THE OXFORD HANDBOOK OF

BANKING

Edited by

ALLEN N. BERGER

PHILIP MOLYNEUX

and

JOHN O. S. WILSON

OXFORD
UNIVERSITY PRESS

CHAPTER 2

THE ROLES OF BANKS IN FINANCIAL SYSTEMS

FRANKLIN ALLEN
ELENA CARLETTI[1]

Introduction

Understanding the many roles that banks play in the financial system is one of the fundamental issues in theoretical economics and finance. The crisis that started in the summer of 2007 underlines just how important banks are to the economy. The efficiency of the process through which savings are channeled into productive activities is crucial for growth and general welfare. Banks are one part of this process. Figure 2.1 gives an overview of the functioning of a financial system. Lenders of funds are primarily households and firms. These lenders can supply funds to the ultimate borrowers, who are mainly firms, governments, and households, in two ways. The first is through financial markets, which consist of money markets, bond markets, and equity markets. The second is through banks and

[1] We are grateful to the editors for helpful comments.

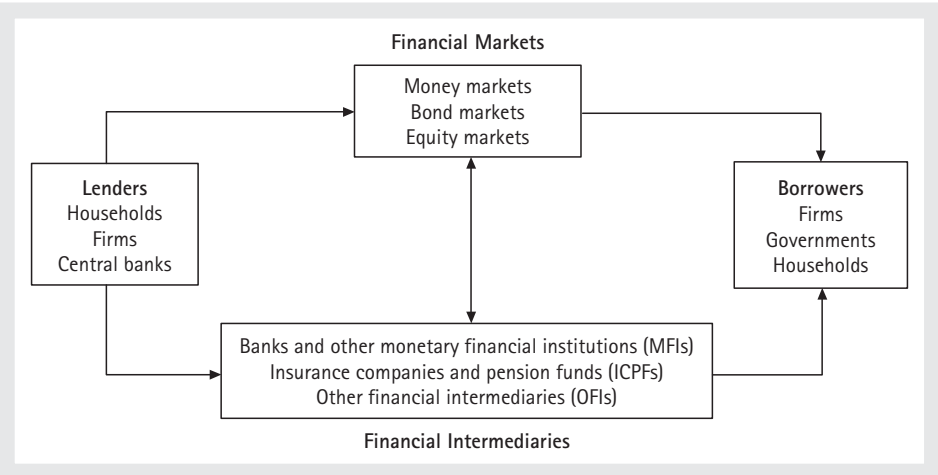

Fig. 2.1. An overview of the financial system

Source: Allen, Chui, and Maddaloni, (2004) p. 491.

other financial intermediaries such as money market funds, mutual funds, insurance companies, and pension funds.

Despite the trend of globalization in recent years, the importance of banks in different economies varies significantly. Figure 2.2 shows a comparison of the long-term financing structure of the Euro area, the UK, the US, Japan, and non-Japan Asia[2] in 1995 and 2003. The figures are given as a percentage of GDP. Bank loans consist of domestic credit to the private sector. The figures in the stock market column are the total market capitalization. The bond market figures are divided into public- and private sector bonds.

It can be seen from Figure 2.2a that in 1995 the euro area had small stock markets but large bank loans and in that sense could be considered as bank-based. However, it also had a significant bond market in terms of both public- and private sector debt. The UK was significantly different, with a large stock market and bank loans but a small bond market, particularly in terms of private sector debt.[3] In some sense it seems to be both market-based and bank-based. The main features of the US financial structure are a small amount of bank loans, a significant stock market, and a much larger bond market than any of the other areas in relative terms. It is the most market-based economy. Japan has significant amounts of finance in all categories. It is very much a bank- and market-based economy. Non-Japan Asia is more similar to the UK: bank loans and the stock market are important but the bond market is not.

[2] This includes Hong Kong, Indonesia, Korea, Malaysia, the Philippines, Singapore, Taiwan, and Thailand.

[3] The UK used to have a significant corporate bond market but this died during the 1970s when inflation was high. It has not revived in recent years despite the reduction in inflation.

