AN INTRODUCTION TO FINANCIAL CRISES*

Franklin Allen
University of Pennsylvania

and

Douglas Gale
New York University

August 14, 2007

* This essay was prepared for the volume on Financial Crises that we edited in THE INTERNATIONAL
LIBRARY OF CRITICAL WRITINGS IN ECONOMICS – Series Editor: Mark Blaug published by
Edward Elgar.
Crisis have been a feature of the financial landscape for hundreds of years. They often appear without warning as the sub-prime mortgage crisis of August 2007 illustrates.\(^1\) A good example of a crisis is provided by the Asian crisis of 1997-98. The ‘Dragons’ (Hong Kong, Singapore, South Korea, and Taiwan,) and the ‘Tigers’ (Indonesia, Malaysia, the Philippines, and Thailand) were models of successful economic development. From the early 1950’s until the eve of the crisis in 1997, their economies grew at high rates. In 1997 the Thai baht came under sustained pressure and, on July 2, the government stopped defending it. The value of the currency immediately fell 14 percent in the onshore market and 19 percent in the offshore market (Fourçans and Franck, 2003, Chapter 10). This marked the beginning of the Asian financial crises.

The next currencies to come under pressure were the Philippine peso and the Malaysian ringitt. The Philippine central bank tried to defend the peso by raising interest rates. Despite the government’s action it lost $1.5 billion in foreign reserves. On July 11, the government let the peso float and it promptly fell 11.5 percent. The Malaysian central bank stopped defending the ringitt on July 11. The Indonesian central bank stopped defending the rupee on August 14.

The countries known as the Dragons were also affected by the spreading crisis. At the beginning of August, Singapore decided not to defend its currency and by the end of September the Singapore dollar had fallen 8 percent. Taiwan also decided to let their currency depreciate and were not much affected. Hong Kong had a currency board that pegged the exchange rate to the US dollar. The Hong Kong dollar came under attack, but the currency board was able to maintain the peg. Initially, the South Korean won appreciated against other South East Asian currencies; however, in November the won

\(^1\) This was starting at the time of writing this essay.
lost 25 percent of its value. By the time the crises ended, the dollar had appreciated against the Malaysian, Philippine, Thai, South Korean, and Indonesian currencies, by 52, 52, 78, 107, and 151 percent respectively.

Although the turbulence in the currency markets subsided by the end of 1997, the real effects of the crisis continued to be felt throughout the region. Many financial institutions and industrial and commercial firms went bankrupt and output fell sharply. Overall, the crisis was extremely painful for the countries involved.

There are many other examples of crises, which are not confined to emerging markets but occur in developed economies as well. The Scandinavian crises of the early 1990’s are a good example. As we shall see, crises can cause great damage to a country’s economy and result in lost output equivalent to several years GDP. This book of readings provides the reader with an introduction to the literature on this important topic. This introduction provides an overview of the book and reviews each section in turn. A more complete account of the literature and a framework for analyzing crises is developed in Allen and Gale (2007).

I. Crises in Historical Perspective

Bordo, Eichengreen, Klingebiel and Martinez-Peria (2001, reprinted as Reading 1) provide a systematic comparison of recent crises, such as the European Monetary System Crisis of 1992-93, the Mexican crisis of 1994-5, and the Asian crises of 1997-98, with earlier crises. They divide 120 years into four periods.

1. The Gold Standard Era 1880-1913
2. The Interwar Years 1919-1939
3. The Bretton Woods Period 1945-1971


The data set for the first three periods comprises 21 countries and, for the recent period, data for the original 21 as well as an expanded group of 56.

Bordo et al. define a banking crisis as a period of financial distress that is severe enough to result in the erosion of most or all of the capital in the banking system. A currency crisis is defined as a forced change in parity, abandonment of a pegged exchange rate, or an international rescue. A twin crisis consists of both a banking crisis and a currency crisis. Bordo et al. define the duration of a crisis as the amount of time before GDP growth returns to the trend rate of GDP growth for the five years preceding the start of the crisis. Finally, they measure the depth of the crisis by summing the output loss relative to trend for the duration of the crisis.

Figure 1 shows the frequency of crises in the four periods. Using the data for the original 21 countries, it can be seen that the Interwar Years (1919-1939) were the worst of the four periods. This is perhaps not surprising in view of the fact that the Great Depression affected most countries in that period. Banking crises, in particular, were more prevalent during this period than in the other periods.

The Bretton Woods Period (1945-1971) was very different from the other periods. After the Great Depression, policymakers in most countries were determined not to allow such an event to happen again. They imposed severe regulations on banks or brought them under state control to prevent them from taking much risk. As a result, banking crises were almost completely eliminated during the Bretton Woods Period. A twin crisis occurred in Brazil in 1962 but, apart from that episode, there were no banking crises
during the entire period. There were frequent currency crises, but these mostly occurred because macroeconomic policies were inconsistent with the level of the fixed exchange rates set in the Bretton Woods system.

Interestingly, the most benign of the four periods was the Gold Standard Era (1880-1913). Banking crises did occur in this period, but they were infrequent. Currency crises and twin crises also occurred much less frequently than in subsequent periods. The global financial system was fairly open at this time, which suggests that globalization does not inevitably lead to crises. The incidence of crises in the Recent Period was not as high as in the Interwar Years, but was nevertheless fairly high. Banking and twin crises were more frequent in the Recent Period than in every period except the Interwar Years and currency crises were much more frequent. The incidence of crises in the Recent Period is even more striking if we use the sample of 56 countries as the basis of comparison. The countries added to create the larger sample are mostly emerging markets, which suggests that emerging markets are more prone to crises, particularly currency crises, than the older sample of 21 countries.

