The banking panics in the United States in the 1930s: some lessons for today

Michael Bordo* and John Landon-Lane**

Abstract In this paper we discuss the lessons learned from the US banking panics in the 1930s for the response by the Federal Reserve to the crisis of 2008. We revisit the debate over illiquidity versus insolvency in the banking crises of the 1930s and provide empirical evidence that the banking crises largely reflected illiquidity shocks. In the recent crisis the Fed under Bernanke had well learned the lesson from the banking panics of the 1930s of conducting expansionary monetary policy to meet demands for liquidity. However, unlike in the 1930s, the deeper problem of the recent crisis was not illiquidity but insolvency and especially the fear of insolvency of counterparties. A number of virtually insolvent US banks deemed too big and too interconnected to fail were rescued by fiscal bail-outs.

Key words: banking panics, illiquidity, insolvency, too big to fail

JEL classification: E52, N12

I. Introduction: the Friedman and Schwartz hypothesis and the subsequent debate

In this paper we raise and answer some questions about the recent financial crisis in light of the experience of the Great Depression of the 1930s. We ask what was similar and what different between now and then, and examine the implications of the 1930s banking panics for policy towards recent banking crises.

The Great Depression was by far the greatest economic event of the twentieth century and comparisons to it were rife during the recent Great Recession. Milton Friedman and Anna

*Rutgers University and NBER, e-mail: bordo@econ.rutgers.edu
**Rutgers University, e-mail: lane@econ.rutgers.edu


doi: 10.1093/oxrep/grq027

© The Authors 2010. Published by Oxford University Press.
For permissions please e-mail: journals.permissions@oup.com.
Schwartz’s *A Monetary History of the United States 1867 to 1960* (1963) has long been viewed as the classic treatment of the Great Depression in the United States. They labelled the downturn in the United States from August 1929 to March 1933 the Great Contraction. Since that event a voluminous literature has debated its causes in the United States and its transmission around the world.

Friedman and Schwartz (1963) challenged the prevailing Keynesian view and attributed the Great Contraction from 1929 to 1933 to a collapse of the money supply by one-third, brought about by a failure of Federal Reserve policy. The story they tell begins with the Fed tightening policy in early 1928 to stem the Wall Street boom. Fed officials believing in the real bills doctrine were concerned that the asset-price boom would lead to inflation. The subsequent downturn beginning in August 1929 was soon followed by the stock-market crash in October. Friedman and Schwartz, unlike Galbraith (1955), did not view the crash as the cause of the subsequent depression. They saw it as an exacerbating factor (whereby adverse expectations led the public to attempt to increase their liquidity) in the decline in activity in the first year of the Contraction.

The real problem arose with a series of four banking panics beginning in October 1930 and ending with Roosevelt’s national banking holiday in March 1933. According to Friedman and Schwartz, the banking panics worked through the money multiplier to reduce the money stock (via a decrease in the public’s deposit-to-currency ratio). The panic in turn reflected what Friedman and Schwartz called a ‘contagion of fear’ as the public, fearful of being last in line to convert their deposits into currency, staged runs on the banking system, leading to massive bank failures. In today’s terms it would be a ‘liquidity shock’. The collapse in money supply in turn led to a decline in spending and, in the face of nominal rigidities, especially of sticky money wages, a decline in employment and output. The process was aggravated by banks dumping their earning assets in a fire sale and by debt deflation. Both forces reduced the value of banks’ collateral and weakened their balance sheets, in turn leading to weakening and insolvency of banks with initially sound assets.

According to Friedman and Schwartz, had the Fed acted as a proper lender of last resort, as it was established to be in the Federal Reserve Act of 1913, then it would have offset the effects of the banking panics on the money stock and prevented the Great Contraction. Since the publication of *A Monetary History*, a voluminous literature has arisen over the issues whether the banking panics were really panics in the sense of illiquidity shocks or whether they reflected endogenous insolvency responses to a recession caused by other forces, such as a collapse of autonomous expenditures or productivity shocks. If the panics really reflected insolvency rather than liquidity shocks, then the case Friedman and Schwartz made that expansionary monetary policy could have avoided the Great contraction would be considerably weakened.

---

1 At the time the consensus view was that the slump was a consequence of the speculative boom of the 1920s. The boom was regarded as a manifestation of deep-seated structural imbalances seen in overinvestment. Indeed, according to the Austrian view which prevailed in the interwar period, depressions were part of the normal operation of the business cycle. Policy prescriptions from this view included tight money, tight fiscal policy, and wage cuts to restore balance. Keynes (1936) rejected these prescriptions and the classical view that eventually a return to full employment would be achieved by falling wages and prices. He attributed the slump to a collapse of aggregate demand, especially private investment. His policy prescription was to use fiscal policy—both pump priming and massive government expenditures. In the post-Second World War era, Keynesian views dominated the economics profession and the explanation given for the depression emphasized different components of aggregate expenditure.
Ben Bernanke, the current Chairman of the Federal Reserve, also attributed the Great Contraction to monetary forces and especially the collapse of the banking system. However, he placed less emphasis on the effects via the quantity theory of money on spending as argued by Friedman and Schwartz, and more on the consequences of the collapse of the banking system in raising the cost of financial intermediation.