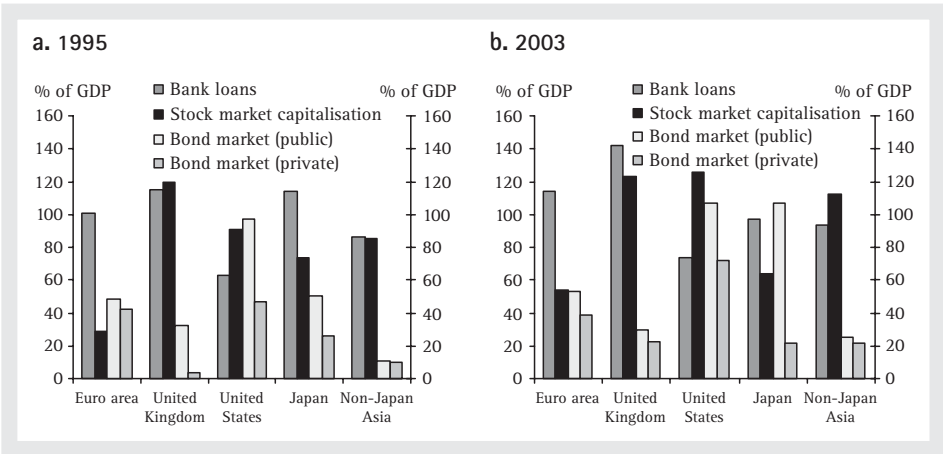

Fig. 2.2. Size of the financial markets by country/region

Source: Allen, Chui, and Maddaloni (2004) p. 492. *Original sources*: CEIC Data Ltd, International Financial Statistics, and national sources.

Figure 2.2b shows the situation in 2003, several years after the Asian crises. It can be seen that the structure is basically the same. The main difference is that Japanese government debt has increased significantly. One interesting feature is that the financial structure in non-Japan Asia has not changed significantly despite the Asian crises.

Figure 2.2 focuses on the claims that are issued by borrowers. Another way of considering the importance of banks is to look at household assets. These are shown in Figure 2.3a. This shows that all the economies are distinctly different. Households in the euro area own significantly fewer financial assets than in the other economies, with a total of 192 percent of GDP compared with 306 percent, 327 percent, and 267 percent for the UK, the US, and Japan, respectively. In terms of the composition of assets there are also large differences. In the euro area, assets held in banks are the most important, insurance and pension funds are next, with direct holdings of shares after that. One striking thing is that household portfolios in the UK are very similar to those in the euro area, with one significant difference: the investment in insurance and pension funds is dramatically higher. This is presumably a result of the difference in public sector pension schemes. In the UK, the basic pension from the state is minimal, while in the euro area, state pensions are usually generous. The US is an outlier in terms of the direct holdings of shares and other equity. Also, households have relatively little in banks. Meanwhile, Japan is an outlier in terms of the amount of assets held in banks where households hold much more in this form than households in other countries. In fact, the Japanese post office bank is the largest deposit taker in the world. Japanese households also have significant amounts in insurance and pension funds. This is to a large extent in insurance companies that offer debt-like contracts. Given the

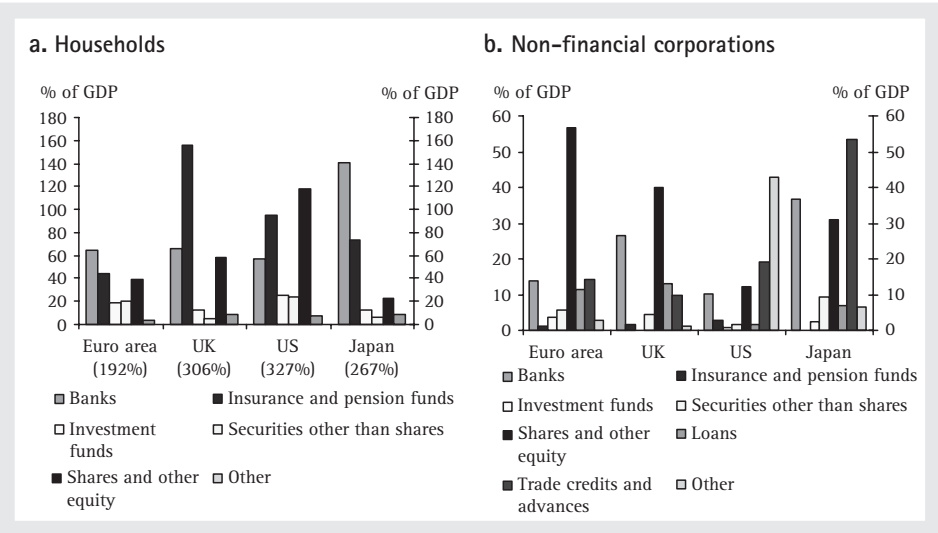

Fig. 2.3. Portfolio allocation (average 1995–2002)

Source: Allen, Chui, and Maddaloni (2004) p. 493. *Original sources*: ECB, Federal Reserve Board and Bank of Japan.

small holdings of shares and other equity, the Japanese bear significantly less financial risk than the households in the US and UK. The US has somewhat less intermediation than the other economies, although the total amount of intermediation is significant in all economies.