Table 1 shows the average duration and depth of crises, broken out by type of crisis and by period and sample. Perhaps the most striking feature of Table 1 is the short duration and shallow depth of crises during the Bretton Woods Period. Another distinctive feature is the greater severity of twin crises in terms of lost output. As might be expected, crises were more severe during the Interwar Years than in the other periods. Although they did not last longer than in the other periods, the cumulative loss in output from these crises was higher. During the Gold Standard Era, by contrast with other
periods, the duration and cumulative loss of crises were unremarkable. In the Recent Period, twin crises were characterized by long durations and significant output losses.

Figure 2 shows the effect on recessions of having a bank or currency crisis at the same time. It can be seen that recessions that coincide with crises have a much higher loss of GDP than recessions that do not coincide with crises. This was particularly true in the Interwar Years. Also, the average recovery time is somewhat higher in recessions that coincide with crises than in recessions that do not coincide with crises.

In summary, the analysis of Bordo et al. (2001) suggests the following conclusions.

- Banking crises, currency crises, and twin crises have occurred under a variety of different monetary and regulatory regimes.
- Over the last 120 years, crises have been followed by economic downturns lasting from 2 to 3 years and costing 5 to 10 percent of GDP, on average.
- Twin crises are associated with particularly large output losses.
- Recessions that coincide with crises are more severe than recessions that do not coincide with crises.
- In retrospect, the Bretton Woods Period (1945-1971) appears to be quite special. Countries either regulated bank balance sheets to prevent them from taking on much risk or achieved the same aim through direct ownership of banks. These measures were successful in suppressing banking crises and there was only one twin crisis during this period.
The Interwar Years were also special. Banking crises and currency crises were widespread. Moreover, the output losses from banking and currency crises were severe, especially when they coincided to form a twin crisis.

The Recent Period does indeed appear more crisis-prone than any other period except for the Interwar Years. In particular, it seems more crisis-prone than the Gold Standard Era, the last time that capital markets were globalized as they are now.

Explaining Crises

Two broad approaches have been adopted by economists in their attempt to understand the crises that occurred in these periods. Both approaches have a long history. One view, well expounded by Kindleberger (1978), is that crises occur spontaneously as the result of mob psychology or panic. If everyone expects a crisis and acts as if one is about to occur, then the crisis becomes a self-fulfilling prophecy. Conversely, if no one expects a crisis, this expectation is also self-fulfilling and no crisis occurs. A second view asserts that crises are an intrinsic part of the business cycle and result from shocks to economic fundamentals (see, e.g., Mitchell, 1941). When the economy goes into a recession or depression, asset returns are expected to fall. Borrowers will have difficulty repaying loans and depositors, anticipating an increase in defaults or non-performing loans, will try to protect their wealth by withdrawing banks deposits. Banks are caught between the illiquidity of their assets (loans) and the liquidity of their liabilities (deposits) and may become insolvent. The result is the same as in the panic story, but the cause is different.
Cries in the National Banking Era and Early 1930’s

There is a large empirical literature on historical banking crises, which we briefly touch on here. Sprague (1910) is the classic study of crises in the US during the National Banking Era (1864-1914). It was commissioned by the National Monetary Commission, after the severe crisis of 1907, as part of an investigation of the desirability of establishing a central bank in the US. Friedman and Schwartz (1963) have written a comprehensive monetary history of the US from 1867-1960. Among other things, they argue that banking panics can have severe effects on the real economy. In the banking panics of the early 1930’s, bank distress developed quickly and had a large effect on output. Friedman and Schwartz argue that the crises were panic-based and offer as evidence the absence of downturns in relevant macroeconomic variables prior to the crises.

Gorton (1988, reprinted as Reading 2) conducts an empirical study to differentiate between the panic view and the business cycle view of banking crises, using data from the National Banking Era. He found evidence consistent with the view that banking panics can be predicted by the business cycle. Table 2 shows recessions and crises occurring in the US during the National Banking Era. It also shows the percentage changes in the currency/deposit ratio and in pig iron production (as a proxy for GDP) during each crisis. The five worst recessions, as measured by the change in pig iron production, were all accompanied by crises. Out of the total of eleven recessions, seven were accompanied by crises. Using the liabilities of failed businesses as a leading economic indicator, Gorton shows that crises were predictable events: whenever his leading economic indicator reached a certain threshold, a panic ensued. The stylized facts
uncovered by Gorton thus suggest that, at least during the US National Banking Era, banking crises were triggered by the business cycle rather than some extraneous random variable.

Calomiris and Gorton (1991) consider a broad range of evidence from the US National Banking Era and conclude that the data do not support the panic view. They find that, for the five episodes they focus on, stock prices fell the most during the pre-panic periods, suggesting that the crises were caused by fundamental factors.