Thus the issue of whether illiquidity shocks triggering banking panics was at the heart of the Great Contraction is of crucial importance for the role of monetary policy in dealing with banking crises such as we recently witnessed. Indeed, in the crisis of 2007–8 the Bernanke Fed apparently learned the lesson from Friedman and Schwartz (Bernanke, 2002) by following expansionary monetary and credit policy and to a large extent prevented a repeat of the 1930s experience. Although there was not a classic Friedman and Schwartz banking panic in 2007–8, there was a panic in the shadow banking system and, unlike in the 1930s, many US banks deemed too big and too interconnected to fail were plagued with insolvency and were rescued by fiscal bail-outs.

In this paper we revisit the debate over illiquidity versus insolvency in the Great Contraction. Section II discusses the recent debate and presents some econometric evidence that suggests that illiquidity shocks dominated in the banking panics in 1930 and 1931 while the last panic of 1933 was largely an insolvency event. In section III we examine why the US had so many bank failures and was so prone to banking failures in its history. Sections IV and V compare the financial crises of the 1930s in the US to the crisis of 2007–8. Section VI concludes with some lessons for policy.

II. The recent debate over US banking panics in the 1930s: illiquidity versus insolvency

In this section we survey recent literature on whether the clusters of bank failures that occurred between 1930 and 1933 were really panics in the sense of illiquidity shocks. This has important implications for the causes of the Great Depression. If the clusters of bank failures were really panics, then it would support the original Friedman and Schwartz explanation. If the clusters of bank failures primarily reflected insolvency then other factors, such as a decline in autonomous expenditures or negative productivity shocks (Prescott, 1999), must explain the Great Contraction. We present some econometric evidence largely in support of the Friedman and Schwartz position.

Friedman and Schwartz viewed the banking panics as largely the consequence of illiquidity, especially in 1930–1. Their key evidence was a decline in the deposit–currency ratio, which lowered the money multiplier, money supply, and nominal spending. They describe the panic in the autumn of 1930 as leading to ‘a contagion of fear’ especially after the failure of the Bank of United States in New York City in December. They also discussed the effects of the initial banking panic leading to contagion by banks dumping their earning assets in a ‘fire sale’ in order to build up their reserves. This, in turn, led to the failure of otherwise solvent banks. Wicker (1996) disputes whether the 1930 panic and the spring 1931

---

Friedman and Schwartz panics were national in scope, but agrees with them that all four banking panics were liquidity shocks.

By contrast both Temin (1976) and White (1984), the latter using disaggregated data on a sample of national banks, argued that the original 1930 banking panic was not a liquidity event but a solvency event occurring in banks in agricultural regions in the South and the Midwest which had been weakened by the recession. These small unit banks came out of the 1920s in a fragile state, reflecting declining agricultural prices and oversupply after the First World War. As in Wicker (1980), they identify the locus of the crisis as the collapse, on 7 November 1930, of the Caldwell investment bank holding company of Nashville, Tennessee, a chain bank (in which one holding company had a controlling interest in a chain of banks), and its correspondent network across a half dozen states.

Calomiris and Mason (2003), following the approach taken in Calomiris and Mason (1997) to analyse a local banking panic in Chicago in June 1932, use disaggregated data on all of the individual member banks of the Federal Reserve System to directly address the question of whether the clusters of banking failures of 1930–3 reflected illiquidity or insolvency. Based on a survival-duration model on 8,700 individual banks, they relate the timing of bank failures to various characteristics of the banks as well as to local, regional and national shocks. They find that a list of fundamentals (including bank size, the presence of branch banking, net worth relative to assets as a measure of leverage, reliance on demand debt, market power, the value of the portfolio, loan quality, and the share of agriculture), as well as several macro variables, largely explains the timing of the bank failures. When they add into the regression as regressors the Friedman and Schwartz panic windows (or Wicker’s amendments to them), they turn out to be of minimal significance. Thus they conclude that, with the exception of the 1933 banking panic, which, as Wicker (1996) argued, reflected a cumulative series of state bank suspensions in January and February leading to the national banking holiday on 6 March, that illiquidity was inconsequential.