Figure 2.3b shows the assets of non-financial corporations. These again underline significant differences across the economies. The euro area and the UK are quite similar except for the amount of shares and other equity held and the amount of trade credits. These are both larger in the euro area than in the UK. The US has much less investment than the other countries except for the 'other' category. This includes holdings of other assets, which are not identified explicitly in the flow of funds data.[4] Japan is perhaps the most different. It has significantly more assets in banks and more trade credit than other countries.

The implication of Figures 2.2 and 2.3 is that the importance of banks and their roles are significantly different in different economies. We start by considering the basic rationales for the existence of banks. The following sections consider the monitoring role of banks and their risk sharing role. The bearing of risks by banks can have important implications for financial stability. The next sections consider banking crises and the contagion between banks that can occur in a crisis. Then we

[4] The column representing 'other' assets is 'unidentified miscellaneous' assets. It is a residual item, arising after accounting for all asset or liability items reported by classified flow of funds' sectors. (In other words, accounting items that do not represent claims on another party are all classified as 'other'.) One example would be the accounting value of goodwill after M&A activities.

consider the role of banks in spurring growth, the corporate governance role of banks, and relationship banking.

Delegated monitoring and banks

An argument that is often put forward in favor of bank-based systems is that banks allow various informational problems to be solved. One important problem is if borrowers must take some action to make proper use of the funds they have borrowed. This action could be the level of effort or choice of project from among various different risky alternatives. The borrower can always claim that a low outcome is due to bad luck rather than from not taking the correct action. Lenders cannot observe the borrower's action unless they pay a fixed cost to monitor the borrower. In a financial market with many lenders, there is a free-rider problem. Each lender is small, so it is not worth paying the fixed cost. Everybody would like to free-ride, leaving it to someone else to bear the monitoring cost. As a result, no monitoring will be done.

A possible solution is to hire a single monitor to check what the borrower is doing. The problem then becomes one of monitoring the monitor, to make sure that she actually monitors the borrowers. Diamond (1984) develops a model of delegated monitoring to solve this problem. Intermediaries have a diversified portfolio of projects for which they provide finance. They pre-commit to monitor borrowers by promising lenders a fixed return. If the intermediary does not monitor, then it will be unable to pay the promised return to lenders. Diamond's model thus illustrates how banks have an incentive to act as a delegated monitor and produce the information necessary for an efficient allocation of resources.

Boot and Thakor (1997) develop a model of financial system architecture that builds on this view of banks as delegated monitors. They assume there are three types of information problem. The first is that there is incomplete information about the future projects a firm has available to it. Outside investors can gather information about these possibilities. The second problem is that lenders cannot observe whether borrowers invest the funds in a risky or safe project. The third problem is the likelihood that borrowers will have the opportunity to invest in a risky project. Boot and Thakor are able to show that the first problem can best be solved by a financial market and the second and third problems can best be solved by intermediaries. They argue that banks will predominate in an emerging financial system, while the informational advantages of markets may allow them to develop in a mature financial system.

The risk sharing role of banks

One of the most important functions of the financial system is to share risk and it is often argued that financial markets are well suited to achieve this aim. As shown in Figure 2.3 and discussed in the Introduction, if both direct holdings of equities and indirect holdings in insurance companies and mutual funds are taken account of, a large amount of household assets is held in equity and only a small amount in banks in the US and UK. In both countries households are exposed to substantial amounts of risk through their holdings of equities. At the other extreme, households in Japan are shielded from risk because they ultimately hold a majority of their assets in banks and very little in equities. Although not as safe as in Japan, households' asset holdings in the euro area are much safer than in the US and UK.

Although the proportions of risky assets held by households in the US and UK are much higher than in Japan and the euro area, this does not necessarily mean that the absolute amount of risk borne by households is greater because the amount invested in financial assets could be higher in the latter countries. However, it can be seen from Figure 2.2 that the euro area has a significantly lower amount of financial assets relative to GDP. Thus, taking into account the amount of wealth held in financial assets increases the differences in the amount of risk borne by households in the different countries, rather than reducing it. Not only do households hold much higher proportions in risky securities in the US and UK, they also hold more financial assets.

