Financial Crises: Theory and Evidence

Franklin Allen
University of Pennsylvania

Ana Babus
Cambridge University

Elena Carletti
European University Institute

June 8, 2009
1. Introduction

Financial crises have been pervasive phenomena throughout history. Bordo et al. (2001) find that their frequency in recent decades has been double that of the Bretton Woods Period (1945-1971) and the Gold Standard Era (1880-1993), comparable only to the Great Depression. Nevertheless, the financial crisis that started in the summer of 2007 came as a great surprise to most people. What initially was seen as difficulties in the US subprime mortgage market, rapidly escalated and spilled over to financial markets all over the world. The crisis has changed the financial landscape worldwide and its costs are yet to be evaluated.

The purpose of this paper is to concisely survey the literature on financial crises. Despite its severity and its ample effects, the current crisis is similar to past crises in many dimensions. In a recent series of papers, Reinhart and Rogoff (2008a, 2008b, 2009) document the effects of banking crises using an extensive data set of high and middle-to-low income countries. They find that systemic banking crises are typically preceded by credit booms and asset price bubbles. This is consistent with Herring and Wachter (2003) who show that many financial crises are the result of bubbles in real estate markets. In addition, Reinhart and Rogoff find that crises result, on average, in a 35% real drop in housing prices spread over a period of 6 years. Equity prices fall 55% over 3 ½ years. Output falls by 9% over two years, while unemployment rises 7% over a period of 4 years. Central government debt rises 86% compared to its pre-crisis level. While Reinhart and Rogoff stress that the major episodes are sufficiently far apart that policymakers and investors typically believe that “this time is different,” they warn that the global nature of this crisis will make it far more difficult for many countries to grow their way out.
A thorough overview of the events preceding and during the current financial crisis is provided in Adrian and Shin (2009), Brunnermeier (2009), Greenlaw et al. (2008), and Taylor (2008). Its seeds can be traced to the low interest rate policies adopted by the Federal Reserve and other central banks after the collapse of the technology stock bubble. In addition, the appetite of Asian central banks for (debt) securities contributed to lax credit. These factors helped fuel a dramatic increase in house prices in the U.S. and several other countries such as the U.K., Ireland and Spain. In 2006 this bubble reached its peak in the U.S. and house prices there and elsewhere started to fall. Mayer, Pence and Sherlund (2009) and Nadauld and Sherlund (2008) provide excellent accounts over the developments of the housing market preceding the crisis.

The fall in house prices led to a fall in the prices of securitized subprime mortgages, affecting financial markets worldwide. In August 2007 the interbank markets, particularly for terms longer than a few days, experienced considerable pressures and central banks were forced to inject massive liquidity. Conditions in collateralized markets have also changed significantly. Haircuts increased and low quality collateral became more difficult to borrow against. The Federal Reserve and other central banks introduced a wide range of measures to try to improve the functioning of the money markets. During the fall of 2007, the prices of subprime securitizations continued to fall and many financial institutions started to come under strain. In March of 2008 the Federal Reserve bailed out Bear Stern through an arranged merger with J. P. Morgan. Public funds and guarantees were required to induce J. P. Morgan to engage in the transaction.
Although the financial system and in particular banks came under tremendous pressure during this time, the real economy was not much affected. All that changed in September 2008 when Lehman’s demise forced markets to re-assess risk. While Lehman's bankruptcy induced substantial losses to several counterparties, its more disruptive consequence was the signal it sent to the international markets. Re-assessing risks previously overlooked, investors withdrew from the markets and liquidity dried up.

In the months that followed and the first quarter of 2009 economic activity in the U.S. and many other countries declined significantly. Unemployment rose dramatically as a result. There is a consensus to qualify the crisis as the worst since the Great Depression.

This survey aims to provide a relation between the existing knowledge on financial crises and the current events. We begin by reviewing the research on banking crises in Section 2. The role of liquidity in crises is treated in Section 3, while Section 4 considers the issue of contagion. Section 5 discusses the literature on bubbles and crises. Finally, Section 6 provides suggestions for future directions for research.¹

2. Banking crises and the economy

As financial intermediaries, banks channel funds from depositors and short term capital markets to those that have investment opportunities. By borrowing and lending from large groups, they can benefit from a diversified portfolio and offer risk sharing to depositors.

¹ We do not cover currency crises as this factor has not played an important role yet in the current crisis. Excellent surveys and analyses of currency crises are contained in Flood and Marion (1999), Krugman (2000) and Fourçans and Franck (2003).
Traditionally, intermediaries also act as delegated monitors, as in Diamond (1984), restructure loans to discipline borrowers, as in Gorton and Kahn (1994), or perform an important role in maturity transformation, as we will describe in the next section. Moreover, when the predominant source of external funding for firms is bank loans, banks become central to business activity as in Allen and Gale (2000a).

Financial distress in the banking system appears to be a concern for the economy as a whole. For instance, Dell’Ariccia, Detragiache and Rajan (2008) provide evidence that bank distress contributes to a decline in credit and to low GDP growth by showing that sectors more dependent on external finance perform relatively worse during banking crises. The effects are stronger in developing countries, in countries with less access to foreign finance, and where banking crises were more severe (see also Krozner, Laeven and Klingebiel (2007)).