Richardson (2007) provides a new comprehensive data source on the reasons for bank suspensions from the archives of the Federal Reserve Board of Governors including all Fed member banks and non-member banks (both state and local) from August 1929 to just before the bank holiday in March 1933. He also distinguished between temporary and permanent suspensions. Based on answers to a questionnaire used by bank examiners after each bank suspension, Richardson put together a complete list of the causes of each suspension. The categories include: depositor runs, declining asset prices, the failure of correspondents, mergers, mismanagement, and defalcations. Richardson then classified each bank suspension into categories reflecting illiquidity, insolvency, or both. With these data he then constructed indices of illiquidity and insolvency. His data show that 60 per cent of the suspensions during the period reflected insolvency, 40 per cent illiquidity. Moreover, he shows that the ratio of illiquidity to insolvency spikes during the Friedman and Schwartz (and also Wicker) panic windows (see Figure 1). This evidence in some respects complements the Friedman and Schwartz and Wicker stories and those of Temin and White. During the panics, illiquidity rises relative to insolvency; between the panics insolvency increases relative to illiquidity. Consistent with the Friedman and Schwartz stories, the panics were driven by illiquidity shocks seen in increased hoarding, but after the panics, in the face of deteriorating economic conditions, bank insolvencies continued to rise. This is consistent with the evidence of Temin and White. The failures continued through the contraction until the banking holiday of the week of 6 March 1933 (with the exception of the spring of 1932 while the Fed was temporarily engaged in open-market purchases).
Figure 1: Bank failures and suspensions

Ratio of Bank Suspensions due to Liquidity over All Bank Suspensions

Ratio of Bank Suspensions due to Insolvency over All Bank Suspensions

Difference between Liquidity Ratio and Insolvency Ratio (L/T-I/T)
Richardson (2006) backs up the illiquidity story with detailed evidence on the 1930 banking panic. As described in Wicker (1980), the failure of Caldwell and Co. in November was the signature event of this crisis. Richardson uses his new database to identify the cascade of failures through the correspondent bank networks based on the Caldwell banks. During this period, most small rural banks maintained deposits on reserve with larger city banks that, in turn, would clear their cheques through big city clearinghouses and/or the Federal Reserve System. When Caldwell collapsed, so did the correspondent network. Moreover, Richardson and Troost (2009) clearly show that when the tidal wave from Caldwell hit the banks of the state of Mississippi in December, the banks in the southern half of the state, under the jurisdiction of the Federal Reserve Bank of Atlanta, fared much better (had a lower failure rate) than those in the northern half, under the jurisdiction of the Federal Reserve Bank of St Louis. The Atlanta Fed followed Bagehot’s Rule, discounting freely the securities of illiquid but solvent member banks. The St Louis Fed followed the real bills doctrine and was reluctant to open the discount window to its member banks in trouble. This pattern holds up when the authors control for fundamentals using a framework like that in Calomiris and Mason (2003).3

Finally, Christiano et al. (2003) build a dynamic stochastic general equilibrium (DSGE) model of the Great Contraction, incorporating monetary and financial shocks. They find that the key propagation channels explaining the slump were the decline in the deposit–currency ratio, amplified by Bernanke et al.’s (1996) financial accelerator. The liquidity shock reduced funding for firms, lowering investment and firms’ net worth. At the same time, the increased currency hoarding reduced consumption expenditure. Their simulations, like those of McCallum (1990) and Bordo et al. (1995) show that expansionary open-market purchases could have offset these shocks.

In sum, the debate over illiquidity versus insolvency in the failures of US banks hinges on the use of aggregate versus disaggregated data. Aggregate data tend to favour illiquidity and the presence of and importance of banking panics in creating the Great Contraction. Disaggregated data tend to focus on insolvency driven by the recession and to downplay the role of the panics in creating the Great Contraction. However, the recent more comprehensive data unearthed by Richardson, as well as the Christiano et al. model, suggest that the original Friedman and Schwartz story may well prevail.

(i) Empirical evidence

In Bordo and Landon-Lane (2010) a vector autoregression (VAR) analysis of the determinants of bank failures is undertaken using the data of Richardson (2007). The VAR used includes the following six variables: bank failures/suspensions due to illiquidity; banks failures/suspensions due to insolvency; total bank failures/suspensions; the growth rate of money supply; the change in the unemployment rate; and the quality spread. The variables are ordered as listed above so that bank failures/suspensions will contemporaneously affect money, unemployment, and the quality spread.

The most important assumption is the ordering of the bank failures/suspensions due to illiquidity series before the bank failures/suspensions due to insolvency series in the VAR. In Richardson (2007) banks that fail or are suspended for reasons of illiquidity are

---

3 Carlson (2008) shows that during the panic, banks that would otherwise have merged with stronger banks rather than fail were prevented from doing so.
It is possible to imagine a situation where a bank run (an illiquidity shock) may cause banks that are otherwise solvent to fail due to illiquidity. Insolvent banks may also be caught up in the bank run and therefore it is natural to think that bank failures due to illiquidity will contemporaneously affect banks failures/suspensions due to insolvency.4

The failure of insolvent banks would not immediately affect illiquid but otherwise solvent banks, at least in the short run. However, the solvency shock may also cause, through contagion, a run on otherwise healthy banks, especially if there was a run-up of closures of insolvent banks preceding the bank run. Our identifying assumption is that if the insolvency shock causes a bank run then this will happen with a time lag. That is, the identifying assumption is that the illiquidity shock might cause some insolvent banks to fail contemporaneously, whereas the insolvency shock will lead to failures due to illiquidity only with a lag. The final variable is total bank failures and is not exactly equal to the sum of the previous two bank failure series. This is because not all bank failures are attributed to illiquidity or insolvency, as noted in the previous paragraph.