How can one explain these differences in the amount of risk households are apparently exposed to in different financial systems? Standard financial theory suggests that the main purpose of financial markets is to improve risk sharing. Financial markets in the US and UK are more developed by most measures than in Japan and the euro area. How can it be that households are exposed to much more risk in the US and UK than in Japan and the euro area?

Allen and Gale (1997; and 2000a: chap. 6) have provided a resolution to this paradox. They point out that traditional financial theory has little to say about hedging non-diversifiable risks. It assumes that the set of assets is given and focuses on the efficient sharing of these risks through exchange. For example, the standard diversification argument requires individuals to exchange assets so that each investor holds a relatively small amount of any one risk. Risks will also be traded so that more risk-averse people bear less risk than people who are more risk-tolerant. This kind of risk sharing is termed cross-sectional risk sharing, because it is achieved through exchanges of risk among individuals at a given point in time. However, importantly, these strategies do not eliminate macroeconomic shocks that affect all assets in a similar way.

Departing from the traditional approach, Allen and Gale focus on the intertemporal smoothing of risks that cannot be diversified at a given point in time.

They argue that such risks can be averaged over time in a way that reduces their impact on individual welfare through intertemporal smoothing by banks. This involves banks building up reserves when the returns on the banks' assets are high and running them down when they are low. The banks can thus pay a relatively constant amount each period and do not impose very much risk on depositors. The authors show that the incentives for engaging in intertemporal smoothing are very different in market-based financial systems. Incomplete financial markets, on the one hand, may not allow effective intertemporal smoothing. The problem is that the long-lived asset 'crowds out' the storage technology because it can be bought and sold for the same price and in addition it pays a dividend. Long-lived banks, on the other hand, can achieve intertemporal smoothing as explained above. However, for this result to hold it is necessary that the banks are not subject to substantial competition from financial markets. In fact, competition from financial markets can lead to disintermediation and the unraveling of intertemporal smoothing provided by long-lived institutions.

Banking crises

Banks perform an important role in terms of maturity transformation. They collect demandable deposits and raise funds in the short-term capital markets and invest them in long-term assets. This maturity mismatch allows them to offer risk sharing to depositors but also exposes them to the possibility that all depositors withdraw their money early. Runs can involve the withdrawal of funds by depositors (retail runs) or the drying up of liquidity in the short-term capital markets (wholesale runs). In the case of the run on Northern Rock in the UK in late 2007, both occurred. These runs can originate in two ways. They can either occur spontaneously as a panic resulting from 'mob psychology' or 'mass hysteria' (e.g., Kindleberger, 1978) or they may arise from fundamental causes that are part of the business cycle (see, e.g., Mitchell, 1941).

The 'panics' view suggests that crises are random events, unrelated to changes in the real economy. The seminal papers developed by Bryant (1980) and Diamond and Dybvig (1983) show bank runs are self-fulfilling prophecies. Given the assumption of first-come, first-served and costly liquidation of some assets, there are multiple equilibria. If everybody believes no panic will occur only those with genuine liquidity needs will withdraw their funds and these demands can be met without costly liquidation of assets. However, if everybody believes a crisis will occur then it becomes a self-fulfilling prophecy as people rush to avoid being last in line. Which of these two equilibria occurs depends on extraneous variables or

'sunspots'. Although sunspots have no effect on the real data of the economy, they affect depositors' beliefs in a way that turns out to be self-fulfilling.

The key issue in theories of panics is which equilibrium is selected and in particular what is the equilibrium selection mechanism. Sunspots are convenient pedagogically but this explanation does not have much content. It does not explain why the sunspot should be used as a coordination device. There is no real account of what triggers a crisis. This is particularly a problem if there is a desire to use the theory for policy analysis.

Carlsson and van Damme (1993) showed how the introduction of a small amount of asymmetric information could eliminate the multiplicity of equilibria in coordination games. They called the games with asymmetric information about fundamentals 'global games'. Their work showed that the existence of multiple equilibria depends on the players having common knowledge about the fundamentals of the game. Introducing noise ensures that the fundamentals are no longer common knowledge and thus prevents the coordination that is essential to multiplicity. Morris and Shin (1998) applied this approach to models of currency crises. Rochet and Vives (2004) and Goldstein and Pauzner (2005) have applied the same technique to banking crises.

Using a global games approach to ensure the uniqueness of equilibrium is theoretically appealing. However, what is really needed in addition to logical consistency is empirical evidence that such an approach is valid. In an important recent contribution, Chen, Goldstein, and Jiang (2007) develop a global games model of mutual fund withdrawals. Using a detailed data set they find evidence consistent with their model. This represents significant evidence supporting the global games approach.

