation, or disaggregated data consisting of samples of national banks, or panels of banks from within individual cities, states, or Federal Reserve districts. Temin (1976) estimates a regression of state-level bank-failure data on characteristics of the state and national economies, including proxies for poor loans and agricultural distress (see also White, 1984). More recently, Calomiris and Joseph (2003) analyze a panel of data for all Federal Reserve member banks. These studies reveal significant correlations between the characteristics of banks, the environments in which they operated, and their chances of surviving the contraction.

Differences in data sources are one reason why this debate has not been resolved. The different sources illuminate different dimensions of the banking industry and highlight different points in time and space, leading to different depictions of events. Another reason is that the extant sources share several weaknesses. All of the sources provide imperfect information about changes in banks' status. The extant evidence distinguishes neither temporary from terminal suspensions, nor voluntary from involuntary liquidations, nor institutions afflicted by illiquidity from banks suffering insolvency. The extant evidence contains information neither on the causes of bank suspensions nor the number of bank mergers. The extant evidence neglects consolidations of banks in financial difficulties and seeking to avoid suspension. The samples are neither random nor representative. The microdata from bank balance sheets comes almost entirely from national and Federal Reserve member banks, whose experience differed from state-chartered institutions. The microdata indicate the financial health of the institution only at annual (or longer) intervals, and not at the time of distress. For aggregate data, the smallest period of aggregation at the national level is the month and at the Federal Reserve District is the year. Key terms remain undefined, leaving much open to interpretation.

This essay examines a new source of evidence that provides a comprehensive view of the ways in which banks exited from the banking industry and the reasons that banks

suspended operations. The new source covers all banks—including Federal Reserve members and nonmembers, national and state, incorporated and private—from the onset of the contraction in 1929 and until the national banking holiday in March 1933. The new source contains observations for an array of events, including closings, reopenings, and consolidations, and indicates the cause of suspension for all banks. The source is the archives of the Federal Reserve Board of Governors, whose Division of Bank Operations tracked changes in the status of all bank operating in the United States, analyzed the cause of each bank suspension, and recorded its conclusion and financial information for each bank on (or near) the date of suspension.¹

This essay analyzes aggregate series derived from the St. 6386 database using non-parametric methods. The goal is filter the data as little as possible, allowing the evidence—and the hundreds of bank examiners, accountants, receivers, economists, Federal Reserve agents, and others who collaboratively created the data set—to speak for itself. Summary statistics, charts, and graphs demonstrate that some types of bank distress were less prevalent than the conventional wisdom suggests; whereas other types of bank distress, which according to the conventional wisdom seldom (or never) occurred, did in fact occur and may have played a pivotal role propagating the banking crisis. For example, [Bernanke \(1983, pp. 259–260\)](#) and [Friedman and Anna \(1971, pp. 311–312\)](#) assert that temporary bank suspensions did not occur during the contraction; however, temporary bank suspensions did, in fact, occur in large numbers during the banking crises in the fall of 1930 and 1931.

The evidence presented in this essay helps to answer a fundamental question concerning the collapse of the banking system: why did banks fail? This question can be restated in terms of the current academic debate as: did illiquidity or insolvency cause the collapse of the financial system? The answer is that the nature of the banking crisis changed over time. Before October 1930, the pattern of failures resembled the pattern that prevailed during the 1920s. Small, rural banks with large loan losses failed at a gradually rising rate. In November 1930, the collapse of correspondent networks triggered banking panics in the center of the nation. The largest counter-party cascade stemmed from the collapse of the Caldwell conglomerate. Runs rose in number and severity after prominent financial institutions in New York and Los Angeles closed amid scandals covered prominently in the national press. More than a third of the banks which closed their doors to depositors during this crisis soon resumed normal operations. Following Britain's departure from the gold standard in September 1931, the depression deepened. Asset values declined. Insolvency loomed as the largest threat facing depository institutions. During the financial crisis in the winter of 1933, almost all of the banks that failed were liquidated at a loss.

Overall, between the Great Crash and the Banking Holiday, both illiquidity and insolvency were substantial sources of bank distress. Nearly three-fourths of the banks that closed their doors due to financial difficulties were insolvent. Slightly more than one quarter were solvent, and without outside financial assistance, reopened for business, or repaid all of their depositors and creditors, or merged at face value with other institutions. Heavy withdrawals were a primary cause of nearly one half of all bank suspensions and a contributing cause of another one sixth. Frozen and devalued assets were a primary cause

¹ For additional information on this new data, see the series of essays that I have written to introduce the source, [Richardson \(2006a,b,c,d\)](#) and [Richardson and Troost \(2006\)](#).

of approximately one half of all bank suspensions and a contributing cause of another one quarter.

The rest of this essay substantiates this depiction of the banking crises of the Great Contraction. Section 2 describes the data, defines key concepts, and discusses issues important for interpreting empirical work based upon this new and unique source. Section 3 presents chronological patterns for the four categories of bank distress: terminal suspensions, temporary suspensions, consolidations due to financial difficulties, and voluntary liquidations. Section 4 presents chronological patterns for the principal causes of bank suspensions: heavy withdrawals, problematic assets, closure of correspondent banks, defalcations, and mismanagement. The data demonstrate that surges in bank suspensions coincided with periods of increased illiquidity. Section 5 discusses the implications these chronological patterns, which corroborate some conjectures, cast doubt on others, and raise new questions concerning the causes and consequences of the collapse of the banking system during the Great Depression.

2. Data source and definitions

From 1929 to 1933, the Federal Reserve Board of Governors' Division of Bank Operations recorded information about changes in bank status on three forms. *Form St. 6386a* reported bank consolidations. *Form St. 6386b* reported bank suspensions. *Form St. 6386c* reported all other bank changes. The Division of Bank Operations gathered these data through a nationwide reporting network that it established during the 1920s. The network's purpose was to collect uniform and comprehensive information about all banks operating in the United States—national and state, member and nonmember, public and private. The complete series of St. 6386 forms survives in the National Archives of the United States.²

As part of its data collecting endeavors, the Board of Governors developed a lexicon for classifying changes in bank status and bank distress. In the Federal Reserve's lexicon, a suspension was a bank that closed its doors to depositors and ceased conducting normal banking business for at least one business day. Some, but not all, suspended banks reopened for business. Liquidations were permanent suspensions. A liquidating bank closed its doors to the public, surrendered its charter, and repaid depositors, usually under the auspices of a court appointed officer known as a receiver. A voluntary liquidation was a category of closure in which banks ceased operations and rapidly arranged to repay depositors the full value of their deposits. Voluntary liquidations did not require the services of receivers and were not classified as suspensions. A consolidation (or merger) was the corporate union of two or more banks into one bank which continued operations as a single business entity and under a single charter. The categories of bank distress were