Friedman and Schwartz (1963) attach great importance to four banking crises in the early 1930’s in advancing their view that bank failures were the result of panics. They suggest these crises were important ‘shocks’ to the real economy and were significant factors in causing the Great Depression. The four crises were: (i) the crisis in late 1930 following the failure of the Bank of the United States, (ii) the crisis from March to August 1931, (iii) the crisis following Britain’s departure from the Gold Standard, which lasted from September 21, 1931 until the end of the year and (iv) the crisis that lasted from the end of 1932 until the bank holiday declared when President Roosevelt came to office in early 1933.

Calomiris and Mason (2003, reprinted here as Reading 3) undertake a detailed study of the four crises. They develop an econometric model of bank failures using a broad range of data on fundamentals at the local, regional and national level. They estimate a stable model that is able to predict bank failures in terms of fundamentals during the first three crises, suggesting that panics did not play a significant role. The model breaks down in the fourth crisis, however, suggesting that panics may have played
a role there. Overall, their conclusions are in sharp contrast to those of Friedman and Schwarz.

II. Empirical Studies of Crises

As the historical overview above suggests, one of the most important consequences of financial crises is the heavy costs they impose on the real economy. The literature on the costs of crises and their resolution is large (see, e.g., Bordo et al. (2001), Hoggarth et al. (2002), Roubini and Setser (2004), and Honohan and Laeven (2005)). Much of the debate in this literature is concerned with exactly how to measure costs. The early literature focused on fiscal costs, that is, how much it costs the government to recapitalize banks and reimburse insured depositors; however, these ‘costs’ are mostly transfers rather than true economic costs. The later literature focuses more on the loss of output relative to a benchmark such as trend growth rate.

When we look at the time series of the costs of crises, we notice two things: the first is the high average cost; the second is the large variation in costs. Boyd, Kwak and Smith (2005, reprinted here as Reading 4) estimate the average present value of losses in a number of different ways. Depending on the method used, the mean loss measures between 63 percent and 302 percent of the value of real per capita GDP in the year before the crisis starts. The range of losses is also large. In Canada, France, Germany, and the US, which experienced mild, non-systemic crises, there was no significant slowdown in growth; in other countries, the slowdown and discounted loss in output were extremely high. In Hong Kong, for example, the discounted losses were 1,041 percent of the value of real output the year before the crisis of 1997.
It is the large average costs and the high variability of these costs that makes policymakers so averse to crises and willing to go to great lengths to avoid them. However, it is not clear that this policy is always optimal. There are also costs associated with regulation and some crises may not be very costly. In any case, one must weigh the costs and benefits of avoiding crises.

In an important paper, Kaminsky and Reinhart (1999, reprinted as Reading 5) investigate the relationship between banking crises and currency crises. They find that in the 1970’s, when financial systems were highly regulated, currency crises did not coincide with banking crises. After the financial liberalization of the 1980’s, however, currency crises and banking crises became intertwined. In the usual sequence of events, problems in the banking sector are followed by a currency crisis, which in turn exacerbates and deepens the banking crisis. Although banking crises typically precede currency crises, the common cause of both is usually a fall in asset values caused by a recession or weakness in the economy. Often the crisis is part of a boom-bust cycle that follows financial liberalization. It appears to be rare that banking crises and currency crises occur when economic fundamentals are sound.

One of the most interesting aspects of the Asian crises that started with the July 2, 1997 flotation of the Thai baht was how quickly the other countries were affected. Within a short period of time, the crisis had spread to Indonesia, Korea, Malaysia and the Philippines. Similarly, after the start of Russian crisis in August 1998, Hong Kong, Brazil, Mexico, and many other emerging markets were quickly affected. However, in other cases, such as the Brazilian devaluation of February 1, 1999 or the Argentine default and abandonment of convertibility in December 2001, there were very few
spillover effects. Kaminsky, Reinhart, and Vegh (2003, reprinted as Reading 6) ask why international contagion occurs in some cases, but not in others. They identify three factors, which they name the “unholy trinity”, that explain contagion. These are: an abrupt reversal in capital flows, a surprise announcement, and a leveraged common creditor. Their paper provides an interesting theoretical and empirical perspective on the spread of recent crises from country to country.

III. Models of Banking Crises

The first models of banking crises were developed by Bryant (1980, reprinted here as Reading 7) and Diamond and Dybvig (1983, reprinted here as Reading 8). Both models assume a continuum of consumers with random demand for liquidity. Optimal insurance against these liquidity shocks can be provided by deposit contracts that promise depositors a fixed payment depending on whether they withdraw early or late. The fixed liabilities in the form of deposit contracts make the banks vulnerable to bank runs. In the Diamond-Dybvig model, bank runs are triggered by sunspots, whereas in Bryant’s model they are triggered by aggregate loan risk and asymmetric information about loan payoffs. Both papers provide a justification for deposit insurance. Diamond and Dybvig (1983) argue that deposit insurance eliminates the possibility of runs by removing the incentive for patient depositors to join bank runs. Furthermore, there is no cost since, in the absence of runs, there is no need to make insurance payouts to depositors. In Bryant’s model, deposit insurance is desirable because it eliminates incentives to gather costly information that is not socially useful.
Panic-Based Models

Cone (1983) and Jacklin (1987) point out that the liquidity insurance provided by banks in the Diamond-Dybig and Bryant models is impossible if depositors have access to capital markets. In fact, if depositors have access to capital markets, banks cannot provide insurance to depositors. This is a special case of the general observation that access to markets undermines many insurance schemes by allowing the insured to arbitrage against the insurer.