An alternative to the sunspot view is that banking crises 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 US 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.

An extensive number of authors have developed models of banking crises caused by aggregate risk. For example, Chari and Jagannathan (1988) focus on a signal extraction problem where part of the population observes a signal about future returns. Others must then try to deduce from observed withdrawals whether an unfavorable signal was received by this group or whether liquidity needs happen to

be high. Chari and Jagannathan are able to show that crises occur not only when the outlook is poor but also when liquidity needs turn out to be high.

Building on the empirical work of Gorton (1988) that nineteenth-century banking crises were predicted by leading economic indicators, 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 the returns are sufficiently low they will withdraw their money in anticipation of low returns. There is thus a crisis.

Allen and Gale (2004b) 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. This would be possible if securities contingent on all the possible combinations of aggregate liquidity and asset return shocks, or in other words all the states of nature, were available. Similarly, the risk sharing contracts between intermediaries and consumers are said to be complete if the payoffs can be explicitly conditioned on all the possible combinations of aggregate liquidity and asset return shocks. 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 of liquidity demand and asset returns. Allen and Gale (2004b) show that the laissez-faire allocation of resources is efficient provided markets are complete. This is the case even if contracts are incomplete. However, crises are inefficient if markets are incomplete. In this case financial fragility and contagion can occur.

The crisis that started in 2007 provides a dramatic example of how damaging banking crises can be. The causes for its occurrence are not fully understood yet, but many attribute them to the bad incentives in the origination of mortgages and their securitization, the provision of ratings for securitizations, and the risk management systems of investment firms. The large global impact of the crisis suggests, however, that the problems with subprime mortgages are a symptom rather than

the cause. One main problem is that there was a bubble, first in stock prices and then in property prices, and the economic system is now suffering the fallout from the collapse of that bubble. The monetary policies of central banks, particularly the US Federal Reserve, appear to have been too loose and have focused far too much on consumer-price inflation while ignoring asset price inflation. Moreover, the Asian crisis of 1997 and the policies of the IMF during that crisis led to a desire among Asian governments to hoard funds. This created important global imbalances that expanded the credit available and helped to fuel the bubble. Allen and Gale (2000c) show how such an expansion of credit can create a bubble.

Whatever are the reasons behind the crisis, its effects have now certainly spread to the real economy. Most industrialized and non-industrialized countries are experiencing problems with many of their industries entering into recession. The problems are multiple. On the one hand, the difficulties of the financial sectors induce intermediaries to tighten their credit standards thus making it more difficult for firms to obtain credit and at good rates. On the other hand, the sharp fall in consumer demand decreases sales and future orders. As in the financial sectors, the problems are not confined to single firms but affect whole industries. The car industry is one dramatic example, but other manufacturing industries, construction, and many more are very much under pressure.

Banks and contagion

The prevalence of financial crises has led many to conclude that the financial sector is unusually susceptible to shocks. One theory is that small shocks can have a large impact. A shock that initially affects only a particular region or sector or perhaps even a few institutions can spread by contagion through interlinkages between banks and financial institutions to the rest of the financial sector and then infect the larger economy.

The theoretical literature on contagion takes two approaches. On the one hand, there is a number of papers that look for contagious effects via direct linkages. Allen and Gale (2000) study how the banking system responds to contagion when banks are connected under different network structures. In a setting where consumers have the Diamond and Dybvig (1983) type of liquidity preferences, banks perfectly insure against liquidity shocks by exchanging interbank deposits. The connections created by swapping deposits 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 to contagion since

the proportion of the losses in one bank's portfolio is transferred to more banks through interbank agreements.

Other models capture well the network externalities created from an individual bank risk. Freixas, Parigi, and Rochet (2000) consider the case of banks that face liquidity needs as consumers are uncertain about where they are to consume. In their model, the connections between banks are realized through interbank credit lines that enable these institutions to hedge regional liquidity shocks. In the same way as in Allen and Gale (2000), interbank connections enhance the resilience of the system to the insolvency of a particular bank. The 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.

Dasgupta (2004) uses a global games approach to show how a unique equilibrium with contagion can arise when banks hold cross deposits. In the same spirit, Brusco and Castiglionesi (2007) show that there is a positive probability of bankruptcy and propagation of a crisis across regions when banks keep interbank deposits and may engage in excessive risk taking if they are insufficiently capitalized.