Modern banking systems have increased in complexity over the last two decades. Despite running off-balance sheet vehicles or using various financial instruments to transfer credit risk, banks remained equally sensitive to panics and runs as they were at the beginning of the previous century. As Gorton (2008) points out, in the summer of 2007 holders of short-term liabilities refused to fund banks, expecting losses on subprime and subprime-related securities. As in the classic panics of the 19th and early 20th century, there were runs on banks. The difference is that modern runs typically involve the drying up of liquidity in the short term capital markets (a wholesale run) instead of or in addition to depositor withdrawals.

Academic research proposes two distinct theories to explain the origins of banking panics. One line of argument maintains that panics are undesirable events caused by random deposit withdrawals unrelated to changes in the real economy. In the influential work of Bryant
(1980) and Diamond and Dybvig (1983) bank runs are self-fulfilling prophecies. In these models, agents have uncertain needs for consumption in an environment in which long-term investments are costly to liquidate. If depositors believe that other depositors will withdraw then all agents find it rational to redeem their claims and a panic occurs. Another equilibrium exists where everybody believes no panic will occur and agents withdraw their funds according to their consumption needs. In this case, their demand can be met without costly liquidation of assets.

While it explains how panics may occur, the theory is silent on which of the two equilibria will be selected. Depositors’ beliefs are self-fulfilling and are coordinated by “sunspots.” Sunspots are convenient pedagogically but they do not have much predictive power. Since there is no real account of what triggers a crisis, it is difficult to use the theory for any policy analysis.

A selection mechanism that applies to this type of coordination games is introduced in Carlsson and van Damme (1993). The authors analyze incomplete information games where the actual payoff structure is randomly drawn from a given class of games and where each player makes a noisy observation of the game to be played. Such games are called global games. In a global games setting, the lack of common knowledge about the underlying payoff structure selects the risk dominant equilibrium to be the unique equilibrium of the game. Morris and Shin (1998) successfully applied this approach to coordination games in the context of currency crises, when there is uncertainty about economic fundamentals. Rochet and Vives (2004) and Goldstein and Pauzner (2005) have used global games to study to banking crises. An important recent contribution by Chen, Goldstein, and Jiang (2007) establishes the empirical applicability of the global games approach. The authors develop a global games model of mutual fund
withdrawals, where strategic complementarities among investors generate fragility in financial markets. Using a detailed data set, they find that consistent with their model funds with illiquid assets exhibit stronger sensitivity of outflows to bad past performance than funds with liquid assets.

He and Xiong (2009) depart from the static framework and analyze a dynamic model of bank runs. A coordination problem arises between creditors whose debt contracts with a firm mature at different times. In deciding whether to roll over his debt, each creditor faces the firm’s future rollover risk with other creditors. There is a unique equilibrium in which preemptive debt runs occur through a rat race among the creditors who coordinate their rollover decisions based on the publicly observable time-varying firm fundamental.

The second set of theories of banking crises is that they are a natural outgrowth of the business cycle. An economic downturn will reduce the value of bank assets, raising the possibility that banks are unable to meet their commitments. If depositors receive information about an impending downturn in the cycle, they will anticipate financial difficulties in the banking sector and try to withdraw their funds, as in Jacklin and Bhattacharya (1988). This attempt will precipitate the crisis. According to this interpretation, crises are not random events but a response of depositors to the arrival of sufficiently negative information on the unfolding economic circumstances. This view is consistent with the evidence in Gorton (1988) that in the U.S. in the late nineteenth and early twentieth centuries, a leading economic indicator based on the liabilities of failed businesses could accurately predict the occurrence of banking crises.

Building on the empirical work of Gorton (1988), Allen and Gale (1998) develop a model that is consistent with the business cycle view of the origins of banking crises. They assume that
depositors can observe a leading economic indicator that provides public information about future bank asset returns. If there are high returns then depositors are quite willing to keep their funds in the bank. However, if returns are sufficiently low, they will withdraw their money in anticipation of low returns and there is a crisis.

One strand of the business cycle explanation of crises emphasizes the role asymmetry of information plays in triggering banking crisis. In this view, a panic is a form of monitoring. Chari and Jagannathan (1988) focus on a signal extraction problem where some depositors withdraw money for consumption purposes while others withdraw money because they know that the bank is about to fail. In this environment, depositors who cannot distinguish whether there are long lines to withdraw at banks because of consumption needs or because informed depositors are getting out early may also withdraw. Chari and Jagannathan show crises occur not only when the outlook is poor but also when liquidity needs are high despite no one receiving information on future returns.

Calomiris and Kahn (1991) show that the threat of bank liquidation disciplines the banker when he can fraudulently divert resources ex post. The first come-first served constraint provides an incentive for costly information acquisition by depositors. Calomiris and Kahn regard bank runs as always beneficial since they prevent fraud and allow the salvage of some of the bank value. Diamond and Rajan (2001) develop a model in which banks have special skills to ensure that loans are repaid. By issuing demand deposits with a first come-first served feature, 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.
Another strand of the business cycle literature integrates banks into models of the economy. Allen and Gale (2004) 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 provide conditions for regulation to improve the allocation of resources. Allen and Gale explicitly model the interaction of banks and markets. 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.