² The forms may be found in the National Archives, Record Group 82, Federal Reserve Central Subject File, file number 434.-1, "Bank Changes 1921–1954 Districts 1929–1954—Consolidations, Suspensions and Organizations—St. 6386 a,b,c, (By States) 1930–1933." The forms are filed alphabetically by state, name of town or city, and name of bank. Multiple entries for individual banks appear in chronological order. To avoid repeated, lengthy citations in the body of the essay and this appendix, after quotations from the archival data, the information required to locate the form (i.e. name of state, town, bank) is indicated in brackets.

typically construed to be temporary suspensions, terminal suspensions (i.e. liquidations), voluntary liquidations, and consolidations due to financial difficulties.³

The Federal Reserve's lexicon also covered common causes of bank suspensions. The Federal Reserve attributed most suspensions to one (or more) of five common causes. The first was *slow, doubtful, or worthless paper*. The term *worthless paper* indicated an asset with little or no value. The term *doubtful paper* meant an asset unlikely to yield book value. The term *slow paper* meant an asset likely to yield full value in time, but whose repayment lagged or which could not be converted to full cash value at short notice. The second common cause of suspension was *heavy withdrawals*, the typical example being a bank run. The third was *failure a banking correspondent*. Correspondents were banks with ongoing relationships facilitated by deposits of funds. A typical example is a county bank (the client or respondent) which kept its reserve deposits within and cleared its checks through a national bank in a reserve city (the correspondent). The fourth common cause was *mismanagement*. The fifth was *defalcation*, a monetary deficiency in the accounts of a bank due to fraud or breach of trust.

Completing the St. 6386 forms (and determining the cause of a bank's suspension) involved implementing algorithms designed to classify complex transactions. The algorithms evolved over a four-year period, from 1925 to 1929, during which the Division of Bank Operations sought comments from bankers and examiners about the accuracy of its procedures. When determining the causes of failures, the Board of Governors sought to gather information about suspensions from the man on the spot who knew the facts of the issue at hand. The Board gathered information from examiners, receivers, correspondents, state banking departments, court officers, the bank's own management, and local and national publications. This array of sources, now no longer extant, provided the Board of Governors with an array of information, now unavailable to economic researchers, such as the health of a bank's assets on the date of suspension, the deposits lost by the bank in the period preceding suspension, the lawsuits (or criminal charges) pending against bank management, and the links that the failed bank had to other financial institutions. Therefore, the Federal Reserve agents could determine whether the bank had experienced a run, whether closed due to inability to maintain cash flow or due to fear of impending insolvency, whether the loss of correspondent linkages forced it out of operations, or whether the bank closed due to the dishonesty of its management. Modern scholars cannot observe these phenomena.⁴

³ Contemporary publications often employed the terms failure and closure as synonyms, which typically (but not universally) referred to the sum of temporary and terminal suspensions.

⁴ Companion papers discuss the quality of this evidence (Richardson 2004b and 2004d). These essays describe in detail how the Board of Governors determined the cause of each bank's demise. The method involved gathering financial data for each institution; talking with the regulators, examiners, receivers, correspondents, and bankers who knew the facts of the issue at hand; applying a set of algorithms devised to determine the cause of the failure; and reporting the results in a lexicon devised to convey the conclusions clearly, concisely, and consistently. The companion paper provides examples of the algorithms and copies of the blank forms. The companion paper describes when, where, why, and how the Board of Governors collected the necessary information and issues important for interpreting empirical work based upon this new and unique source. The discussion emphasizes the lexicon that the Board of Governors devised during the 1920s to facilitate their analysis, the methods that Federal Reserve agents employed to ascertain the causes of bank suspensions, the scope and scale of the data collection process, and the procedures that the Federal Reserve used to validate and cross-check the evidence.

The following sections examine chronological patterns in the rediscovered archival evidence. Because the database covers the entire population of banks and contains information on all of the ways in which those banks could depart from the banking business, the typical problems of probabilistic inference—such as extrapolating from samples, sample selection, and censoring—are not concerns.

I emphasize that no judgments are made in the analysis about the reasons why particular banks failed. Rather, those judgments were made by contemporary experts possessing far more information about each event than is available to scholars today. In this essay, I report their conclusions, both in their own lexicon and after translating the conclusions into terms used by modern economists.

3. Categories of bank distress

This section examines chronological patterns on categories of bank distress. Table 1 presents annual data. Several patterns appear prominent. The majority of bank changes due to financial distress were terminal suspensions (72.9%). The remaining changes were divided between temporary suspensions (12.3%), consolidations due to financial difficulties (9.7%), and voluntary liquidations (5.2%). Temporary suspensions occurred at a higher rate early in the depression, when more than one in five suspended banks reopened for business, and usually occurred during surges of bank suspensions which contemporary observers classified as banking panics. Consolidations due to financial difficulties followed a different pattern. Struggling banks were least likely to merge in 1930 and most likely to do so in 1931. The rate of voluntary liquidations followed no discernible trend.

Fig. 1 displays patterns of terminal and temporary suspensions from January 1929 to March 1933. The rates of both types of suspensions remained near pre-depression levels until November 1930, when the rates rose suddenly. The rise in temporary suspensions was particularly pronounced. Few temporary suspensions occurred during typical weeks; many weeks witnessed none. Almost all temporary suspensions occurred during waves of bank failures such as the banking crisis of fall 1930, when more than 40% of all institutions which closed their doors to depositors soon reopened for business. The percentage fell as the depression progressed. Less than 20% of the banks which closed their doors during the fall of 1931 later reopened. Only a small fraction of the banks which suspended operations during 1932 and 1933 reopened, and most of those that did manage to resume operations

Table 1
Bank changes due to financial difficulties by category and year January 1929–March 1933

Categories	1929	1930	1931	1932	1933	Total
Suspensions, terminal	433	1078	1958	1300	420	5189
Suspensions, temporary	70	278	316	161	48	874
Consolidations due to financial difficulties	54	136	309	151	38	688
Voluntary liquidations	39	86	110	99	37	371
Total	596	1578	2693	1712	543	7122

Note. Statistics for 1933 include only the months January–March. Figures do not include banks closed due to government proclamations, moratoria, or banking holidays.

Source. National Archives and Records Administration.