The ordering we choose for the last three variables is the following: the first variable is the growth rate of the money supply; the second is the change in the unemployment rate; and the third is the quality spread. The triangular ordering we use implies, then, that each variable contemporaneously affects each variable ordered below it, but not any variable ordered above it in the vector. Thus a change in the growth rate of money supply contemporaneously affects the change in unemployment and the quality spread, while the change in the unemployment rate contemporaneously affects the quality spread. These variables then affect bank failures/suspensions with a lag.

Thus we identify six shocks in total that we interpret as follows: the first shock, is the illiquidity shock; the second is the insolvency shock; while the third is a bank failure/suspension residual shock. It is the shock to banks failures/suspensions that cannot be attributed to either illiquidity or insolvency. The next three shocks are a money growth rate shock, an aggregate real shock to unemployment that is orthogonal to the money growth shock, and a shock to the quality spread that is orthogonal to all the previous shocks. We might consider this shock to be a credit shock. Note that we cannot with this specification identify supply or demand shocks for both the money shocks and aggregate real shocks.

The reduced form VAR is estimated using ordinary least squares with two lags of each variable in each equation. It was determined that the money supply and unemployment series were non-stationary, so that all variables enter the VAR in log-levels except for money supply and the unemployment rate which enter as first differences of the log-level. The sample period used (based on Richardson’s data) finished in February 1933 and so does not include the period of the bank holiday starting on 6 March 1933.

The lag structure was determined using various information criteria and the standard sequential likelihood ratio tests. All information criteria and the sequential likelihood ratio test suggest two lags should be included. The results suggest that there are a large number of significant contemporaneous relationships between the variables. All coefficients are significant except for the effect of the illiquidity and insolvency shocks on the growth rate of money supply.

4 These technically insolvent banks may still be operating owing to asymmetric information between depositors and bank operators.
Orthogonalized impulse response functions were computed, using the ordering described above, in order to determine the effect of the identified shocks on the variables of the system. It is clear from the results reported in Bordo and Landon-Lane (2010) that the illiquidity shock has a large and persistent effect on total bank failures/suspensions. The forecast error variance decompositions show that the illiquidity shock accounts for roughly 50 per cent of the forecast error, with the insolvency shock only accounting for 16 per cent. Thus it appears that the illiquidity shock is very important for explaining total bank failures/suspensions. Money supply shocks also have an effect on total bank failures/suspensions. A positive shock to money growth has the effect of lowering bank failures/suspensions. This result is persistent and occurs for each of the bank failure/suspension series. The effect of money is especially strong and persistent for the bank failures/suspensions due to insolvency series. This result suggests that monetary policy aimed at increasing the growth rate of money may have helped to mitigate some of the bank failures/suspensions that occurred during the early 1930s. This result reinforces the views of McCallum (1990), Bordo et al. (1995), and Christiano et al. (2003).

The other impulse response functions, reported in Bordo and Landon-Lane (2010) support the identification assumption that bank failures/suspensions contemporaneously affect money, unemployment, and the quality spread, and that these three variables feed back into the bank failures/suspensions series with some lag. The illiquidity shock is seen to have a strong direct and indirect effect on total bank failures/suspensions due to insolvency, with the illiquidity shock affecting money supply which in turn affects bank failures/suspensions due to insolvency.

The impulse response functions together with the variance decompositions show that the illiquidity shock is very important in explaining the bank failures/suspensions during the early 1930s. In order to determine if the illiquidity shocks played a role during the particular financial crisis windows identified by Friedman and Schwartz (1963), we now turn to historical decompositions. Figure 2 contains historical decompositions for the total bank failures/suspensions series. Each panel of Figure 2 contains a simulated total bank failures/suspensions series under the hypothesis that only one orthogonalized shock was driving the stochastic component of the data. Thus the panel titled illiquidity shock shows the generated series if there was only an illiquidity shock.

The results of the historical decompositions clearly point to the illiquidity shock playing a significant role in the bank failures during the Friedman and Schwartz crisis windows. The most obvious case is during the first window from October 1930 to January 1931. Here the historical decomposition for the illiquidity series almost completely follows the actual data. The other shocks do not explain this first crisis window at all. For the next two crisis windows that take up most of 1931 the illiquidity shock does generate series that follow the actual series quite well. During these periods the money shock and the insolvency shock generate series that peak around the right time, but they do not generate series that closely follow the actual total bank failures/suspensions series. The only crisis window that the insolvency shock does predict well appears to the final crisis of early 1933. In this case it does appear that the financial crisis in 1933 is more an insolvency story then an illiquidity story.