He and Krishnamurthy (2008) are interested in the role of financial intermediaries in determining asset prices. They develop a dynamic general equilibrium framework where the need for intermediation arises endogenously based on optimal contracting considerations. The model has the feature that low intermediary capital reduces the risk-bearing capacity of the marginal investor, and replicates the observed rise during crises in Sharpe ratios, conditional volatility, correlation in price movements of assets held by the intermediary sector, and fall in riskless interest rates.

Since banks play a crucial role in the economy as whole, a key issue is the extent to which a regulator should intervene to prevent banking crises. For instance, Morrison and White (2005) analyze a general equilibrium model in which the regulator can learn the success probability of banks’ projects, and impose capital adequacy requirements. This way, a regulator with a strong reputation can alleviate moral hazard. Yet, crises of confidence will arise in economies where regulation solves only adverse selection problems. The appropriate policy response may be to tighten capital requirements to improve the quality of surviving banks.
Repullo and Suarez (2008) are concerned with the procyclicality of banking regulations. Banks hold capital buffers to smooth their capacity to lend over the business cycle. They show that the Basel II agreement changes the behavior of these buffers from countercyclical to procyclical. However, the higher buffers maintained in expansions are insufficient to prevent a significant contraction in the supply of credit at the arrival of a recession.

There is a large empirical literature on banking crises. Friedman and Schwartz (1963) have written a comprehensive monetary history of the U.S. from 1867-1960. Friedman and Schwartz argued that the crises were panic-based, as evidenced by the absence of downturns in the relevant macroeconomic time series prior to the crises. This contrasts with Gorton’s (1988) evidence that banking crises in the National Banking Era were predictable, which suggests banking crises are business cycle related. Calomiris and Gorton (1991) provide a wider range of evidence that crises are fundamental-based. Wicker (1980, 1996) shows that, despite the absence of collapses in US national macroeconomic time series, in the first two of the four crises identified by Friedman and Schwartz in the early 1930's there were large regional shocks and attributes the crises to these shocks. Calomiris and Mason (2003) undertake a detailed econometric study of the four crises using a broad range of data and conclude that the first three crises were fundamental-based while the fourth was panic-based.

We have only touched on some highlights of the literature on banking crises here. More complete surveys are provided by Bhattacharya and Thakor (1993), Gorton and Winton (2003), Allen and Gale (2007, Chapter 3), Freixas and Rochet (2008), Rochet (2008), and Degryse, Ongena and Kim (2009).
3. Liquidity and interbank markets

Interbank markets play a key role in financial systems. Their main purpose is to redistribute liquidity in the financial system from the banks that have cash in excess to the ones that have a shortage. In this process, they become the medium for implementing central banks’ monetary policy. Their smooth functioning is essential for maintaining financial stability.

Bhattacharya and Gale (1987) is the pioneering theoretical study in this area. They analyze a setting in which individual banks face privately observed liquidity shocks due to a random proportion of depositors wishing to make early withdrawals. In addition, each bank has private information about the liquid fraction of its portfolio. Since the liquidity shocks are imperfectly correlated across intermediaries, banks coinsure each other through an interbank market. Bhattacharya and Gale show that, even in the absence of an aggregate liquidity shock for the intermediary sector as a whole, banks are induced to under-invest in liquid assets and free-ride on the common pool of liquidity because of the lower return that liquid assets yield. A central bank can mitigate this problem by (even imperfectly) monitoring banks’ asset choices. However, they argue one would not expect to achieve the first-best, as in such an asymmetric information setting it seems unrealistic to assume that a central bank can elicit perfect knowledge of the quality of the assets in overall bank portfolios.

Freixas and Holthausen (2004) analyze the scope for international interbank market integration when cross-border information about banks is less precise than home country information. The timing of consumption needs generates liquidity shocks for the banks, both at the individual and at the aggregate level. Banks can cope with these shocks by investing in a storage technology or can use the interbank market for channelling liquidity. They look at
secured repo and unsecured interbank lending markets since both allow banks to cope with liquidity shocks, and they consider under what conditions segmented or integrated international interbank markets exist. They show that a segmented interbank market is always an equilibrium, while the emergence of an integrated international market depends on the quality of cross-border information. Only if cross-border information is sufficiently precise, is integration of markets possible.

Turbulence in the interbank market in the current financial crisis has spurred a series of new papers. Allen, Carletti and Gale (2009) show that the interbank market is characterized by excessive price volatility when there is a lack of opportunities for banks to hedge aggregate and idiosyncratic liquidity shocks. By using open market operations to fix the short-term interest rate, a central bank can prevent price volatility and implement the constrained efficient solution. Thus, the central bank effectively completes the market, a result in line with the argument of Goodfriend and King (1988) that open market operations are sufficient to address pure liquidity risk on the interbank markets. One implication of the model is that situations where banks stop trading with each other can be a feature of the constrained efficient solution.

Acharya, Gromb and Yorulmazer (2008), and Freixas and Jorge (2008) also study inefficiencies in the interbank market. Acharya et al. (2008) consider that interbank markets are characterized by moral hazard, asymmetric information, and monopoly power in times of crisis. They show that a bank with surplus liquidity has bargaining power vis-à-vis deficit banks which need liquidity to keep funding projects. Surplus banks may strategically provide insufficient lending in the interbank market in order to induce inefficient sales of bank-specific assets by the needy banks, which results in an inefficient allocation of resources. The role of the central bank
is to provide an outside option to the deficit bank for acquiring the needed liquidity. Freixas and Jorge (2008) examine how financial imperfections in the interbank market affect the monetary policy transmission mechanism. In their model, firms face liquidity shocks and rely on bank credit to raise external finance. Firms’ shocks will result in a demand for credit and a liquidity shock for the banks that can be smoothed out through an interbank market. Asymmetry of information the interbank market disrupts the efficient allocation of liquidity to solvent illiquid banks. Consequently, by tightening monetary policy banks with less liquidity are forced to cut down on their lending.