Another crucial assumption of the bank-run models is the so-called ‘sequential-service’ or ‘first-come-first-served’ constraint. An early criticism of the Diamond-Dybvig paper was that banks could survive a bank run by the simple expedient of suspending convertibility. Diamond and Dybvig (1983) responded by arguing that depositors would arrive at the bank sequentially and that the sequential-service constraint would require the banks to pay out as much as possible to the depositors who arrived first. In this way, the bank would exhaust its funds by the time it realized that a bank run was underway and before it could suspend convertibility.

A number of writers have sought to justify the existence of the sequential-service constraint endogenously, rather than by appealing to legal restrictions. Wallace (1988) assumes that the fraction of the population requiring liquidity is random. He also assumes that agents are spatially separated from each other, but are always in contact with the bank. These factors imply that a sequential-service constraint is optimal. Whereas Diamond and Dybvig (1983) were able to demonstrate the optimality of deposit insurance, deposit insurance is not optimal in Wallace’s model. Building on Wallace’s model, Chari (1989) considers what happens if the interbank market does not work well
because of regulatory restrictions of the type in place during the National Banking era in the US. He shows that, in the presence of such regulatory restrictions, banking panics can occur. With a well functioning interbank market, however, runs do not occur.

Calomiris and Kahn (1991, reprinted here as Reading 9) provide another justification for demand deposits and the sequential-service, by showing that they provide depositors with an incentive to monitor the bank. In equilibrium, a fraction of depositors will have an incentive to gather information about the viability of the bank. If the bank is in poor shape, the informed depositors will discover this and withdraw their money. The sudden withdrawals will force the bank to liquidate all of its assets. In order to avoid this outcome, the bank has to maintain the value of its assets and avoid taking excessive risks. Although maintaining idle cash balances is costly for the bank and monitoring is costly for depositors, the entire arrangement is incentive-efficient because it solves the bank’s moral hazard problem.

Diamond and Rajan (2001) provide another example of the optimality of demand deposits and the sequential-service constraint. In their setup, the possibility of bank runs ensures that banks will not renegotiate to extract more rents from entrepreneurs that borrow from the bank.

Models with Real (Intrinsic) Shocks

There is a large variety of models in which bank runs are triggered by real shocks to the economy. In Bryant’s (1980) seminal model, there is asymmetric information about aggregate loan risk. Gorton’s (1985) model assumes that depositors receive a noisy signal about the value of bank assets. If the signal suggests the value of assets is low, a bank run
ensues. Solvent banks suspend convertibility and pay a cost to verify their solvency to investors. Chari and Jagannathan (1988) focus on a signal extraction problem, where part of the population observes a signal about future returns and others must then try to deduce from observed withdrawals whether an unfavorable signal was received or whether liquidity needs happened to be high. Chari and Jagannathan (1988) are able to show bank runs occur not only when the outlook is poor but also when liquidity needs turn out to be high. Jacklin and Bhattacharya (1988) also consider a model where some depositors receive an interim signal about risk. They show that the optimality of bank deposits compared to equities depends on the characteristics of the risky investment. Hellwig (1994) considers a model where the reinvestment rate is random and shows that the risk should be born both by early and late withdrawers. Alonso (1996) demonstrates using numerical examples that contracts where runs occur may be better than contracts which ensure runs do not occur, because bank runs improve risk sharing.

Allen and Gale (1998, reprinted here as Reading 10) develop a model of business cycle risk with symmetric information. They assume that the prospects of banks are observable, but not contractible. Motivated by Gorton’s (1988) empirical findings, they assume that everyone, including depositors, can observe a leading economic indicator that is perfectly correlated with future asset returns. Banks invest in two kinds of assets, a risky, illiquid asset (the long asset) and a safe, liquid asset (the short asset). The short asset can also be held by individuals. Because the long asset is completely illiquid and cannot be liquidated, default causes no deadweight loss and the first-best allocation is achieved through a competitive banking system using demand deposit contracts. Bank
runs are essential to achieving the first best because they introduce the optimal amount of state-contingency into the consumption allocation.

Allen and Gale (1998) consider several extensions of the basic model. If there is a cost to holding the liquid asset outside the banking system, the competitive banking outcome is no longer efficient; however, the first best can achieved if the central bank provides fiat money that can be held by the depositors as a substitute for the short asset. A second extension introduces a market in which the long asset can be sold. The possibility of liquidating the long asset results in a significantly worse outcome. Banks are forced to unload their illiquid assets at fire-sale prices. Although this transfer of value to speculators is ex post efficient, it is inefficient ex ante because of the risk it imposes on the banks and their depositors. Again, efficiency can be restored through a central bank intervention.

Allen and Gale (2004, reprinted here as Reading 11) develop a general equilibrium framework for understanding the normative aspects of crises. This framework is used to investigate the welfare properties of financial systems and to discover conditions under which regulation might improve the allocation of resources. An interesting feature of the Allen-Gale framework is that it explicitly models the interaction of banks and markets. Financial institutions are the main players in financial markets, which allow banks and intermediaries to share risks and liquidity. Individuals do not have direct access to markets; instead, they access markets indirectly by investing in intermediaries. Financial intermediaries and markets play important but distinct roles in the model. Intermediaries provide consumers with insurance against idiosyncratic
liquidity shocks. Markets allow financial intermediaries and their depositors to share risks from aggregate liquidity and asset return shocks.