Recent contributions have linked the risk of contagion to financial innovation and the accounting system in use. The common feature in this analysis is the presence of incomplete markets where liquidity provision is achieved by selling assets in the market when required. Asset prices are determined by the available liquidity or, said differently, by the 'cash in the market'. It is necessary that people hold liquidity and stand ready to buy assets when they are sold. These suppliers of liquidity are no longer compensated for their opportunity cost of providing liquidity state by state. The cost must be made up on average across all states. This implies volatility in the asset prices that can in turn lead to costly and inefficient crises. In order for people to be willing to supply liquidity they must be able to make a profit in some states. In equilibrium, prices of assets will be such that the profit in the states where banks and intermediaries sell assets is sufficient to compensate the providers of liquidity for all the other states where they are not called upon to provide liquidity and simply bear the opportunity cost of holding it. In other words, asset prices are low in the states where banks and intermediaries need liquidity. But, from an efficiency point of view, this is exactly the wrong time for there to be a transfer from the banks and intermediaries who need liquidity to the providers of liquidity. This is because the banks' depositors who need liquidity will already have low income because they have to withdraw early.

Allen and Carletti (2006) rely on cash in the market pricing to show how financial innovation in the form of credit risk transfer can create contagion across sectors and lower welfare relative to the autarky solution. They focus on the

structure of liquidity shocks hitting the banking sector as the main mechanism determining 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 is beneficial as it improves risk sharing across sectors. Differently, when banks face idiosyncratic liquidity shocks, they invest also in the long risk-free asset and trade it in the market. The presence of credit risk transfer turns out now to be detrimental as it induces a higher need of liquidity in the market and consequently 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 which banks use to hedge their liquidity risk. The banks that are selling the long asset receive a lower amount and may be unable to pay their depositors.

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). The intuition is similar to the one in the previous chapter. 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 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 who 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 in their model direct linkages through mutual credit exposures as well, contagion is mainly driven by changes in asset prices.

Complementary to the literature on network effects, Babus (2007) considers a model where banks form links with each other in order to reduce the risk of contagion. The network is formed endogenously and serves as an insurance mechanism. At the base of the link-formation process lies the same intuition developed in Allen and Gale (2000): 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 to being involved in a link prevents banks from forming connections more than required by the connectivity threshold. Banks manage to form

networks where contagion rarely occurs. Castiglionesi and Navarro (2007) are also interested in whether banks manage to decentralize the network structure a social planner finds optimal. 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. For example, Upper and Worms (2004) analyze the German banking system. They show that the failure of a single bank could lead to the breakdown of up to 15 percent of the banking sector in terms of assets. Cocco, Gomes, and Martins (2005) consider Portugal; Furfine (2003) the US; Boss, et al. (2004) Austria; and Degryse and Nguyen (2007) Belgium. Iyer and Peydró-Alcalde (2006) conduct a case study of interbank linkages resulting from a large bank failure due to fraud. Upper (2006) contains a survey of this literature. The main conclusion of the literature is that contagion is usually not a serious risk provided there are not significant price movements in response to the turmoil. If there are, as in Cifuentes, Ferrucci, and Shin (2005) then contagion effects can be significant.

The current crisis illustrates the practical importance of contagion. 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, 2008). The bankruptcy of Lehman Brothers a few months later in September 2008, illustrated quite how damaging contagion can be. The process did not work in quite the way envisaged in the academic literature and occurred despite the judgment 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 on it. Going forward, much more research is needed to understand the many channels of contagion in a crisis.

Banks and growth

Another important role of banks is in spurring growth. There has been a debate on the relative effectiveness of banks compared with financial markets in doing this. This debate was originally conducted in the context of German and UK growth in the late nineteenth and early twentieth centuries. Gerschenkron (1962) argued that the bank-based system in Germany allowed a closer relationship between bankers providing the finance and industrial firms than was possible in the market-based system in the UK. Goldsmith (1969) pointed out that although manufacturing industry grew much faster in Germany than the UK in the late nineteenth and early twentieth centuries the overall growth rates were fairly similar. More recently, Levine (2002) uses a broad database covering forty-eight countries over the period 1980–95. He finds that the distinction between bank-based and market-based systems is not an interesting one for explaining the finance-growth nexus. Rather, elements of a country's legal environment and the quality of its financial services are most important for fostering general economic growth. In contrast, in a study of thirty-six countries from 1980 to 1995, Tadesse (2002) does find a difference between bank-based and market-based financial systems. For underdeveloped financial sectors, bank-based systems outperform market-based systems, while for developed financial sectors, market-based systems outperform bank-based systems. Levine and Zervos (1998) show that higher stock market liquidity or greater bank development lead to higher growth, irrespective of the development of the other. There is some evidence that financial markets and banks are complements rather than substitutes. Demirgüç-Kunt and Maksimovic (1998) show that more-developed stock markets tend to be associated with increased use of bank finance in developing countries.