To summarize, we have estimated a VAR and used a triangular ordering to identify a set of shocks including illiquidity and insolvency shocks. The impulse response functions obtained from this orthogonalized VAR make sense and show that the illiquidity shock is an important shock for explaining the observed bank failures/suspensions series. Further, the historical decompositions show that the financial crises of late 1930 and all of 1931 are well modelled as illiquidity crises. The financial crisis of 1933 is better explained as an insolvency crisis.
Thus the evidence suggests that the banking panics of the 1930s were largely a liquidity event which massive monetary expansion could have avoided. The key policy lesson from the 1930s experience is that central banks need to attach prominent importance to their role as lenders of last resort. As we discuss in sections IV and V below, the Fed and other central banks learned this lesson well in the crisis of 2007–8.
III. Why did the US have so many banking panics?

We have argued that the signature event in the US Great Contraction was the series of banking panics from 1930 to 1933. But this was nothing new in US financial history. From the early nineteenth century until 1914, the US had a banking panic every decade. There is a voluminous literature on US financial stability, and the lessons that come from that literature are that the high incidence of banking instability reflected two forces: unit banking and the absence of an effective lender of last resort.

(i) Unit banking

Fear of the concentration of economic power largely explains why states generally prohibited branch banking and why since the demise of the Second Bank of the United States in 1836, there was until quite recently no interstate banking (White, 1983). Unit banks, because their portfolios were geographically constrained, were highly subject to local idiosyncratic shocks. Branching banks, especially those which extended across regions, can better diversify their portfolios and protect themselves against local/regional shocks.

A comparison between the experience of the US and Canadian banking systems makes the case (Bordo et al., 1996). The US until the 1920s has had predominantly unit banking and until very recently a prohibition on interstate banking. Canada since the late nineteenth century has had nationwide branch banking. Canada only adopted a central bank in 1934. The US established the Fed in 1914. Canada has had no banking panics since Confederation in 1867, the US has had nine. However, the Canadian chartered banks were always highly regulated and operated very much like a cartel under the guidance of the Canadian Bankers Association and the Department of Finance.

(ii) A lender of last resort

From the demise of the Second Bank of the United States until the establishment of the Federal Reserve in 1914, the US had not had anything like a central bank to act as a lender of last resort, into which the Bank of England had evolved during the nineteenth century (Bordo, 2007). Clearinghouses, established first in New York City in 1857 and other major cities later, on occasion acted as a lender of last resort by pooling the resources of the members and issuing clearinghouse loan certificates as a substitute for scarce high-powered money reserves. However, on several prominent occasions before 1914 the clearinghouses did not allay panics (Timberlake, 2002). Panics were often ended in the National Banking era by the suspension of convertibility of deposits into currency. Also the US Treasury on a few occasions performed lender-of-last-resort functions.

The Federal Reserve was established to serve (among other functions) as a lender of last resort but, as documented above, failed in its task between 1930 and 1933. Discount window lending to member banks was at the prerogative of the individual Federal Reserve banks and, as discussed above, some Reserve banks did not follow through. Moreover, until the establishment of the National Credit Corporation in 1931 (which became the Reconstruction Finance Corporation in 1932), there was no monetary authority to provide assistance to non-member banks (Wicker, 1996). Wicker effectively argues that the panics pre-1914 were
always centred in the New York money market and then spread via the vagaries of the National banking system to the regions. The New York Fed, according to him, learned the lesson of the panics of the national banking system and did prevent panics from breaking out in New York City during the Great Contraction. But, as he argues, it did not develop the tools to deal with the regional banking panics which erupted in 1930 and 1931.

(iii) Recent evidence

There is considerable empirical evidence going back to the nineteenth century on the case linking unit banking to failures and panics (White, 1983). Cross-country regression evidence in Grossman (1994, 2010) finds that during the 1930s countries which had unit banking had a greater incidence of banking instability than those which did not. For the US, Wheelock (1995) finds, based on state- and county-level data, that states that allowed branching had lower bank failure rates than those which did not. However Carlson (2004) (also Calomiris and Mason, 2003) find, based on a panel of individual banks, that state branch banks in the US were less likely to survive the banking panics. The reason Carlson gives is that while state branch banks can diversify against idiosyncratic local shocks better than can unit banks, they were still exposed to the systemic shocks of the 1930s. He argues that branch banks used the diversification opportunities of branching to increase their returns but also followed more risky strategies such as holding lower reserves.

Carlson and Mitchener (2009) show, based on data on Californian banks in the 1930s (California was a state that allowed branch banking), that the entry of large branching networks, by improving the competitive environment actually improved the survival probabilities of unit banks. They explain the divergent results between studies based on individual banks and those based on state- and county-level data by the argument that the US banking system would have been less fragile in the 1930s had states allowed more branching, not because branch banks would have been more diversified but because the system would have had more efficient banks.