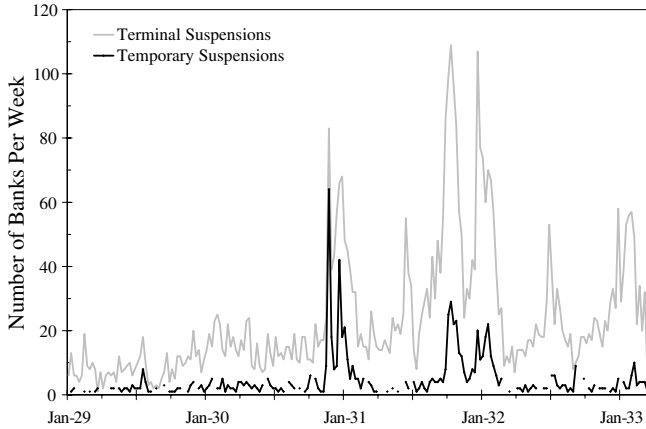


Fig. 1. Terminal and Temporary Bank Suspensions, January 1929–March 1933. Definitions: A bank suspension occurred when a bank closed its doors to depositors and ceased conducting normal banking business for at least one business day. Banks that suspended terminally never reopened. Banks that suspended temporarily did reopen. Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Record Administration. See footnote 1 for details.

did so after receiving infusions of cash from stockholders, depositors, the Reconstruction Finance Corporation, and other sources.

Fig. 2 displays patterns of voluntary liquidations and consolidations due to financial difficulties. Kernel-smoothing illuminates trends over time. Consolidations rose during

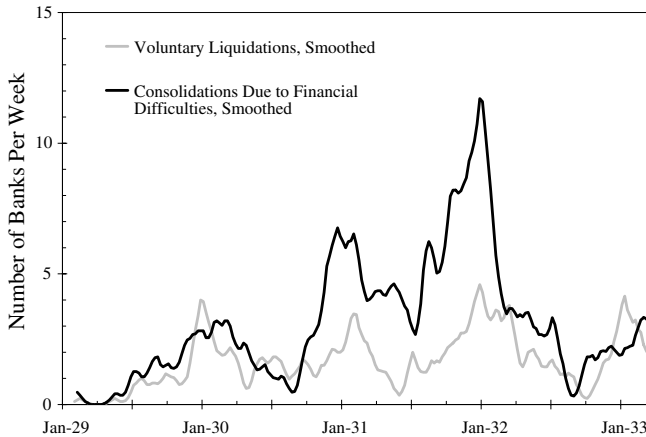


Fig. 2. Consolidations Due to Financial Difficulties and Voluntary Liquidations January 1929–March 1933. Definitions: A voluntary liquidation occurred when ceased operations and rapidly arranged to repay depositors the full value of their deposits. A consolidation was the corporate union of two or more banks into one bank which continued operations as a single business entity and under a single charter. Notes: Consolidations and liquidations are smoothed using Epanechnikov kernel filters with bandwidth equal to four weeks. Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Record Administration.

the waves of bank failures and peaked at the end of 1932. Voluntary liquidations varied greatly from week to week, and no trend is discernible.

Fig. 3 indicates rates at which banks departed from the banking business each year. The horizontal axis divides the rates into solvencies and insolvencies. The share of solvent departures appears above the axis. The numerator of this ratio consists of banks in each year which (i) voluntarily liquidated and (ii) consolidated due to financial difficulties (dark gray area) plus (iii) the number of mergers of banks not in financial distress (light gray area). The number of mergers is a lower-bound approximation, because the series equals ‘mergers’ minus ‘new bank openings.’ The denominator is the number of banks in operation at the beginning of each year.

The share of insolvent departures appears below the axis in Fig. 3. The numerator is the number of terminal suspensions. The denominator is the number of banks in operation at the beginning of each year. The values for 1933 include banks which ceased operations in January, February, and March with the exception of those closed due to banking holidays declared by state and federal governments. The dotted outlines indicate the annualized rate of departure during the first three months of 1933, or in other words, what would have happened had the departure rate during the first quarter continued throughout the year. The last column indicates the rates of departure averaged over the entire contraction.

Fig. 3 shows that for the period January 1929–March 1933, departures from the banking business were split about evenly between solvent and insolvent. Solvent departures—principally mergers but also voluntary liquidations and consolidations under duress—were concentrated during the first two years of the contraction (i.e. 1929 and 1930) and during the first half of 1932, a period during which the Federal Reserve System pursued an expansionary policy. Insolvent departures gradually increased during 1930 and peaked in 1931

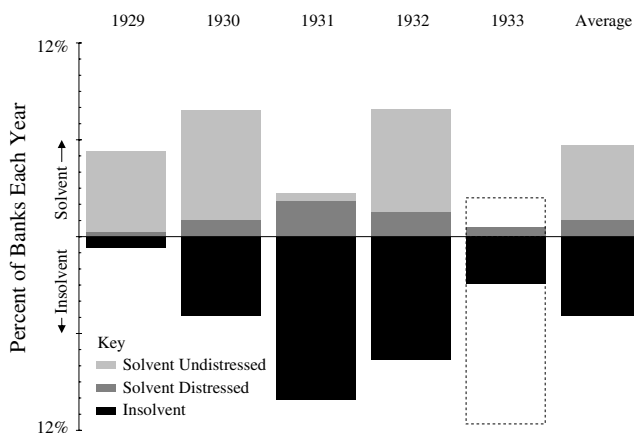


Fig. 3. Solvency and Insolvency among Banks Departing the Banking Business Each Year January 1929–March 1933. Definitions: Solvent departures consist of distressed and undistressed departures. Solvent distressed departures include banks that voluntarily liquidated and consolidated due to financial difficulties (summed together in the dark grey area). Solvent undistressed departures banks that merged without undergoing financial distress (light grey area). Insolvent departures consist of terminal suspensions (black area). Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Records Administration and FRB’37.

following Britain's departure from the gold standard, to which the Federal Reserve reacted by raising the discount rate and contracting the money supply. Insolvent departures continued at a high rate in 1932 and spiked upwards during the months before the Roosevelt administration declared a national banking holiday in March 1933.

Together, Figs. 1–3 demonstrate that during the depression, bank distress occurred in many forms. Some forms for which data have been available—such as the liquidation of insolvent banks—have been studied extensively. Other forms for which data have not been extant—such as temporary suspensions and consolidations due to financial difficulties—have not been analyzed. The hitherto unobserved forms of distress were concentrated during surges of bank suspensions which contemporaries characterized as panics. The concentration suggests that the causes of suspensions during panics differed from the causes during ordinary times. The next section, which examines evidence on the causes of bank suspensions, strengthens that supposition.