Motivated by the current financial crisis, several papers seek to explain market freezes. Diamond and Rajan (2009) relate the seizing up of term credit with the overhang of illiquid securities. When banks have a significant quantity of assets with a limited set of potential buyers, shocks in future liquidity demands may trigger sales at fire sale prices. The prospect of a future fire sale of the bank’s assets depresses their current value. In these conditions, banks prefer holding on to the illiquid assets and risking a fire sale and insolvency than selling the asset and ensuring its own stability in the future, since the states in which the depressed asset value recovers are precisely the states in which the bank survives. In turn, this creates high expected returns to holding cash or liquid securities across the financial system and an aversion to locking up money in term loans.

Acharya, Gale and Yorulmazer (2009) show that freezes in markets for rollover debt, such as asset-backed commercial paper, depend on how information about the quality of the asset is revealed. When there is a constant probability that “bad news” is revealed each period and, in the absence of bad news, the value of the assets is high. By contrast, when there is a
constant probability that “good news” is revealed each period and, in the absence of good news, the value of the assets is low. In the latter scenario, the debt capacity of the assets is below the fundamental value and is decreasing in the liquidation cost and frequency of rollovers. In the limit, as the number of rollovers becomes unbounded, the debt capacity goes to zero even for an arbitrarily small default risk.

Another explanation for market freezes relies on asymmetric information. Heider, Hoerova and Holthausen (2009) analyze the functioning of interbank markets when there is asymmetric information. As banks face individual liquidity shocks, there is a role for an interbank market in which banks with surplus liquidity can lend to those with liquidity shortage. An interbank loan may not be repaid, however, because the long-term investment is risky, thus, giving rise to counterparty risk. Asymmetric information about counterparty risk can elevate the interbank market spreads and in extreme situations lead to a total interbank market break down. Bolton, Santos and Scheinkman (2008) have also provided a theory of liquidity provision with asymmetric information. In their model there is an adverse selection problem due to the superior information that intermediaries have about the assets they hold. When they sell they must do so at a discount and this becomes greater the longer they hold an asset. If an intermediary is hit by a liquidity shock the problem it faces is whether to sell its assets now at a discount or to try and ride out the crisis. The danger of doing this is that it runs the risk of having to sell at a greater discount if the crisis lasts longer than expected. In the immediate trading equilibrium, intermediaries sell assets immediately to ensure they have enough liquidity. In the delayed trading equilibrium intermediaries try to ride out the crisis and only sell if they are forced to. For some parameter values only the immediate trading equilibrium exists while for others both do and in this case the delayed trading equilibrium is Pareto superior.
A different mechanism for market crashes is proposed by Huang and Wang (2008a, b). Instead of relying on the presence of information asymmetry among investors about the fundamentals, they show that purely idiosyncratic and non-fundamental shocks can cause market crashes if capital flow is costly. Agents trade to smooth out idiosyncratic shocks to their wealth. Since there is no aggregate uncertainty, their trades will be perfectly synchronized and matched, and there will be no need for liquidity if market presence is costless. In this case, the market-clearing price always reflects the fundamental value of the asset, and idiosyncratic shocks generate trading but have no impact on prices. In contrast, when market presence is costly, the need for liquidity arises endogenously and idiosyncratic shocks can affect prices via two channels: first trading becomes infrequent which makes traders more risk averse, and second the gains from trading for potential sellers are always larger than the gains from trading for potential buyers. The asymmetry in their appetite to trade leads to order imbalances in the form of excess supply, and the price has to decrease in response.

Two studies isolate illiquidity risk from other confounding effects. Morris and Shin (2009) define “illiquidity risk” the probability of a default due to a run when the institution would otherwise have been solvent. They show that is difference between “asset insolvency risk”, as the conditional probability of default due to deterioration of asset quality if there is no run by short term creditors, and “total credit risk,” as the unconditional probability of default, either because of a (short term) creditor run or (long run) asset insolvency. Brunnermeier and Pedersen (2008) distinguish between market liquidity and funding liquidity. Market liquidity reflects how difficult is to raise money by selling the asset, instead of by borrowing against it. Traders provide market liquidity, and their ability to do so depends on their availability of funding. Conversely, traders’ funding, i.e., their capital and margin requirements, depends on the
assets’ market liquidity. They show that, under certain conditions, margins are destabilizing and market liquidity and funding liquidity are mutually reinforcing, leading to liquidity spirals.