There is a large theoretical literature on the relative merits of bank-based and market-based systems for innovation and growth. Bhattacharya and Chiesa (1995) consider a model of R&D incentives and financing. In a market system, lenders learn the value of each firm's R&D at the interim stage after R&D has been undertaken but before production takes place. The lenders can share the information among the firms and will do so if it is in their interest. Bhattacharya and Chiesa show that their incentives to do this correspond to maximizing the aggregate value of the firms' R&D projects. Also, a collusive agreement can be structured so that only one firm actually produces at the production stage. However, this collusion creates a free-rider problem and reduces incentives to undertake the R&D at the first stage. If this incentive problem is severe enough, bilateral financing may be preferable. Under this arrangement, each firm is financed by one bank and there is no scope for information sharing. As a result, each firm's R&D information remains proprietary.

Allen and Gale (1999; and 2000a: chap. 13) ask whether financial markets or banks are better at providing finance for projects where there is diversity of opinion as in the development of new technologies. Diversity of opinion arises from differences in

prior beliefs, rather than differences in information. The advantage of financial markets is that they allow people with similar views to join together to finance projects. This will be optimal provided the costs necessary for each investor to form an opinion before investment decisions are made are sufficiently low. Finance can be provided by the market even when there is great diversity of opinion among investors. Intermediated finance involves delegating the financing decision to a manager who expends the cost necessary to form an opinion. There is an agency problem in that the manager may not have the same prior as the investor. This type of delegation turns out to be optimal when the costs of forming an opinion are high and there is likely to be considerable agreement in any case. The analysis suggests that market-based systems will lead to more innovation than bank-based systems.

The corporate governance role of banks

The importance of equity ownership by financial institutions in Japan and Germany, and the lack of a strong market for corporate control in these countries have led to the suggestion that the agency problem in these countries is solved by banks acting as outside monitors for large corporations. In Japan, this system of monitoring is known as the main bank system. The characteristics of this system are the long-term relationship between a bank and its client firm, the holding of both debt and equity by the bank, and the active intervention of the bank should its client become financially distressed. It has been widely argued that this main bank relationship ensures that the bank acts as delegated monitor and helps to overcome the agency problem between managers and the firm. However, the empirical evidence on the effectiveness of the main bank system is mixed (see, e.g., Hoshi, Kashyap, and Scharfstein, 1990; Hoshi, Kashyap, and Scharfstein, 1993; Aoki and Patrick, 1994; and Hayashi, 2000). Overall, the main bank system appears important in times of financial distress, but less important when a firm is doing well.

In Germany, the counterpart of the main bank system is the hausbank system. Banks tend to have very close ties with industry and form long-run relationships with firms not only because of the loans they make and the shares they directly own but also because of the proxies they are able to exercise. A number of studies have provided evidence on the effectiveness of the outside monitoring of German banks (see, e.g., Gorton and Schmid, 2000).

In an important book, Edwards and Fischer (1994) have argued that in Germany the corporate governance role of banks has been overemphasized in the literature. They provide a variety of evidence that banks do not have the degree of influence as

lenders, shareholders, or voters of proxies that is usually supposed. For example, they find that the number of votes controlled in a company is only weakly related to the number of representatives the bank has on the supervisory board. Hellwig (1991; 1994) also provides a number of theoretical arguments concerning the disadvantages of the banking system in Germany.

Relationship banking

There is a growing literature that analyzes the advantages and disadvantages of relationships in banking (see, for reviews, Boot, 2000; Gorton and Winton, 2003; and Degryse and Ongena, 2008). If, on the one hand, close and durable relationships provide better access to firms and ameliorate some of the information problems characterizing lending relationships, on the other hand, they also involve inefficiencies related to the hold-up and the soft-budget-constraint problems. The hold-up problem refers to the possibility that a relationship bank uses the superior private information it possesses about the firm to extract rents, thus distorting entrepreneurial incentives and causing inefficient investment choices (Sharpe, 1990; Rajan, 1992; and von Thadden, 1995). The soft-budget-constraint problem concerns the inability of a relationship lender to commit itself to a particular course of action in advance. Although it is optimal to threaten to terminate the availability of credit in advance, once the borrower has defaulted the first loan becomes a 'sunk cost'. If the firm has another good project we should expect that the lender will continue to extend credit, even if the borrower defaults. Renegotiation thus creates a time-consistency problem. The threat to terminate credit creates good incentives for the borrower to avoid the risk of default. Termination of credit is not Pareto-efficient *ex post*, but the incentive effect makes both parties better off. However, if the borrower anticipates that the lender will not carry out the threat in practice, the incentive effect disappears. Although the lender's behavior is now *ex post* optimal, both parties may be worse off *ex ante*.