4. Causes of bank suspensions

The previous section documented when and how banks departed from the banking business. This section determines why banks did so by analyzing chronological patterns in the causes of bank suspensions. Table 2 presents annual figures. Row (a) indicates the total number of banks suspending operations. Row (a.1) indicates the total number of banks suspending operations due to problems with assets and liabilities. Imbalances often arose on one side of the balance sheet, but could arise simultaneously on both sides. Rows (a.1.1) to (a.1.5) display the possible combinations: (a.1.1) slow, doubtful, or worthless assets listed as the primary cause of suspension; (a.1.2) assets listed as the primary cause and heavy withdrawals listed as a contributing cause; (a.1.3) both withdrawals and assets listed as primary causes; (a.1.4) heavy withdrawals listed as the primary cause and assets as the contributing cause; and (a.1.5) withdrawals alone as the primary cause of suspension. Row (a.2) indicates the number of banks suspending due to the closure of a correspondent. Row (a.3) indicates the number of banks suspending due to defalcations,

Table 2
Causes of suspensions January 1929–March 1933

Causes of suspensions	1929	1930	1931	1932	1933	Total
(a.1) Assets and liabilities	242	848	1720	1125	337	4272
(a.1.1) Assets	102	316	457	354	91	1320
(a.1.2) Assets (p) and withdrawals (c)	44	142	294	250	61	791
(a.1.3) Assets (p) and withdrawals (p)	16	71	163	74	22	346
(a.1.4) Withdrawals (p) and assets (c)	38	145	487	294	116	1080
(a.1.5) Withdrawals	42	174	319	153	47	735
(a.2) Closure of correspondent	28	140	97	68	24	357
(a.3) Defalcation and mismanagement	143	153	159	76	16	547
(a.4) Other and multiple causes	90	215	298	193	91	887
Total	503	1356	2274	1462	468	6063

Note. (p) indicates a primary cause. (c) indicates a contributing cause. Statistics for 1933 include only the months January–March. Figures do not include banks closed due to government proclamations, moratoria, or banking holidays. Rows (a.1.1) to (a.1.5) indicate subtotals of line (a.1).

Source. National Archives and Records Administration.

mismanagement, excessive loans to officers and directors, excessive investment in building and facilities, and similar circumstances. Row (a.4) indicates the number of banks suspending due to other or multiple causes. The latter consists in part of complex cases which do not fit into the mutually exclusive categories above, often because they spanned two or more classifications. An example is a poorly managed bank which failed to enforce collections on its slow farm loans and which experienced runs after local newspapers revealed that its president embezzled funds from savings accounts.

In Table 2, several patterns appear prominent. The share of suspensions due primarily to problems on the asset side of the balance sheet fell through 1931, rose in 1932, and fell again in 1933. The share of suspensions due primarily to withdrawals rose through 1931, fell in 1932, and rose again in 1933. The ratio of assets over withdrawals varied over time. Withdrawals' role in the collapse of the banking system peaked in 1931, while before and after, asset problems caused the majority of suspensions. The share of suspensions due to defalcation and mismanagement fell throughout the depression, perhaps because corrupt and poorly-managed institutions lacked the resources needed to weather the initial storm. The share of suspensions due to the closure of correspondents peaked during 1930.

Fig. 4, which plots the number of suspensions each week due to the closure of correspondents (and for sake of comparison also plots total changes due to financial distress), reveals why. The typical week witnessed few, if any, suspensions from this source. The weekly mode and median were zero. The mean was one. The weekly number rose during July of 1929, when the Mediterranean fruit fly epidemic produced a banking panic in Florida, but remained near zero until November 1930, when it spiked sharply upwards.

Fig. 5 paints an analogous picture for suspensions whose primary cause was heavy withdrawals. Suspensions due to withdrawals were rare during the initial year of the depression but rose following the collapse of correspondent networks in November 1930. During 1931, when bank suspensions surged repeatedly, withdrawals were a

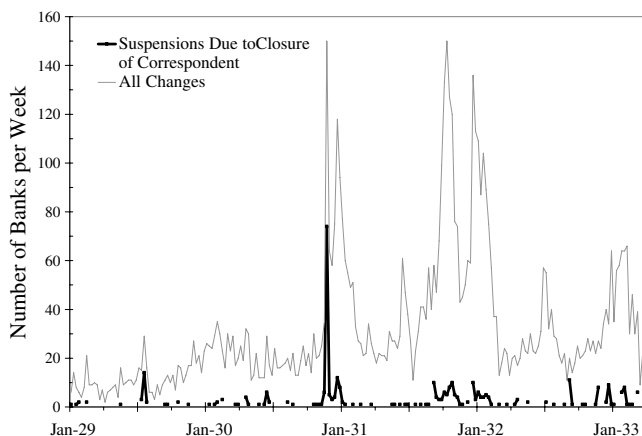


Fig. 4. Suspensions Due to Closures of Correspondents, January 1929–March 1933. Definitions: The series *All Changes* indicates for each week the total number of bank changes due to financial distress for all reasons. The series *Suspensions Due to Closure of Correspondent* indicates for each week the number of banks for which the principal cause of suspension was the closure of a correspondent. Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Record Administration.

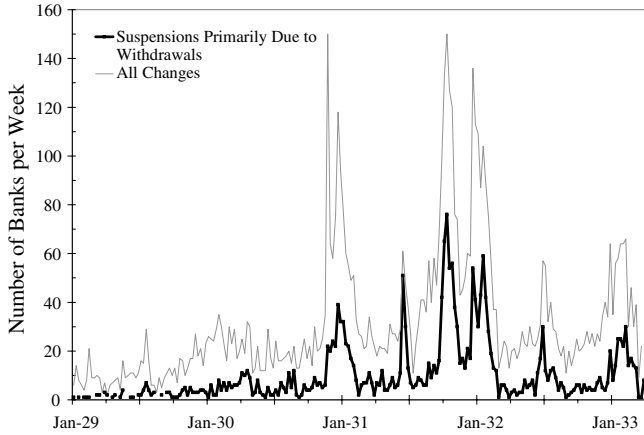


Fig. 5. Suspensions Due to Heavy Withdrawals (as primary cause), January 1929–March 1933. Definitions: The series *All Changes* indicates for each week the total number of bank changes due to financial distress for all reasons. The series *Suspensions Primarily Due to Withdrawals* indicates for each week the number of banks for which the principal cause of suspension was heavy withdrawals. Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Records Administration.

significant cause of suspensions. Thereafter, withdrawals continued to cause suspensions during periods of acute banking distress, but the ratio of withdrawals relative to other causes declined.