Most models of banking crises ignore the role of money. Ultimately, banks contract with depositors in nominal terms. 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 (2006) introduce money and nominal deposit contracts into the model in Diamond and Rajan (2001) 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.
Interest rates also play a role in determining the liquidity of interbank markets. Freixas, Martin and Skeie (2009) suggest that inducing low interbank market rates in states of financial disruptions is an optimal policy response of the central bank. Although generally banks can provide better risk-sharing possibilities and more liquidity than (incomplete) markets, during financial disruption, banks themselves face considerable uncertainty regarding their own idiosyncratic liquidity needs. An interbank market can achieve the optimal allocation, which implies efficient risk sharing to consumers and insuring banks against idiosyncratic liquidity shocks. In the optimum, however, the interest rate on this market must be state-contingent and low in states of financial disruption. This suggests a role for a central bank which can implement the efficient allocation by setting the interest rates in the interbank market. Diamond and Rajan (2008) consider a model where banks fund long term illiquid projects by borrowing short term from households. But when household needs for funds are high, banks will have to call in loans to long gestation projects in order to generate the resources to pay them. Interest rates will rise sharply to equate the household demand for consumption goods and the supply of these goods from terminated projects. Debtors will have to shut down illiquid projects, and the net worth of the bank decreases, leading in the limit to runs. The optimal policy response may require authorities to commit to raising rates when low, to offset incentives for banks to make more illiquid loans.

4. Contagion

The prevalence of financial crises has led many to conclude that the financial sector is unusually susceptible to shocks. A shock that initially affects only a particular region or sector or perhaps even a few institutions can become systemic and then infect the larger economy.
The literature on contagion takes two approaches: examining direct linkages and indirect balance-sheet linkages. In looking for contagious effects via direct linkages, early research by Allen and Gale (2000b) studied how the banking system responds to contagion when banks are connected under different network structures. Banks perfectly insure against liquidity shocks by exchanging interbank deposits. The connections created by swapping deposits, however, expose the system to contagion. The authors show that incomplete networks are more prone to contagion than complete structures. Better connected networks are more resilient since the proportion of the losses in one bank's portfolio is transferred to more banks through interbank agreements. To show this, they take the case of an incomplete network where the failure of a bank may trigger the failure of the entire banking system. They prove that, for the same set of parameters, if banks are connected in a complete structure, then the system is more resilient with regard to contagious effects.

The research that followed, although using stylized models, captured well the network externalities created from individual bank risk. Freixas, Parigi and Rochet (2000) considered the case of banks that face liquidity shocks due to uncertainty about where consumers will withdraw funds. In their model, the connections between banks are realized through interbank credit lines that enable these institutions to hedge regional liquidity shocks. As in Allen and Gale’s study, more interbank connections enhance the resilience of the system to the insolvency of a particular bank. One drawback is that this weakens the incentives to close inefficient banks. Moreover, the authors find that the stability of the banking system depends crucially on whether many depositors choose to consume at the location of a bank that functions as a money center or not.
Concerned with the optimal financial network, Leitner (2005) constructs a model where the success of an agent's investment in a project depends on the investments of other agents she is linked to. Since endowments are randomly distributed across agents, an agent may not have enough cash to make the necessary investment. In this case, agents may be willing to bail out other agents to prevent the collapse of the whole network. Leitner examines the design of optimal financial networks that minimize the trade-off between risk sharing and the potential for collapse. In a related paper, Kahn and Santos (2008) investigate whether banks choose the optimal degree of mutual insurance against liquidity shocks. They show that when there is a shortage of exogenously supplied liquidity, which can be supplemented by bank liquidity creation, the banks generally fail to find the correct degree of interdependence. In aggregate, they become too risky. Dasgupta (2004) also explores how linkages between banks, represented by crossholding of deposits, can be a source of contagious breakdowns. The study examines how depositors who receive a private signal about banks' fundamentals may wish to withdraw their deposits if they believe that enough other depositors will do the same. To eliminate the multiplicity of equilibria the author uses the concept of global games. Dasgupta isolated a unique equilibrium, depending on the value of the fundamentals. In the same spirit, Brusco and Castiglionesi (2007) show that there is a positive probability of bankruptcy and propagation of a crises across regions when banks keep interbank deposits and may engage in excessive risk taking if they are not enough capitalized.

Parallel to this literature, other researchers applied network techniques developed in mathematics and theoretical physics to study contagion. For instance, Eisenberg and Noe (2001) investigate default by firms that are part of a single clearing mechanism. First, the authors show the existence of a clearing payment vector that defines the level of connections between firms.
Next, they develop an algorithm that allows them to evaluate the effects that small shocks have on the system. This algorithm produces a natural measure of systemic risk based on how many waves of defaults are required to induce a given firm in the system to fail. Similarly, Minguez-Afonso and Shin (2007) use lattice-theoretic methods to study liquidity and systemic risk in high-value payment systems, such as for the settlement of accounts receivable and payable among industrial firms, and interbank payment systems. Gai and Kapadia (2007) develop a model of contagion in financial networks and use similar techniques as the epidemiological literature on spread of disease in networks to assess the fragility of the financial system. As with Allen and Gale, they find that greater connectivity reduces the likelihood of widespread default. However, shocks may have a significantly larger impact on the financial system when they occur.

The tools of the network literature have been applied also to analyze the issue of network formation. Babus (2007) proposes a model where banks form links with each other as an insurance mechanism to reduce the risk of contagion. At the base of the link formation process lies the same intuition developed in Allen and Gale (2000b): better connected networks are more resilient to contagion. The model predicts a connectivity threshold above which contagion does not occur, and banks form links to reach this threshold. However, an implicit cost associated with being involved in a link prevents banks from forming more connections than required by the connectivity threshold. Banks manage to form networks where contagion rarely occurs.