Financial markets are said to be complete if it is possible for intermediaries to hedge all aggregate risks in the financial markets, for example, if there are state-contingent Arrow securities for every aggregate state of nature. Similarly, the risk-sharing contracts between intermediaries and consumers are said to be complete if the payoffs are explicitly conditioned on aggregate risks, that is, they are contingent on all the aggregate states of nature. An example of an incomplete contract would be something like debt, where the payoff on the contract does not depend explicitly on the aggregate state. Allen and Gale (2004) show that the laissez-faire allocation of resources is *incentive-efficient* if markets are complete and contracts are complete. In this case, Adam Smith’s invisible hand works in the sense that a central planner with the same information could not Pareto-improve on the market’s allocation. Incentive efficiency is the appropriate notion of efficiency because the depositors’ idiosyncratic liquidity shocks are private information, so both intermediaries and the central planner are constrained to use incentive-compatible contracts.

In this idealized world, there is no possibility of market failure and hence no possibility of welfare-improving government intervention. The result is analogous to the First Fundamental Theorem of Welfare Economics, extended to an economy with financial intermediation. There is no possibility of default either, because banks and other intermediaries can use state-contingent markets to make their net liabilities contingent on the aggregate state of nature.
Many of the contracts we observe in practice, such as debt and deposit contracts, are incomplete; however, even in this case there is an analog to the First Theorem. If we assume that the central planner is constrained, like the banks, to use incomplete contracts, it is impossible to Pareto-improve on the laisser-faire allocation of resources as long as the markets for aggregate risk are complete. In this case, we say that the laisser-faire allocation is constrained efficient. Again, the invisible hand has worked its magic in the market and there is no need for government intervention. Unlike the equilibrium with complete contracts, an equilibrium with incomplete contracts may involve default. For example, if a bank uses a deposit contract, it may be optimal for the bank to make promises that it cannot keep in some states of nature. Then there can be a banking crisis, in the sense that banks default and are forced to liquidate assets. This is not a market failure because, as we have seen, the incidence of financial crises is constrained efficient.

The proceeding benchmarks show that financial crises are not always a bad thing; however, the assumption of complete markets is crucial here. Without this assumption, market failure is possible and some regulation to prevent or manage financial crises may be needed. The form of the optimal intervention is an open question. It may not resemble the regulation we see in practice.

Equilibrium Selection

Because of the presence of multiple equilibria, especially in models that assume banking panics are sunspot phenomena, equilibrium selection becomes a critical issue. Models without a unique equilibrium have limited predictive power and ambiguous comparative static properties.
Diamond and Dybvig (1983) appeal, among other things, to sunspots as the coordination device, but they do not model the sunspot equilibrium fully. They leave open the possibility of multiple equilibria after the initial decisions about investments and deposit contracts have been made by the bank. Postlewaite and Vives (1987) develop a model of a unique equilibrium that incorporates intrinsic uncertainty and does not rely on sunspots. Allen and Gale (1998, 2004) focus on fundamental shocks as the driver of financial crises. They assume that a non-crisis equilibrium is selected if such an equilibrium exists. In other words, crises are considered only if they are essential, that is, unavoidable.

Carlsson and van Damme (1993) introduce the notion of global games, which are games with asymmetric information about fundamentals. They show that the introduction of a small amount of asymmetric information can eliminate the multiplicity of equilibria in coordination games. Morris and Shin (1998, reprinted here as Reading 19) apply the global games approach in the context of currency crises to obtain a unique equilibrium. They link the panic-based and fundamental-based approaches by showing how the probability of a crisis depends on the fundamentals. Morris and Shin (2003) provide an excellent overview of global games. Allen and Morris (2001) develop a simple example to show how these ideas can be applied to banking crises. Rochet and Vives (2004, reprinted as Reading 12) use the unique equilibrium resulting from their global games approach to undertake policy analysis. They consider the role of ex ante regulation of solvency and liquidity ratios and ex post provision of liquidity by the central bank. Goldstein and Pausner (2005, reprinted as Reading 13) use the global games approach to
show how the probability of panic-based runs can be made endogenous and related to the parameters of the banking contract.

**Bubbles and Crises**

Banking crises often follow collapses in asset prices after what appears to have been a ‘bubble’. Asset price bubbles can arise for many reasons, but one important factor is the amount of liquidity provided in the form of money or credit by the central bank. Kindleberger (1978; p. 54) emphasizes the role of this factor in his history of bubbles: “Speculative manias gather speed through expansion of money and credit or perhaps, in some cases, get started because of an initial expansion of money and credit.”

In many recent cases where asset prices have first risen and then collapsed dramatically, an expansion of credit following financial liberalization appears to have been an important factor. Perhaps the best known example of this type of phenomenon is the dramatic rise in real estate and stock prices that occurred in Japan in the late 1980’s and the subsequent collapse in 1990. The next few years were marked by defaults and retrenchment in the financial system. The real economy was adversely affected in the aftermath of the bubble and growth rates during the 1990’s were barely positive or negative, in contrast to most of the post war period when the economy grew at much faster rates.