Fig. 6 examines the share of suspensions due to correspondents, withdrawals, and assets for the depression as a whole. The left-hand pie chart presents data for banks which suspended operations temporarily. The right-hand pie chart presents data for banks which suspended operations permanently. The shades of gray indicate the percentage of banks suspending for different reasons. The color *black* denotes banks which suspended solely

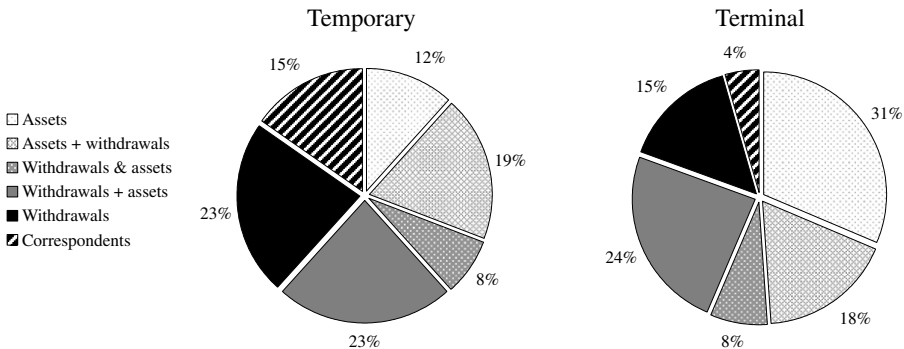


Fig. 6. Causes of Suspensions, Percent of Totals, January 1929–March 1933. Note: The pie charts above display the proportion of suspensions attributed to the 6 causes listed above occurring between January 1929 and March 1933 except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Suspensions attributed to defalcation, mismanagement, multiple, and other causes were excluded from the calculations. Source: National Archives and Records Administration.

due to heavy withdrawals (in key, WITHDRAWALS) but which were in fine financial shape before depositors' demanded repayment of deposits. Most of these banks experienced sudden and severe runs. *Three-quarters gray* denotes banks forced out of operations by heavy withdrawals and also afflicted by asset troubles that exacerbated their condition, but which were not severe enough to have necessitated suspension in the absence of withdrawals (WITHDRAWAL + assets). *One-half gray* indicates banks in double trouble. Problems severe enough to necessitate suspension existed on both sides of the balance sheet (WITHDRAWALS & ASSETS). *One-quarter gray* indicates banks for which slow, doubtful, and worthless assets forced them to close their doors and where withdrawals exacerbated the situation. Examples include (i) banks forced by continuous, slow withdrawals to sell their profitable assets leaving them with a problematic portfolio, and (ii) banks headed towards failure whose demise was accelerated by depositors' demands (ASSETS + withdrawals). *Lightly-dotted gray* indicates banks forced to suspend solely due to problematic assets (ASSETS).

Fig. 6 demonstrates that for the contraction as a whole, heavy withdrawals were the primary cause of more than a half of all temporary suspensions. The closure of counterparties caused another sixth of all temporary suspensions. Asset problems were the primary cause of less than a third of the temporary suspensions. For terminal suspensions, the pattern differed. Asset problems were the primary cause of more than half of all bank suspensions and a contributing cause of another one-quarter. More than eight in ten of the banks which went out of business, in other words, were judged to have problematic assets.

Figs. 4–6 illuminate the causes of the initial surge of bank suspensions during the fall of 1930 and winter of 1931. Before the surge, the pattern of bank suspensions resembled the pattern of failures throughout the 1920s. Banks failed at a steady rate. The principal cause was problems with asset quality. The pattern changed dramatically in November 1930, when the rate of suspension rose suddenly. The collapse of correspondent networks emanating from Caldwell and Company caused the initial increase. Thereafter, runs (or fear of runs) forced scores of banks to close their doors, and adverse circumstances pushed many weak banks into insolvency. The number of runs increased after the failure of The Bank of United States, which was the fourth largest bank in New York City, the eighth largest bank in the nation, and possessed one of the nation's largest branch networks (57 branches).

During the crisis that began in the fall of 1930, forty percent of the institutions that closed their doors to the public soon reopened. Since they reopened without receiving infusions of capital or other outside assistance and survived the remainder of the depression, they were solvent. Moreover, some banks that closed permanently at this time might have been solvent, if they had not been forced to liquidate when similar institutions in neighboring communities dumped assets on the market, driving down prices, and when a national liquidity crunch forced hundreds of banks (which suspended operations) and thousands of others (beset by withdrawals but able to continue without interruption) to sell bonds and securities, which were their most readily marketable assets, and whose prices fell when their supply suddenly expanded. These events are symptoms of a banking system suddenly suffering withdrawals and lacking liquidity.

Lack of liquidity also bedeviled the banking system during the spring of 1931, when suspension surged in Illinois in an event that Friedman and Anna (1971) named the Second Banking Crisis. During this period, the Federal Reserve determined heavy withdrawals to have been the primary cause and slow or frozen assets to have been a contributing cause

for almost all bank suspensions. On most of the forms, Federal Reserve agents stated laconically that the bank closed its doors after depleting its cash reserves. These comments indicate an epidemic of illiquidity. Banks' assets were frozen. Depositors wanted cash. Banks closed when they failed to make ends meet.

The pattern of causation changed as time passed. In September, Britain abandoned the gold standard. Fear that the United States might also devalue its currency triggered the flight of financial assets abroad. In October, to combat the external drain and defend gold, the Federal Reserve raised the discount rate from 1.5 to 3.5% and sold substantial quantities of government securities, which reduced the money supply and raised market interest rates. These actions weakened the financial positions of banks throughout the nation. Depositors grew increasingly apprehensive about the health of depository institutions. Bank suspensions surged once again, and suspension rates remained elevated through the winter of 1932. During this period, banks failed for a wide array of reasons. Heavy withdrawals from healthy banks caused 20 percent of the suspensions. Heavy withdrawals from banks beset by portfolio problems caused 40 percent of the suspensions. Depreciated assets caused 24 percent of the suspensions. For many of these banks, written comments indicated that droughts, declines in agricultural prices, depressions of local industries, and depreciations of bonds' values and securities' prices had reduced the value of the bank's portfolio so far as to necessitate suspension.

During the spring of 1932, the Federal Reserve initiated an open-market purchase program in an attempt to reinvigorate banks and re-inflate the economy. The attempt lasted only a few months. During that period, suspensions fell to their lowest level since the fall of 1930. Withdrawals also eased. Liquidity appears to have been available. The majority of the banks that suspended operations did so principally due to problems on the asset side of their balance sheet.

This pattern held true even during the month of June, when a banking crisis plagued the city of Chicago and surrounding environs. During that panic, some solvent banks experienced runs, but the solvent institutions which suspended operations soon reopened. Devalued assets were the primary cause of most terminal suspensions, while withdrawals were noted as only a contributing condition, if they were mentioned at all. Written comments on the suspension forms indicate that banks were struggling due to declines in the values of securities, bonds, and real estate and losses due to uncollectible and defaulted loans. In other words, during this event, most banks which experienced runs were either insolvent or teetering on the edge of bankruptcy. Illiquidity and runs did not destroy solvent institutions. Real economic forces caused their demise.