Castiglionesi and Navarro (2007) are also interested in decentralizing the network of banks that is optimal from a social planner perspective. In a setting where banks invest on behalf of depositors and there are positive network externalities on the investment returns, fragility arises
when banks that are not sufficiently capitalized gamble with depositors' money. When the probability of bankruptcy is low, the decentralized solution approximates the first best.

Besides the theoretical investigations, there has been a substantial interest in looking for evidence of contagious failures of financial institutions resulting from the mutual claims they have on one another. Most of these papers use balance sheet information to estimate bilateral credit relationships for different banking systems. Subsequently, the stability of the interbank market is tested by simulating the breakdown of a single bank. Upper and Worms (2004) analyze the German banking system, Cocco, Gomes, and Martins (2005) consider Portugal, Furfine (2003) the US, Boss, Elsinger, Thurner, and Summer (2004) Austria, and Degryse and Nguyen (2007) Belgium. These papers find that the banking systems demonstrate high resilience, even to large shocks. For instance, simulations of the worst case scenarios for German system show the failure of a single bank could lead to the breakdown of up to 15% of the banking sector based on assets. Since these results depend heavily on how the linkages between banks are estimated and they abstract from any type of behavioral feedback (Upper 2006), it is likely that they provide downward bias estimator of contagious outcomes. Mistrulli (2007) confirms this when analyzing contagion propagates within the Italian interbank market using actual bilateral exposures. Iyer and Peydró-Alcalde (2006) find second order behavioral feedback in a study of interbank linkages at the time of the failure of a large Indian bank: banks with higher interbank exposure to the failed bank experience higher deposit withdrawals.

The second approach to modeling contagion focuses on indirect balance-sheet linkages. Lagunoff and Schreft (2001) construct a model where agents are linked in the sense that the return on an agent's portfolio depends on the portfolio allocations of other agents. In their model,
agents that are subject to shocks reallocate their portfolios, thus breaking some linkages. Two related types of financial crisis can occur in response. One occurs gradually as losses spread, breaking more links. The other type occurs instantaneously when forward-looking agents preemptively shift to safer portfolios to avoid future losses from contagion. Similarly, de Vries (2005) shows that there is dependency between banks' portfolios, given the fat tail property of the underlying assets, and this carries the potential for systemic breakdown. Cifuentes, Ferrucci, and Shin (2005) present a model where financial institutions are connected via portfolio holdings. The network is complete as everyone holds the same asset. Although the authors incorporate direct linkages through mutual credit exposures, contagion is mainly driven by changes in asset prices.

Several papers document empirically how indirect connections between financial institutions pose problems for systemic risk. Adrian and Brunnermeier (2009) actually propose a new measure for systemic risk that is conditional on an institution (or the whole financial sector) being under distress. Their concern is confirmed by Boyson, Stahel and Stulz (2008) who find that the average probability that a hedge fund style index has extreme poor performance increases with the number of other hedge funds with extreme poor performance. Similarly, Jorion and Zhang (2009) find evidence of credit contagion via counterparty effects.

Recent contributions have linked the risk of contagion to financial innovation. Parlour and Winton (2008) analyze the choice a bank has to lay off credit risk between credit default swaps (CDS) and loan sales. With a CDS, the originating bank retains the loan's control rights but no longer has an incentive to monitor; with loan sales, control rights pass to the buyer of the loan, who can then monitor, though in a less-informed manner. The authors show that when
capital costs are low, loan sales are dominant, while when capital costs are high, CDS and loan sales may co-exist. Shin (2009) studies the impact of securitization on financial stability. Since securitization allows credit expansion through higher leverage of the entire financial system, it may drive down lending standards enhancing fragility.

Allen and Carletti (2006) show how financial innovation in the form of credit risk transfer can create contagion across sectors and lower welfare relative to the autarky solution. In their model asset prices are determined by the available of liquidity or, by “cash in the market”. The structure of liquidity shocks hitting the banking sector determines the mechanism for contagion. When banks face a uniform demand for liquidity, they keep a sufficient amount of the short term asset and do not need to raise additional liquidity in the market. In this case, credit risk transfer improves risk sharing across sectors. However, when banks face idiosyncratic liquidity shocks, there is scope to invest in the long risk-free asset that can be traded in the market. Transferring credit risk is now detrimental as it induces a higher need of liquidity in the market and a greater variability in the asset prices. This in turn affects banks' ability to face their liquidity shocks as it implies a severe reduction in the price of the long asset that banks use to hedge their liquidity risk. The effect of introducing credit risk transfer depends crucially also on the accounting system in use, be it historical cost or mark-to-market accounting, as shown by Allen and Carletti (2008). When banks need to liquidate a long-term asset on an illiquid market, it may not be desirable to value such assets according to market values as it reflects the price volatility needed to induce liquidity provision.