The recent financial crisis, although not a classic banking panic, did exhibit a large number of bank failures (mostly in small banks, as in the past, although there were a few large ones such as Countrywide). Most of the bank failures were resolved by the Federal Deposit Insurance Corporation (FDIC) set up in the aftermath of the Great Contraction. So, again, some lessons were learned from the 1930s experience. However, unlike in the 1930s, as we discuss below, a number of very large banks which became insolvent and were deemed too big and too interconnected to fail were bailed out by massive capital injections and partial nationalization.

IV. A comparison of the financial crisis in the US to the 2007–8 crisis

Many people have invoked the experience during the Great Contraction, and especially the banking crises of 1930–3, as a good comparison to the financial crisis and Great Recession of 2007–9. In several descriptive figures in this section we compare the behaviour of some key variables between the two events. We demarcate the crisis windows in the Great Con-
traction using Friedman and Schwartz’s dates. For the recent period we use Gorton’s (2010) characterization of the crisis as starting in the shadow bank repo market in August 2007 (dark grey shading) and then changing to a panic in the Universal banks after Lehman failed in September 2007 (light grey shading). In most respects, e.g. the magnitude of the decline in real GDP and the rise in unemployment (see Figures 3 and 4), the two events are very
different, but there are some parallels between recent events and the 1930s. In Figure 3 we report real GNP for the 1930s and for 2007–9, normalized to be 100 at the start of each period. It is quite clear that the contraction in late 2007 was mild (only about 5 per cent peak to trough) relative to the Great Contraction in the 1930s (roughly 35 per cent peak to trough). The same is clear for unemployment which is depicted in Figure 4. Unemployment
rose from near 0 per cent at the start of the Great Contraction to slightly over 25 per cent by the end of the contraction whereas the rise in unemployment from 4 to 10 per cent for the most recent contraction is small in comparison.

As discussed above, the signature of the Great Contraction was a collapse in the money supply brought about by a collapse in the public’s deposit–currency ratio, a decline in the
banks deposit–reserve ratio, and a drop in the money multiplier (see Figures 5–7). In the recent crisis M2 did not collapse; indeed, it increased reflecting expansionary monetary policy. Moreover, the deposit–currency ratio did not collapse in the recent crisis, it rose. There were no runs on the commercial banks because depositors knew that their deposits were...
protected by federal deposit insurance, which was introduced in 1934 in reaction to the bank runs of the 1930s. The deposit–reserve ratio declined, reflecting an increase in banks’ excess reserves induced by expansionary monetary policy, rather than a scramble for liquidity, as in the 1930s. The money multiplier declined in the recent crisis, largely explained by a massive expansion in the monetary base reflecting the Fed’s doubling of its balance sheet in 2008.
Moreover, although a few banks failed recently, they were minuscule relative to the 1930s, as were deposits in failed banks relative to total deposits (see Figure 8).\footnote{The large spike in 1933 largely represents the bank holiday of 6–10 March in which the entire nation’s banks were closed and an army of examiners determined whether they were solvent or not. At the end of the week one-sixth of the nation’s banks were closed. The relatively large spike in 2008 in the deposits in the failed banks series reflected the failure and reorganization by the FDIC of Countrywide bank. Compared to the case in the 1930s failures, there were no insured depositor losses. With respect to the number of failed banks the current crisis was small with the maximum number of failed banks in the recent crisis being 25 compared with maximums 20 and 100 times greater during the banking crises of the 1930s.}

Thus the recent financial crisis and recession did not constitute a pure Friedman and Schwartz money story. It was not driven by an old-fashioned contagious banking panic. But, as in 1930–3, there was a financial crisis. It reflected a run in August 2007 on the shadow banking system, which was not regulated by the central bank nor covered by the financial safety net. According to Eichengreen (2008), its rapid growth was a consequence of the repeal in 1999 of the Depression-era Glass–Steagall Act of 1935, which had separated commercial from investment banking. These institutions held much lower capital ratios than the traditional commercial banks and hence were considerably more prone to risk. When the crisis hit, they were forced to engage in major deleveraging, involving a fire sale of assets into a falling market, which, in turn, lowered the value of their assets and those of other financial institutions. A similar negative feedback loop occurred during the Great Contraction, according to Friedman and Schwartz.

According to Gorton (2010), the crisis centred in the repo market (sale and repurchase agreements) which had been collateralized by opaque (subprime) mortgage-backed securities by which investment banks and some universal banks had been funded. The repo crisis continued through 2008 and then morphed into an investment/universal bank crisis after the failure of Lehman Brothers in September 2008. The crisis led to a credit crunch, which led to a serious, but, compared to the Great Contraction, not that serious, recession (see Figures 3 and 4). The recession was attenuated in 2009 by expansionary monetary and fiscal policy.

Finally, Figure 9 compares the Baa 10-year composite Treasury bond spread between the two historical episodes. This spread is often used as a measure of credit-market turmoil (Bordo and Haubrich, 2010). As can be seen, the spike in the spread in 2008 is not very different from that observed in the early 1930s.