In November 1932, Franklin Roosevelt won the presidential election. Expectations that the new Democratic administration would abandon the gold standard induced foreigners to ship funds abroad and Americans to convert deposits to precious metals. The Federal Reserve lacked the resources to counter these external and internal drains. Some districts (for example, Atlanta) extended credit freely to ailing institutions. Other districts (for example, New York) had to choose between aiding banks and preserving their dwindling stocks of gold. Policies lacked consistency across districts, and some districts (for example, Chicago) refused to cooperate with others despite their statutory obligation to do so. Bank suspensions rose steadily during this period. Banks failed for many, and usually multiple, reasons. Three-out-of-four suspensions suffered both asset problems and heavy withdrawals. Only a handful of banks that suspended operations ever reopened, and in every case, their resurrection required outside financial assistance.

In Figs. 7–9, I address a key question concerning the contraction: did (a) illiquidity and contagion or (b) fundamentals and falling asset values cause banks to fail? The answer is that both were important sources of bank distress. Failures due to illiquidity were concentrated during times when bank suspensions surged, during which illiquidity caused between one-third and one-half of all suspensions. Failures caused by declining asset

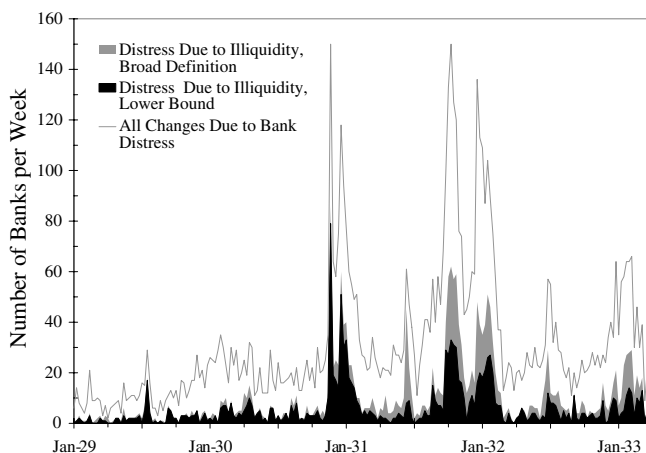


Fig. 7. Suspensions Due to Illiquidity, January 1929–March 1933. Definitions: The series *All Changes Due to Bank Distress* indicates for each week the total number of bank changes due to financial distress for all reasons. For remaining definitions, see Section 4. Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Records Administration.

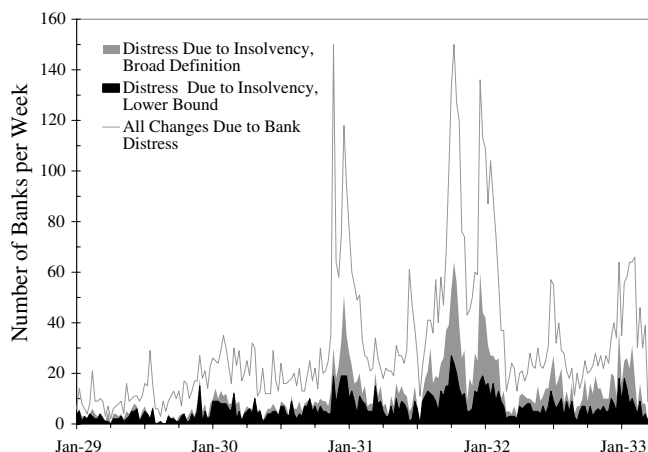


Fig. 8. Suspensions Due to Insolvency, January 1929–March 1933. Definitions: The series *All Changes* indicates for each week the total number of bank changes due to financial distress for all reasons. For remaining definitions, see Section 4. Note: Figures for 1933 include only changes occurring in January–March except those which occurred to institutions closed by government proclamation of banking moratoria or holidays. Source: National Archives and Records Administration.

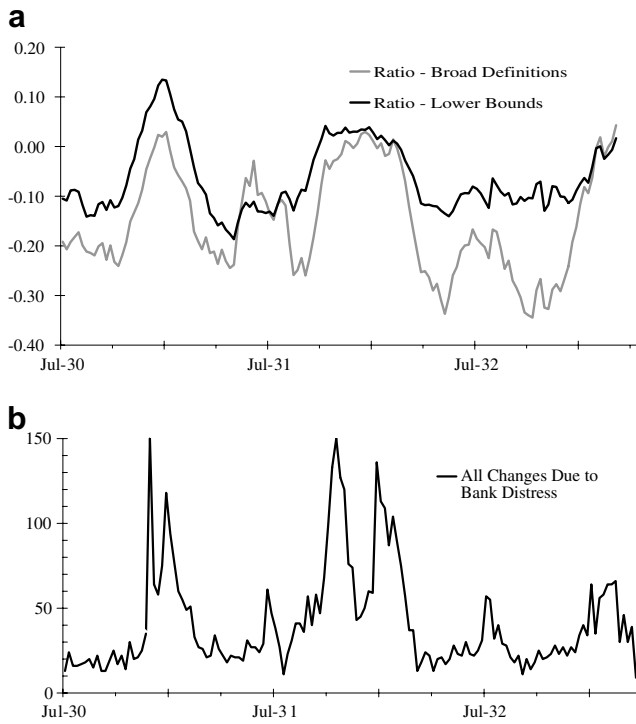


Fig. 9. Illiquidity Relative to Insolvency as the Cause of Bank Suspensions July 1930–March 1933 (a) Share of suspension due to illiquidity minus share of suspensions due to insolvency, Thirteen-week moving averages. (b) Number of bank changes due to distress each week. Definitions and Notes: The plotted series in Fig. 9(a) are $\sum_{t-6}^{t+6} R_t$. See Section 4 for details. Source: National Archives and Records Administration.

values and loan defaults predominated before November 1930, during the lulls between crises, after Britain abandoned the gold standard, and during periods when the Federal Reserve pursued contractionary monetary policy. Outside of panic periods, in other words, contagion and illiquidity caused a small fraction of all suspensions, while insolvency caused the preponderance.

Fig. 7 provides two estimates of banks failing due to contagion and illiquidity. The lower bound is 1,376 or approximately 22% of the 6,063 banks which suspended operations between January 1929 and March 1933. The lower bound estimate consists of all banks which (i) temporarily suspended operations and reopened within three months without outside financial assistance, (ii) permanently suspended operations after experiencing heavy withdrawals and for no other reason (that is examiners judged their assets to be in good condition), (iii) permanently suspended operations after being forced to close by the closure of a counterparty. A more complete accounting is 2,739 or approximately 45 percent of all suspended banks. The estimate consists of banks in categories (i), (ii), and (iii) plus banks which suspended operations permanently (iv) after experiencing heavy withdrawals and possessing assets which examiners judged to be frozen or slow, (v) because they had exhausted their borrowing capacity, were unable to borrow, or failed to obtain loans from their correspondent, or (vi) because they had exhausted their cash reserves. Written comments on the St. 6386 forms identify banks in categories (v) and (vi).