The Japanese and other examples suggest a relationship between asset-price bubbles and the provision of liquidity. These examples also illustrate the severe strains on the banking sector that can result when the bubble is burst. Banks that hold stocks and real estate or that have made loans to the owners of these assets, often come under severe
pressure from withdrawals because their liabilities are fixed and falling prices have reduced the value of their assets. Banks in this situation are forced to call in loans and liquidate assets, which in turn may exacerbate the problem of falling asset prices. These ‘negative bubbles’ can be very damaging to banks and other financial institutions. The damage to the financial sector will in turn make the problems of the real economy more severe than they need have been.

Despite the practical importance of the relationship between liquidity and asset price bubbles, there is no generally agreed upon theory of what underlies these relationships. Allen and Gale (2000a, reprinted here as Reading 14) provide a theory of bubbles and crises that relies on the existence of an agency problem. Many investors in real estate and stock markets obtain their investment funds from external sources. If the ultimate providers of funds are unable to observe the characteristics of the investment, there is a classic asset-substitution problem. Asset substitution increases the return to investment in risky assets and causes investors to bid up prices above their fundamental values. A crucial determinant of asset prices is thus the amount of credit provided by the financial system.

Financial liberalization, by expanding the volume of credit and creating uncertainty about the future path of credit expansion, can interact with the agency problem and lead to a bubble in asset prices. When the bubble bursts, either because returns are low or because the central bank tightens credit, banks are put under severe strain. Many of their liabilities are fixed while their assets fall in value. Depositors and other claimants may decide to withdraw their funds in anticipation of problems to come. This will force banks to liquidate some of their assets, resulting in a further fall in asset
prices in an illiquid market. It can be shown that the price of risky assets is determined by ‘cash-in-the-market’ pricing in some states and can fall below their fundamental value. This leads to an inefficient allocation of resources. The central bank can eliminate this inefficiency by an appropriate injection of liquidity into the market.

Money and Crises

Most models of banking crises ignore the role of money. Banks contract with depositors in real terms. As discussed above, Allen and Gale (1998) show how the use of nominal deposit contracts and the injection of money by a central bank can prevent crises or, at least, ensure the consequences are innocuous. Variations in the price level act as a substitute for state contingent contracts and allow risk to be shared optimally.

Smith (2002) considers a model in which spatial separation and limited communication introduce a role for money into a standard banking model with early and late consumers. He shows that the lower the inflation rate and the nominal interest rate, the lower is the probability of a banking crisis. Reducing the inflation rate to zero, in line with the Friedman rule, eliminates banking crises. However, this solution is inefficient as it causes banks to hold excess cash reserves at the expense of investment in higher yielding assets.

Diamond and Rajan (2001) develop a model in which banks have special skills to ensure that loans are repaid. By issuing real demand deposits, banks can precommit to recoup their loans. This allows long-term projects to be funded and depositors to consume when they have liquidity needs. However, this arrangement leads to the possibility of a liquidity shortage in which banks curtail credit when there is a real shock.
Diamond and Rajan (2006, reprinted as Reading 15) introduce money and nominal deposit contracts into the model to investigate whether monetary policy can help alleviate this problem. They assume there are two sources of value for money. The first arises from the fact that money can be used to pay taxes (the fiscal value). The second arises from the role of money in facilitating transactions (the transactions demand). They show that the use of money can improve risk sharing, since price adjustments introduce a form of state contingency in contracts. However, this is not the only possibility. In some cases, variations in the transaction value of money can lead to bank failures. Monetary intervention can help ease this problem. If the central bank buys bonds with money, it changes liquidity conditions in the market and allows banks to fund more long term projects than would be possible in the absence of intervention. The model thus provides a different perspective on the operation of monetary policy through bank lending.

We have only touched on some highlights of the literature on banking crises here. For further reading, excellent surveys are provided by Bhattacharya and Thakor (1993), Freixas and Rochet (1997), and Gorton and Winton (2003).

IV. Currency Crises

The first generation currency crisis models were designed to explain the problems experienced by a number of Latin American countries in the 1970’s and early 1980’s. The distinguishing feature of these episodes was that they had their origins in macroeconomic imbalances. The classic papers of Krugman (1979, reprinted as Reading 16) and Flood and Garber (1984, reprinted as Reading 17) show how a fixed exchange rate plus a government budget deficit leads to a currency crisis. In equilibrium, the
exchange rate cannot change discontinuously because this would lead to an arbitrage opportunity. Instead, the exchange rate adjusts continuously so that the real rate of return on domestic currency is equated to the real rate of return on foreign currency. The fiscal deficit is covered by a combination of depletion of foreign reserves and an inflation tax on the domestic money stock. When the exchange rate reaches its long-run equilibrium level, that is, the level that can be sustained without support, there is a speculative attack and reserves are exhausted.

Although the first generation models have many nice features, they have difficulty explaining episodes such as the Exchange Rate Mechanism (ERM) crisis of 1992, in which the pound and the lira dropped out of the mechanism. In the first place, the timing of currency crises is very unpredictable. Second, there are often “discontinuous” jumps in exchange rates. Finally, the models assume that no steps are taken by the government to eliminate deficits.

These problems led to the development of second generation models. For example, Obstfeld (1996) (Reading 18) shows how a conditional government policy can lead to multiple equilibria – one without a speculative attack and one with a speculative attack. The existence of multiple equilibria and uncertainty about the timing of an attack permit a discontinuous jump in the exchange rate. The outcome of the attack depends on the resources the government is willing to commit to maintain the exchange rate.