V. The recent crisis in more detail

The crisis occurred following 2 years of rising policy rates. Its causes include: major changes in regulation, lax regulatory oversight, a relaxation of normal standards of prudent lending, and a period of abnormally low interest rates. The default on a significant fraction of subprime mortgages produced spillover effects around the world, via the securitized mortgage derivatives into which these mortgages were bundled, to the balance sheets of investment banks, hedge funds, and conduits (which are bank-owned but off their balance sheets) which intermediate between mortgage and other asset-backed commercial paper and long-term securities. The uncertainty about the value of the securities collateralized by these mortgages had the effect of spreading uncertainty about the soundness of loans for leveraged buy-outs through the financial system. All of this led to the freezing of the interbank lending...
market in August 2007 and substantial liquidity injections by the Fed and other central banks.

The Fed then both extended and expanded its discount-window facilities and cut the federal funds rate by 300 basis points. The crisis worsened in March 2008 with the rescue of Bear Stearns, an investment bank, by J. P. Morgan, backstopped by funds from the Federal Reserve. The rescue was justified on the grounds that the exposure of Bear Stearns to coun-

Figure 8: Deposits in failed banks as a proportion of total deposits
terparties was so extensive that a worse crisis would follow if it were not bailed out. The March crisis also led to the creation of a number of new discount-window facilities whereby investment banks could access the window and which broadened the collateral acceptable for discounting. The next major event was a Federal Reserve–Treasury bail-out and partial nationalization of the insolvent government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, in July 2008, on the grounds that they were crucial to the functioning of the mortgage market.

**Figure 9:** Quality spread (Baa 10-year T-Bill)
Events took a turn for the worse in September 2008 when the Treasury and Fed allowed the investment bank Lehman Brothers to fail, in order to discourage the belief that all insolvent institutions would be saved, in an attempt to prevent moral hazard. It was argued that Lehman was both in worse shape and less exposed to counterparty risk than Bear Stearns. The next day the authorities bailed out and nationalized the insurance giant AIG, fearing the systemic consequences for collateralized default swaps (insurance contracts on securities) if it were allowed to fail. The fall-out from the Lehman bankruptcy then turned the liquidity crisis into a fully fledged global credit crunch and stock-market crash, as interbank lending effectively seized up on the fear that no banks were safe.

In the ensuing atmosphere of panic, along with Fed liquidity assistance to the commercial paper market and the extension of the safety net to money market mutual funds, the US Treasury sponsored its Troubled Asset Relief Plan (TARP), whereby $700 billion could be devoted to the purchase of heavily discounted mortgage-backed and other securities to remove them from the banks’ balance sheets and restore bank lending. As it later turned out, most of the funds were used to recapitalize the banks.

In early October 2008 the crisis spread to Europe and to the emerging-market countries as the global interbank market ceased functioning. The UK authorities responded by pumping equity into British banks, guaranteeing all interbank deposits and providing massive liquidity. The EU countries responded in kind. And on 13 October 2008 the US Treasury followed suit with a plan to inject $250 billion into the US banks, to provide insurance of senior interbank debt and unlimited deposit insurance for non-interest-bearing deposits. These actions ended the crisis. Expansionary Federal Reserve policy at the end of 2008, lowering the funds rate close to zero, followed by a policy of quantitative easing: the open-market purchases of long-term Treasury bonds and mortgage-backed securities finally attenuated the recession by the summer of 2009.

Unlike the liquidity panics of the Great Contraction, the deepest problem facing the financial system was insolvency. This was only recognized by the Fed after the September 2008 crisis. The problem stemmed from the difficulty of pricing securities backed by a pool of assets, whether mortgage loans, student loans, commercial paper issues, or credit card receivables. Pricing securities based on a pool of assets is difficult because the quality of individual components of the pool varies and, unless each component is individually examined and evaluated, no accurate price of the security can be determined.

As a result, the credit market, confronted by financial firms whose portfolios were filled with securities of uncertain value, derivatives that were so complex the art of pricing them had not been mastered, was plagued by the inability to determine which firms were solvent and which were not. Lenders were unwilling to extend loans when they could not be sure that a borrower was creditworthy. This was a serious shortcoming of the securitization process that was responsible for the paralysis of the credit market.

Finally, another hallmark of the recent crisis which was not present in the Great Contraction is that the Fed and other US monetary authorities engaged in a series of bail-outs of incipient and actual insolvent firms deemed too systemically connected to fail. These included Bear Stearns in March 2008, the GSEs in July, and AIG in September. Lehman Brothers had been allowed to fail in September on the grounds that it was both insolvent and not as systemically important as the others and, as was stated well after the event, that the Fed did not have the legal authority to bail it out. The extension of the ‘too big to fail’ doctrine, which had begun in 1984 with the bail-out of Continental Illinois bank, may be the source of future crises.
VI. Conclusion: some policy lessons from history

In this paper we have re-examined the issue of the role of the banking panics between 1930 and 1933 in creating the Great Contraction. We focused on the debate between those following the Friedman and Schwartz view that the banking crises were illiquidity shocks and those following the approach of Temin and others who view the clusters of banking failures as not being liquidity-driven panics but insolvencies caused by the recession. Our survey of the evidence suggests that the banking crises did reflect contagious illiquidity, but also that endogenous insolvency was important between the panics. Bank failures regardless of their genesis contributed to the depression by reducing the money supply and by crippling the credit mechanism.