Fig. 8 provides two estimates of the number of banks failing due to insolvency and declining asset values. The lower bound is 1458 or approximately 24% of all banks that suspended operations during the contraction. The lower bound consists of all banks that suspended operations because their assets were slow, doubtful, or worthless and for no other reason. A broader definition yields the figure 3,055 or just over one-half of all suspended banks. This broader accounting includes all banks in the lower bound plus banks whose failure was attributed to (a) the accumulation of slow, doubtful, or worthless paper with heavy withdrawals as a contributing cause, (b) the depreciation of assets, (c) the depreciation of real estate, (d) the depreciation of stocks and bonds, (e) inability to collect on loans, (f) impaired capital, (g) local crop failures, (h) low crop prices, or (j) local industrial or agricultural depression. The banks in categories (b) through (j) were identified by written comments on the St. 6386 forms.

A comparison of Figs. 7 and 8 illuminates differences between the patterns of failures due to illiquidity and insolvency. Fig. 9 highlights these differences. The figure's bottom half indicates the total number of banks experiencing distress in each week from July 1930 to March 1933. The top half indicates the share of banks failing due to illiquidity relative to the share of banks failing due to insolvency. This share is calculated with the following formulas,

$$R_t = (L_t/Z_t) - (S_t/Z_t) \quad (1)$$

And

$$\sum_{t-6}^{t+6} R_t/13 = \sum_{t-6}^{t+6} ((L_t/Z_t) - (S_t/Z_t))/13 \quad (2)$$

Here, R_t is the share under consideration, which I refer to as the illiquidity-insolvency ratio. L_t equals number of suspensions due to illiquidity in week t . S_t equals number of suspensions due to illiquidity in week t . Z_t equals the total number of suspensions for all causes in week t . Note also that $L_t + S_t \neq Z_t$, because a fraction of banks failed for reasons that could not be clearly identified as illiquidity or insolvency, such as defalcations on the part of the management, and a fraction of banks suspended while they suffered from both illiquidity and insolvency.

The sum, $\sum_{t-6}^{t+6} R_t/13$, is a 13-week moving average of the illiquidity-insolvency ratio. The moving average is useful for separating high-frequency oscillations from broader trends. The number of banks in distress oscillated widely from week to week. The mean of the series was 32. The standard deviation of the series was 27. The largest increase from one week to the next was 115. The largest decrease was 86. The illiquidity-insolvency ratio also varied widely from week to week. For the lower-bound estimate, the mean was -0.071 . The standard deviation was 0.179. For the broader definition, the mean was -0.142 . The standard deviation was 0.286. Both series also exhibited trends. The moving averages plotted in the top half of Fig. 9 illuminate these trends. The lower-bound estimate of the illiquidity-insolvency ratio rises during the fall of each 1930, 1931, and 1932. The broader definition rises during the same periods and also during the spring of each year. To focus on these trends, this essay employs a moving average of thirteen weeks, because the periods of heightened distress lasted, on average, for three months (that is, approximately 13 weeks). Series smoothed with different filters and bandwidths yield similar results.

Fig. 9 illuminates several pertinent patterns. First, periods of acute bank distress were correlated with periods when illiquidity was a more common cause of bank distress. Banking crises were, in other words, periods when lack of liquidity forced large numbers of banks to close their doors. The banking crisis in the fall of 1931, in particular, was the peak period of illiquidity relative to insolvency. Second, during non-panic periods, insolvency was the principal threat to depository institutions. When depositors were not withdrawing funds *en masse*, in other words, the majority (and often the preponderance) of the banks which ceased operations did so because their assets no longer covered their liabilities and bankruptcy loomed on the horizon.

These conclusions can be put on a firmer footing by examining the statistical properties of the series. Table 3 indicates the number of banks in distress during weeks when the share of banks suspending due to illiquidity exceeded the share suspending due to insolvency (that is, R_t above 0) and weeks when the insolvent share exceeded the illiquid share (that is, R_t below 0). On average, the number of banks in distress during the weeks in which illiquidity was the principal form of distress exceeded by sixty to eighty percent the number of banks in distress during the weeks in which insolvency was the principal form of distress.

As Fig. 9 and Table 3 suggest, the two series—the number of banks in distress and the illiquidity-insolvency ratio—are highly correlated. The correlations exist for both the raw and smoothed series and for both the lower-bound and broader estimates. For the lower-bound estimate of the insolvency-liquidity ratio, the correlation coefficient (calculated using Pearson's product-moment method) with the number of banks in distress is 0.36. For the lower-bound's moving average, the correlation coefficient is 0.57. For the broader definition, the correlation coefficient is 0.25. For the broader definition's moving average, the correlation coefficient is 0.42.

Table 4 summarizes these correlations. Its four columns report the results of ordinary-least-squares regressions in which the standard errors have been corrected for autocorrelation using the Newey-West procedure. In each case, the dependent variable is the number of banks in distress. The unit of observation is the week, beginning with the seventh week of 1929, continuing for 210 weeks, and ending with the ninth week of 1933. The first six and last six weeks are dropped from the sample to accommodate the thirteen-week moving average. The independent variable in each case is a measure of the illiquidity-insolvency ratio. In all four regressions, the coefficient on the illiquidity-insolvency ratio is significant in statistical and practical terms. I cannot reject the hypothesis that periods

Table 3
Bank Distress and the illiquidity-insolvency ratio, January 1929–March 1933

R_t	# of weeks	# of Banks in distress	
		Average	Standard deviation
Lower bound estimates			
Above 0	69	46.8	37.2
Below 0	138	25.7	17.0
Broad definitions			
Above 0	62	44.4	36.7
Below 0	151	27.5	21.8

Definitions and notes. See Section 4.

Table 4

The Correlation between Bank Distress and the Illiquidity-Insolvency Ratio Ordinary Least Squares Regressions with Newey-West Standard Errors

	(1)	(2)	(3)	(4)
<i>Dependent Variable: Number of Banks in Distress</i>				
<i>Each Week</i>				
Constant	36.8 (5.0)	48.3 (6.8)	36.3 (5.4)	49.0 (9.6)
Independent Variable, illiquidity–insolvency ratio				
Lower bound	59.3 (20.2)			
Lower bound moving average		229.1 (55.2)		
Broad definition			24.6 (10.4)	
Broad definition moving average				112.9 (43.2)
Number of observations	210	210	210	210
R-squared	0.13	0.33	0.06	0.18

Notes. Bold faced indicates significant at the 1% level. Italics indicates significance at the 5% level. Standard errors have been calculated using the Newey-West procedure with lags of up to six weeks.

when larger number of banks experienced distress coincided with periods when illiquidity caused a larger portion of suspensions. The signs and significance levels of the coefficients vary little when the dependent variable is changed to other measures of bank distress, such as the number of terminal bank suspensions, the number of temporary bank suspensions, or the total number of bank suspensions. In sum, the correlation is clear. Periods of increased bank distress were correlated with periods of heightened illiquidity from the fall of 1930 until the winter of 1933.