The current crisis made it clear we need a broader view on financial systems to capture externalities between institutions. The usual justification for intervention by central banks and
governments to prevent the bankruptcy of systemic financial institutions is that this will prevent contagion. This was the argument used by the Federal Reserve for intervening to ensure Bear Sterns did not go bankrupt in March 2008, for example (see Bernanke (2008a)). The bankruptcy of Lehman Brothers a few months later in September of 2008 illustrated how damaging contagion can be. The process did not work in quite the way envisaged in the academic literature and was not accounted for in the decision of the Federal Reserve and Treasure that Lehman should not be saved. The first spillover was to the money market mutual fund sector. Reserve Capital “broke the buck” as it held a significant amount of paper issued by Lehman. This led to many withdrawals from other money market mutual funds and four days after Lehman announced bankruptcy, the government was forced to announce guarantees for the entire sector. After seeing Lehman Brothers collapse, confidence in the creditworthiness of banks and other financial institutions and firms fell significantly and this is when the financial crisis started to spill over into the real economy and had such a damaging effect.

5. Bubbles and crises

Banking crises often follow collapses in asset prices after what appears to have been a ‘bubble’. This is in contrast to standard neoclassical theory and the efficient markets hypothesis which precludes the existence of bubbles. The global crisis that started in 2007 provides a stark example.

Asset price bubbles can arise for many reasons, but one important factor is the amount of liquidity provided by the central bank as money or credit. 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.” With the current crisis, the monetary policies of central banks particularly the U.S. Federal Reserve appear to have been too loose and have focused too much on consumer price inflation, ignoring asset price inflation. Moreover, the Asian crisis of 1997 and the policies of the IMF adopted then induced Asian governments to hoard funds. This created important global imbalances that expanded the credit available and helped to fuel the bubble. The bubble burst triggered problems with subprime mortgages, which in turn posed difficulties for the banking system and then spread to the real economy. When the prices of securitized subprime mortgages fell too low (see Bank of England (2008) for evidence prices were below fundamentals), a negative bubble was created.

The sequence of events in the current crisis is, in fact, often observed. Kaminsky and Reinhart (1999) study a wide range of crises in 20 countries including 5 industrial and 15 emerging ones. A common precursor to most of the crises considered is financial liberalization and significant credit expansion. These are followed by an average rise in the price of stocks of about 40 percent per year above that occurring in normal times. The prices of real estate and other assets also increase significantly. At some point the bubble bursts and the stock and real estate markets collapse. In many cases banks and other intermediaries were overexposed to the equity and real estate markets and about a year later on average a banking crisis ensues. This is often accompanied by an exchange rate crisis as governments choose between lowering interest rates to ease the banking crisis or raising interest rates to defend the currency. Finally, a significant fall in output occurs and the recession lasts for an average of about a year and a half.

There are a number of theories that can explain how bubbles can arise (see, e.g., Tirole (1982, 1985), Allen and Gorton (1993), Allen, Morris and Postlewaite (1993), Abreu and
Brunnermeier (2003), Scheinkman and Xiong (2003), Brunnermeier and Nagel (2004), Hong, Scheinkman and Xiong (2008), and Brunnermeier (2001) for an overview). Here we focus on theories that are explicitly related to crises. Allen and Gale (2000c) develop a model of boom and bust 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, there is a financial crisis.

Makarov and Plantin (2009) analyze changes in house prices in an economy where banks grant mortgages to liquidity-constrained households to finance a fixed supply of homes. Households’ aggregate debt capacity drives the aggregate demand for homes. Home supply at a given date stems from foreclosures in case of default, sales motivated by the acquisition of a larger home, and sales that follow exogenous moving decisions. Market-clearing home prices in turn drive aggregate debt capacity. The model generates interesting insights into the impact of lower refinancing costs and housing bubbles on equilibrium outcomes.

There has been a substantial literature attempting to understand how sales of assets can lead prices to fall too low so that there is a negative bubble. Bernanke and Gertler (1989) and
Bernanke, Gertler and Gilchrist (1996) develop the notion of the financial accelerator. They show that credit market conditions can amplify and propagate shocks. Asymmetric information leads to an agency problem between borrowers and lenders. They show that a negative shock to the borrowers’ wealth is amplified because of the nature of the principal-agent relationship between lenders and borrowers. Another very influential paper is Kiyotaki and Moore (1997) who show that small shocks can lead to large effects because of the role of collateral. A shock that lowers asset prices lowers the value of collateral. This means that less borrowing is possible, asset prices are further lowered and so on in a downward spiral. Disruptions in liquidity provision can be the shock that initially lowers asset prices and starts the problem.

While Kiyotaki and Moore (1997) take the ratio of the amount that can be borrowed against collateral as given, Geanakoplos (1997, 2003, 2009) and Fostel and Geanakoplos (2008) show how the interest rate and amount of collateral are simultaneously determined. In practice in crises, the amounts that can be borrowed against different kinds of collateral vary significantly and this is an important part of why shocks are amplified by the financial sector.

Caballero and Krishnamurthy (2001) distinguish between collateral that can be used when borrowing domestically and collateral that can be used when borrowing internationally. In emerging economies the latter is particularly important. They show that the two types of collateral can interact in important ways. If domestic collateral is liquidated at fire sale prices, this can lead to wasted international collateral. When both international and domestic collateral are limited, buffers to deal with adverse shocks will be limited and the effects of such shocks will be severe.
Another seminal contribution is Holmstrom and Tirole (1998). In their model entrepreneurs operate firms. These entrepreneurs need to provide costly effort for the firm to be successful. In order to ensure they are willing to do this, they need to be provided with part of the equity of the firm. This limits the ability of the firm to raise funds by issuing securities to outside investors. If a firm is hit by a liquidity shock and needs more funds to continue, it may be unable to raise them in the market. If it cannot continue because of this, then it may go bankrupt and this can cause a significant loss in welfare. The occurrence of this event is more likely when credit markets are disrupted. In order to overcome this problem, the firm may need to hold liquid securities that it can sell in the event of a liquidity shock. If the private supply of such securities is insufficient, the government may be able to improve welfare by issuing government debt that can be held by firms. Now when firms are hit by a shock they will have sufficient liquidity to continue.