As in the banking crisis literature, equilibrium selection is an important issue. We have already mentioned the work of Carlsson and van Damme (1993) on global games. Their work shows that the existence of multiple equilibria depends on the players having common knowledge about the fundamentals of the game. Introducing noise ensures that
fundamentals are no long common knowledge and thus prevents the coordination that is essential for multiplicity. Morris and Shin (1998, reprinted as Reading 19) make the important step of showing how this global games approach can be applied to the analysis of currency crises. This approach leads to uniqueness of equilibrium which in turn allows comparative static policy analysis.

Despite Kaminsky and Reinhart’s (1999) research showing the intimate connection between banking crises and currency crises, the literatures on the two topics have, for the most part, developed independently. Important exceptions are the papers by Chang and Velasco (2000; 2001, reprinted as Reading 20). The first paper develops a model of currency and banking crises based on the Diamond-Dybvig model of bank runs. Chang and Velasco introduce money as an argument in the utility function. The central bank controls the ratio of currency to consumption. Different exchange rate regimes correspond to different rules for regulating the currency-consumption ratio. There is no intrinsic uncertainty in these models: banking and currency crises are “sunspot” phenomena. In other words, there are at least two equilibria, a ‘good’ equilibrium and a ‘bad’ equilibrium. In the good equilibrium, impatient consumers withdraw early and receive the proceeds from short-term assets and patient consumers wait to withdraw and receive the proceeds from long-term assets. In the ‘bad’ equilibrium, every one believes a crisis will occur and tries to withdraw early and long-term assets have to be liquidated at a loss.

Chang and Velasco (2001) show that the existence of the bad equilibrium depends on the exchange rate regime in force. In some regimes, only the ‘good’ equilibrium
exists; in other regimes, there exists a ‘bad’ equilibrium in addition to the ‘good’
equilibrium. The selection of equilibrium, ‘good’ or ‘bad,’ is not explained or modeled.

There is a large literature on currency crises. Flood and Marion (1999) provide a
good survey. Krugman (2000) contains a number of analyses of historic and recent
currency crises. Fourçans and Franck (2003) is an excellent book on the subject. Chui
and Gai (2005) explain the global games approach to analyzing crises.

V. Financial Contagion

Financial contagion refers to the process by which a shock in one part of the financial
system spreads to other parts through a series of ‘linkages.’ There are many kinds of
inter-linkages in the financial system. Interbank claims provide one channel of contagion.
If a bank fails, the financial institutions holding claims on the bank will be weakened.
Information provides another channel for contagion. A fall in prices on one market may
be interpreted as a negative signal about fundamentals. If these fundamentals are
common to other markets, the expected returns and hence prices on those markets will
also fall. Similarly, if one currency depreciates, other countries with common
fundamentals may find that their currency also depreciates.

Banks are also linked together via payments systems. These linkages can facilitate
Their model is based on a locational model of payment systems developed by
McAndrews and Roberds (1995). In a net payment system, banks extend credit to each
other within the day and settle their net position at the end of the day. This exposes banks
to the possibility of contagion if the failure of one institution triggers a chain reaction. In
a gross settlement system, transactions are settled on a one-to-one basis with central bank money. There is no risk of contagion, but banks have to hold large reserve balances. A net payment system is preferred when the probability of banks defaulting on payments is small; when the opportunity cost of holding central bank money reserves is high; and when the proportion of transactions between different locations is high. Freixas, Parigi and Rochet (2000) use the same model to examine the conditions under which ‘gridlock’ occurs. They show that there can be gridlock when the depositors in one bank withdraw their funds, anticipating that other banks cannot meet their netting obligations if all their depositors have also withdrawn their funds. In other work, Rochet and Tirole (1996a) consider the role of the too-big-to-fail policy in preventing contagion. Furfine (2003) examines interbank payment flows in the US and concludes that the risk of contagion from this source is small.

Allen and Gale (2000b, reprinted as Reading 21) focus on a different channel of contagion. In their model, banks in different regions or sectors of the banking system have negatively correlated liquidity shocks. It is optimal for banks to hold deposits in banks in other regions or sectors in order to provide liquidity if demand is unusually high. When one region suffers a banking crisis, the other regions suffer a loss because their claims on banks in the troubled region fall in value. If this spillover effect is strong enough, it can cause a crisis in adjacent regions. The crisis gets stronger as it passes from region to region and becomes a contagion. Eisenberg and Noe (2001) derive various results concerning the interconnectedness of institutions.

Lagunoff and Schreft (2001) study the spread of crises in a probabilistic model. Financial linkages are modeled using an overlapping partnership structure. Each project
requires two participants and each participant requires two projects, but partners are assigned to projects so that each individual has different partners for different projects. If one partner withdraws from a project, the project has to be abandoned; the other partner will want to withdraw from his second project, leaving the third partner with only one active project. This creates a domino effect that can spread throughout the system. In fact, the effect is instantaneous: when the probability that one’s partner will withdraw becomes too large, all participants simultaneously withdraw. This is interpreted as a financial crisis.