5. Discussion

Scholars have long debated the causes and consequences of bank failures during the Great Contraction. Some argue that widespread withdrawals, financial contagion, and lack of liquidity were the principal reasons that banks suspended operations. Others argue that declining asset values, large loan losses, and the agricultural and industrial depressions were the principal reason that banks went out of business. Partisans push their theories to the limit. For example, recent work by Calomiris and Joseph (2003) asserts that insolvency was not just the principal or the predominant source of distress, but that insolvency was the only source of distress.

The evidence presented in this essay indicates that mono-causal theories concerning the causes of bank closures are inconsistent with the evidence. Illiquidity was not the sole source of bank distress. Many depository institutions did not experience deposit losses, but failed because the value of their assets declined until the value of their liabilities exceeded the value of their resources. Insolvency was not the sole source of bank distress. Many depository institutions had assets in good condition, but closed their doors because their depositors demanded funds, which the bank could not supply, because it could not quickly convert assets into cash, and it lacked access to sufficient lines of credit.

Both illiquidity and insolvency, in other words, were substantial sources of bank distress. The nature of the banking crisis varied across institutions, time, and place. Different portions of the banking industry experienced different forms of distress. None of the data samples previously studied by scholars has captured the complexity and dynamism of events.

The temporal patterns roughly follow the description of events outlined by [Friedman and Anna \(1971\)](#) with modifications made by [Temin \(1976\)](#) and [Wicker \(1980, 1996\)](#). Before the fall of 1930, bank failures were concentrated in agricultural areas among banks that had experienced large investment losses. The initial banking panic in the fall of 1930—with its cluster of temporary bank suspensions, bank runs, and collapsing correspondent networks—appears to have been a credit crunch triggered by counter-party cascades and propagated by the public's flight from deposits to currency. During later surges in suspensions, distress of depository institutions appears to have been driven in part by depositors' fears for the safety of banks and in part by declines in the values of banks' assets. Between crises, banks that closed their doors typically did so due to problems with their portfolios and fears of impending insolvency.

Why did the nature of the banking crisis change over time? The archival evidence illuminates the initial impetus. During the fall of 1930, Caldwell and Company's correspondent network collapsed. Neighboring correspondent networks collapsed at the same time. Bank runs radiated from the locus of these counter-party cascades. The rate of runs increased after the failure of The Bank of United States. During the spring of 1931, runs struck large numbers of banks in Illinois, particularly several large banking groups centered in the city of Chicago.

What explains the continued contraction of the banking system? Answers to this question can be divided into two classes. The first follows the logic of [Friedman and Anna \(1971\)](#). The liquidity crisis grew like a snowball rolling down a hill. The initial credit crunch forced some banks out of business. Fear of further failures induced depositors to withdraw additional funds, which forced banks to sell more assets, which drove asset prices even lower, which forced more banks to fail, which confirmed depositors' fears in a continuing cycle of self-fulfilling pessimism. The snowball stopped only when Roosevelt declared a national banking holiday in the winter of 1933 and restored confidence in the safety and soundness of the financial system.

The second follows the logic of [Temin \(1989\)](#). Shocks—droughts, tariff wars, contractionary policies, beggar-thy-neighbor devaluations, declines in autonomous consumption and investment—battered the economy in the early 1930s. England's departure from the gold standard was a particularly severe stimulus. These shocks reduced asset values and pushed marginal banks into insolvency. The banking crisis came to an end only after economic recovery began in 1933.

Both of these theories are consistent with the data presented in this essay. Both probably explain a portion of events. Their relative explanatory power remains one of the key unanswered questions concerning the depression. Other questions raised by the evidence presented in this essay are: why did correspondent networks collapse in November and December 1930 but endure even greater shocks during the following three years? Why did some banks which suspended operations reopen for business, while others whose owners and managers also strove to reorganize and reopen, fail to reopen their doors? Why were banks in some states able to avoid suspension by merging with healthy institutions at the last minutes but not in others? Did the policies pursued by state banking departments and Federal Reserve Banks influence the categories and causes of bank distress? Future essays will employ the new archival evidence to answer such questions and to advance our efforts to understand the causes and consequences of banking crises during the Great Contraction.

Acknowledgments

I thank Shagufta Ahmed, Shaista Ahmed, Jacqueline Chattopadhyay, Ching-Yi Chung, Nathan Montgomery, Yuiichi Inomata, Mark Ng, Mitra Pai, Doris Sum, Brandon Tsang, Ian Wagner, and Eve Wang for research assistance. I thank Reid Click, Deborah Kauffman, Ed and Edwin Richardson, and Gloria Richardson for accommodations near the National Archives. I thank Erik Heitfield for the loan of photographic equipment. I thank Marigee Bacolod, Dan Bogart, William Branch, Mark Carlson, Milton Friedman, Michelle Garfinkel, Joseph Mason, Kris Mitchener, Elmus Wicker, two anonymous referees, and the editor of this journal for comments that improved this essay. I thank friends and colleagues for comments, advice, and encouragement.

References

- Bernanke, B.S., 1983. Nonmonetary effects of the financial crisis in the propagation of the great depression. *American Economic Review* 73, 257–276.
- Calomiris, Charles W., Mason, Joseph R., 2003. Fundamentals, panics, and bank distress during the depression. *American Economic Review* 93 (5), 1615–1646.
- Friedman, Milton, Anna, Schwartz, 1971. *A Monetary History of the United States, 1867–1960*. Princeton University Press, Princeton.
- Richardson, Gary, 2006a. The records of the federal reserve board of governors in the national archives of the United States. *Financial History Review* 13 (1), 123–134.
- Richardson, Gary. 2006b. Quarterly Data on the Categories and Causes of Bank Distress during the Great Depression, NBER Working Paper w12715.
- Richardson, Gary. 2006c. Correspondent Clearing and the Banking Panics of the Great Depression, NBER Working Paper w12716.
- Richardson, Gary. 2006d. Bank Distress during the Great Contraction, 1929 to 1933, New Evidence from the Archives of the Board of Governors NBER Working Paper w12590.
- 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 w12591.
- Temin, Peter, 1976. *Did Monetary Forces Cause the Great Depression?* W.W. Norton, New York.
- Temin, Peter, 1989. *Lessons from the Great Depression*. MIT Press, Cambridge.
- White, Eugene, 1984. A reinterpretation of the banking crisis of 1930. *Journal of Economic History* 44 (2), 119–138.
- Wicker, Elmus, 1980. A reconsideration of the causes of the banking panic of 1930. *Journal of Economic History* 40, 571–583.
- Wicker, Elmus, 1996. *The Banking Panics of the Great Depression*. Cambridge University Press, Cambridge.