6. Future directions for research

The financial crisis has generated substantial amounts of new research. Many questions remain to be addressed, however. Perhaps most importantly, we need to gain a better understanding of the market failures that lead to financial crises. Clearly, such insights are necessary to design policies aimed both to prevent such crises and to ameliorate their effects once they occur.

Market failures in financial services are numerous. Three of most important in our view deserve thorough investigation: (i) provision of liquidity; (ii) limits to arbitrage and the mispricing of assets; and (iii) contagion.
One of the striking features of the current crisis has been the illiquidity of the interbank and credit markets. Central banks have responded with a bevy of measures to restore fluidity in these markets. We have discussed above the emerging research that investigates market freezes. However, we also need to understand the optimal policy responses to these failures. Current policies have been introduced on a pragmatic basis and may have unintended consequences. For example, allowing financial institutions to swap asset backed securities for Treasuries at end-of-quarter and end-of-year reporting periods allows “window dressing”. Investors in these financial institutions are then unable to judge the risks that are being taken. Such an abrogation of the market mechanism is unlikely desirable. We need a full analysis of the best way to restore liquidity in financial markets.

As discussed above, it is difficult to reconcile market prices of the toxic mortgage assets with valuations based on fundamentals. In this case, mark-to-market accounting becomes a concern and the question of how to measure regulatory capital remains open. The nature of this particular market failure and the best way to correct it are yet to be explored. The resolution proposed by U.S. Treasury and the Federal Reserve implied having the government purchase assets and act as the arbitrageur of last resort. This was the much maligned TARP plan of the fall of 2008. In early 2009 a related plan was proposed to help fund the private sector to arbitrage these markets. Once again there is a serious issue of whether these are the best policy responses to a market failure that is not yet well understood.

Perhaps the most important of the three market failures is contagion. This is a very common justification for central bank intervention. It was clearly the main drive that prompted the Federal Reserve to save Bear Sterns in March 2008. Bernanke (2008b) argued that failing to
rescue the bank would have risked a whole wave of bank failures and a meltdown in the financial system. The developments after Lehman demise in September 2008 suggest that contagion is indeed a serious problem. However, contagion did not manifest itself as a wave of failures suggesting a more complex phenomenon that is currently not well understood. A full understanding of contagion is necessary before adequate policy responses can be designed.

From the macro perspective, we need to understand better the relationship between monetary policy, credit and asset prices. Financial crises often follow the bursting of bubbles. The bubble associated to the crisis that started in 2007 seems to have been caused by two main factors. First, the interest rates were set by the Federal Reserve in 2003 at a very low level. When the interest rate is set below the rate of increase in house prices, it is quite likely that a bubble in property prices can emerge. However, much more research is needed to understand the precise way in which such a bubble plays out. Second, global imbalances triggered an abundance of credit. This problem has its roots in the Asian Crisis of 1997. The fact that the IMF forced many Asian countries such as South Korea with fundamentally strong economies to raise interest rates and cut government expenditure had a long-lasting effect. The measure was the precise opposite of what the U.S. and Europe did when faced with a similar situation. Asian governments and central banks concluded that self insurance by accumulating reserves was a better way to solve their problem than relying on the IMF. The design of a global financial infrastructure that all countries can rely on to share risk without self insuring is an important issue for research.

The financial services industry is perhaps the most regulated in the world. However, regulations seem to have done little to prevent the crisis. The failure of their main purpose
indicates that a complete overhaul is needed. The basic problem with current financial
regulation is that it is not based on a coherent intellectual framework. A good illustration of this
shortcoming is the Basel agreements. While aimed to develop a framework for regulation of
banks’ accounting capital, the accords fail to explain what the underlying problems are and why
the particular regulations imposed are the best way to deal with these problems. Without
identifying if the market failures stem from coordination problems, incomplete risk sharing
opportunities or other problems, it is difficult to assess whether regulations based on accounting
capital are appropriate. Moreover, the particular ratios specified in the regulation seem to be
rather ad-hoc. Understanding these issues is a major issue for research going forward.

The reoccurrence of financial crises over the centuries suggests that it is unlikely they
will be completely prevented in the future. When they do occur what should governments and
central banks do? The response of many governments in the current crisis of injecting funds
through the purchase of preferred shares is one type of intervention. The justification for
intervening is that large financial institutions are “Too big to fail”. If they were allowed to fail
this would lead to very damaging contagion. However, “Too big to fail” does not imply “Too
big to liquidate”. For example, the government could temporarily nationalize large financial
institutions on the brink of failure to prevent contagion, and later liquidate them in an orderly
fashion. Current policies of supporting failed institutions create bad incentives for large
institutions if they start relying on being saved in future crises. We need to fully analyze and
understand the best form of intervention.
References


Dell’Ariccia, G., E. Detragiache, and R. Rajan (2008)”The Real Effects of Banking Crises”, *Journal of Financial Intermediation* 17, 89-112