Rochet and Tirole (1996b) use monitoring as a means of triggering correlated crises: if one bank fails, it is assumed that other banks have not been properly monitored and a general collapse occurs. Dasgupta (2004) uses a global games approach to show how a unique equilibrium with contagion can arise when banks hold cross deposits. Allen and Carletti (2006) show how contagion can occur through the market for credit risk transfer while Allen and Carletti (2007) show that mark-to-market accounting can exacerbate contagion. Van Rijckeghem and Weder (2000) document linkages through banking centers empirically. Iyer and Peydró-Alcalde (2006) conduct a case study of interbank linkages resulting from a large bank failure due to fraud.

There is a growing literature on contagious currency crises and international contagion. Masson (1999) provides a good overview of the basic issues. He distinguishes between ‘monsoonal’ effects, spillovers and pure contagion. Monsoonal effects occur when there are major economic shifts in industrial countries that impact emerging economies. Spillovers occur when there are links between regions. Pure contagion occurs when there is a change in expectations unrelated to fundamentals. This
type of contagion is associated with multiple equilibria. Eichengreen, Rose and Wyplosz (1996) and Glick and Rose (1999) provide evidence that trade linkages are important factors in the spread of many currency crises. Kaminsky, Reinhart and Vegh (2003) consider the long history of cross-border contagion and consider why it occurs in some cases, but not in other similar situations.

Contagion in financial markets is studied by King and Wadwhani (1990) using a model in which information is correlated between markets. Price changes in one market are perceived to have implications for asset values in other markets. Calvo (2002) and Yuan (2005) use correlated liquidity shocks as a channel for contagion. When investors need cash to meet a margin call, they liquidate assets in a number of markets, spreading the shock throughout those markets. Kodres and Pritsker (2002, reprinted as Reading 22) use a multi-asset, rational expectations model to show how macroeconomic risk factors and country-specific asymmetric information can combine to produce contagion. Kyle and Xiong (2001) present a model of contagion in financial markets due to the existence of a wealth effect. Pavlova and Rigobon (2005) provide a theoretical model of contagion of stock market prices across countries arising from wealth transfers and portfolio constraints.

References for further reading can be found in the surveys by Masson (1999), De Bandt and Hartmann (2002), Karolyi (2003), and Pericoli and Sbracia (2003) and Upper (2006). Dungey and Tambakis (2005) contains a number of papers on various aspects of identifying the occurrence of international contagion.
Concluding Remarks

This brief introduction has sampled the rich and varied literature on financial crises. The articles we have selected for this volume provide important insights into central issues and the evidence and models used to analyze them. Despite the size and scope of the literature, much work remains to be done before these important phenomena are fully understood.
References


Table 1  
Duration and depth of crises  
(From Table 1 of Bordo et al. (2001))

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All countries</td>
<td>21 nations</td>
<td>56 nations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average duration of crises in years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currency crises</td>
<td>2.6</td>
<td>1.9</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Banking crises</td>
<td>2.3</td>
<td>2.4</td>
<td>3.1</td>
<td>2.6</td>
</tr>
<tr>
<td>Twin crises</td>
<td>2.2</td>
<td>2.7</td>
<td>1.0</td>
<td>3.7</td>
</tr>
<tr>
<td>All crises</td>
<td>2.4</td>
<td>2.4</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Average crises depth (cumulative GDP loss in %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currency crises</td>
<td>8.3</td>
<td>14.2</td>
<td>5.2</td>
<td>3.8</td>
</tr>
<tr>
<td>Banking crises</td>
<td>8.4</td>
<td>10.5</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td>Twin Crises</td>
<td>14.5</td>
<td>15.8</td>
<td>15.7</td>
<td>18.6</td>
</tr>
<tr>
<td>All crises</td>
<td>9.8</td>
<td>13.4</td>
<td>5.2</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Notes: a indicates no crises
Source: Authors' calculations
**Table 2**
National Banking Era Panics

<table>
<thead>
<tr>
<th>NBER Cycle Peak-Trough</th>
<th>Panic Date</th>
<th>%Δ(Currency/Deposit)*</th>
<th>%Δ Pig Iron†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct. 1873-Mar. 1879</td>
<td>Sep. 1873</td>
<td>14.53</td>
<td>-51.0</td>
</tr>
<tr>
<td>Mar. 1882-May 1885</td>
<td>Jun. 1884</td>
<td>8.80</td>
<td>-14.0</td>
</tr>
<tr>
<td>Mar. 1887-Apr. 1888</td>
<td>No Panic</td>
<td>3.00</td>
<td>-9.0</td>
</tr>
<tr>
<td></td>
<td>Nov.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jul. 1890-May 1891</td>
<td>1890</td>
<td>9.00</td>
<td>-34.0</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan. 1893-Jun. 1894</td>
<td>1893</td>
<td>16.00</td>
<td>-29.0</td>
</tr>
<tr>
<td>Jun. 1899-Dec. 1900</td>
<td>No Panic</td>
<td>2.78</td>
<td>-6.7</td>
</tr>
<tr>
<td>May 1907-Jun. 1908</td>
<td>Oct. 1907</td>
<td>11.45</td>
<td>-46.5</td>
</tr>
<tr>
<td></td>
<td>Aug.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan. 1913-Dec. 1914</td>
<td>1914</td>
<td>10.39</td>
<td>-47.1</td>
</tr>
</tbody>
</table>

*Percentage change of ratio at panic date to previous year's average
†Measured from peak to trough.
(Adapted from Table 1, Gorton (1988), p. 233)