In Bordo and Landon-Lane (2010) we showed illiquidity played a major role in the financial crises of late 1930 and all 1931. We estimated a VAR and used a triangular ordering to identify a set of shocks including illiquidity and insolvency shocks. The impulse response functions obtained from this orthogonalized VAR make sense and show that the illiquidity shock is an important shock for explaining the observed bank failures/suspensions series. Further, the historical decompositions show that the financial crisis of late 1930 and all of 1931 are well modelled as illiquidity crises. The financial crisis of 1933 is better explained as an insolvency crisis.

The Federal Reserve learned the Friedman and Schwartz lesson from the banking panics of the 1930s of the importance of conducting expansionary open-market policy to meet all of the demands for liquidity (Bernanke, 2002). In the recent crisis the Fed conducted highly expansionary monetary policy in the autumn of 2007 and from late 2008 to the present. Also, based on Bernanke’s (1983) view that the banking collapse led to a failure of the credit-allocation mechanism, the Fed, in conjunction with the Treasury, developed a plethora of extensions to its discount window referred to as credit policy (Goodfriend, 2009) to encompass virtually every kind of collateral in an attempt to unclog the blocked credit markets.

Some argue that for the first three quarters of 2008 Fed monetary policy was actually too tight, seen in a flattening of money growth and the monetary base and high real interest rates (Hetzel, 2009). Although the Fed’s balance sheet surged, the effects on high-powered money were sterilized. This may have reflected concern that rising commodity prices at the time would spark inflationary expectations. By the end of the third quarter of 2008 the sterilization ceased, as evidenced by a doubling of the monetary base.

The Fed’s credit policy involved providing credit directly to markets and firms the Fed deemed most in need of liquidity, and exposed the Fed to the temptation to politicize its selection of the recipients of its credit (Schwartz, 2008). In addition, the Fed’s balance sheet ballooned in 2008 and 2009 with the collateral of risky assets including those of non-banks. These assets were in part backed by the Treasury. The Fed also worked closely with the Treasury in the autumn of 2008 to stabilize the major banks with capital purchases and stress testing. Moreover, the purchase of mortgage-backed securities in 2009 (quantitative easing) combined monetary with fiscal policy. These actions, which many argue helped reduce the spreads and reopen the credit channels, impinged upon the Fed’s independence and created problems for the Fed in the future (Bordo, 2010).

As discussed in section V, the deepest problem of the recent crisis, however, was not illiquidity, as it was in the 1930s, but insolvency, and especially the fear of insolvency of counterparties. This has echoes in the correspondence-banking-induced panic of November 1930 (Richardson, 2006), but very different from the 1930s. The too-big-to-fail doctrine, which had developed in the 1980s, ensured that the monetary authorities would bail out
insolvent large financial firms which were deemed too interconnected to fail. This is a dramatic departure from the original Bagehot’s rule prescription to provide liquidity to illiquid but solvent banks. This new type of systemic risk (Tallman and Wicker, 2009) raises the spectre of moral hazard and future financial crises and future bail-outs.

(i) Policies to prevent the next crisis

The crisis of 2007–8 had similarities to the 1930s experience in that there was a panic in the shadow banking system and the repo market in 2007, as argued by Gorton (2010), but also in investment banks and the universal banking system after Lehman failed in September 2008. But it was not a classic contagious banking panic. The decision to bail out large interconnected financial institutions in the autumn of 2008 does not have much resonance in the 1930s experience. The closest parallel from the 1930s was the Bank of United States, which failed in December 1930. It was one of the largest banks in the country but it was insolvent and it was allowed to fail (Lucia, 1985).

A key concern from the bail-outs of 2008 is that, in the future, the too-big-to-fail doctrine will lead to excessive risk-taking by such firms and future crises and bail-outs. This was a major concern in the debate leading to the recent Dodd Frank Wall Street Reform and Protection Act, passed in July 2010. The Act attempted to address the too-big-to-fail problem by establishing a Financial Stability Oversight Council made up of members from the Federal Reserve Board, The Treasury, the Securities and Exchange Commission, and a number of other financial agencies. The Council was charged with identifying and responding to emerging risks throughout the financial system. The Council would make recommendations to the Federal Reserve to impose increasingly strict rules for capital and leverage and other requirements to prevent banks from becoming too large and systemically exposed. It remains to be seen whether it will be effective in preventing future crises.

References


McCallum, B. (1990), ‘Could a Monetary Base Rule have Prevented the Great Depression?’, Journal of Monetary Economics, 26, 3–21.