Multiple bank relationships can help mitigating the drawbacks of single-bank relationships in terms of the hold-up and the soft-budget-constraint problems. As for the former, borrowing from multiple banks can restore competition among banks and, consequently, improve entrepreneurial incentives (Padilla and Pagano, 1997). As for the latter, Dewatripont and Maskin (1995) argue that by complicating the refinancing process and making it less profitable multiple bank lending allows banks to commit not to extend further inefficient credit. Similarly, Bolton and Scharfstein (1996) show that multiple bank lending reduces entrepreneurial incentives to default strategically because it complicates debt renegotiation.

The number of bank relationships also has important implications for banks' role as monitors. In a context where both firms and banks are subject to moral hazard problems, Carletti (2004) analyzes how the number of bank relationships influences banks' monitoring incentives, the level of loan rates, and a firm's choice between single and multiple bank relationships. Multiple bank lending suffers from duplication of effort and free-riding but it benefits from diseconomies of scale in monitoring, thus involving a lower level of monitoring but not necessarily higher loan rates than single lending. Since banks choose their monitoring effort to maximize their expected profits, they may choose a level of monitoring which is excessive from the firms' perspective. When this is the case, the firm may choose multiple bank relationships in order to reduce the overall level of monitoring. The attractiveness of such a choice increases with the cost of monitoring, the firm's private benefit, and expected profitability. In a similar framework, Carletti, Cerasi, and Daltung (2007) analyzes the circumstances where banks with limited diversification opportunities find it profitable to enter into multiple bank relationships. They show that sharing lending allows banks to diversify better their portfolios but still entails duplication of effort and free-riding. When the benefit of greater diversification dominates, multiple bank lending leads to higher overall monitoring as a way to mitigate the agency problem between banks and depositors and achieve higher banks' expected profits. The attractiveness of multiple bank lending now decreases with the level of banks' (inside) equity and firms' prior profitability, while it increases with the cost of monitoring.

Other rationales for multiple bank relationships relate to firms' desire to reduce liquidity risk and disclose information through credit relationships. Detragiache, Garella, and Guiso (2000) show that, when relationship banks face internal liquidity problems, borrowing from multiple banks can avoid early liquidation of profitable projects. Yosha (1995) suggests that firms may prefer multiple bank lending as a way to disclose confidential information about the quality of their projects and to avoid aggressive behavior by competitors.

As a final remark, note that there are ways other than multiple bank relationships to solve the problem of lack of commitment affecting exclusive bank relationships. For example, financial institutions may develop a valuable reputation for maintaining commitments. In any one case, it is worth incurring the small cost of a sub-optimal action in order to maintain the value of the reputation. Incomplete information about the borrower's type may lead to a similar outcome. If default causes the institution to believe it is more likely that the defaulter is a bad type, then it may be optimal to refuse to deal with a firm after it has defaulted. Institutional strategies such as delegating decisions to agents who are given no discretion to renegotiate may also be an effective commitment device. Several authors (Huberman and Kahn, 1988; Hart and Moore, 1988; Gale, 1991; and Allen and Gale, 2000a: chap. 10) have argued that, under certain circumstances, renegotiation is welfare-improving. In that case, the argument is reversed. Intermediaries

that establish long-term relationships with clients may have an advantage over financial markets precisely because it is easier for them to renegotiate contracts.

Concluding remarks

We have covered a number of roles of banks in the financial system in this chapter. Banks act as delegated monitors and ensure that firms use the resources allocated to them effectively. They also play an important role in sharing risk in the economy by diversifying and smoothing fluctuations over time. These are positive aspects of the roles banks play. However, the fixed nature of the claims they issue can cause fragility in the financial system. Banks are often at the center of financial crises as in the crisis that started in the summer of 2007. They can help spread crises if there is contagion and small shocks can have a large effect on the financial system and the economy. Banks play an important role in providing funds for firms and helping them and the economy to grow. They are also important for corporate governance, particularly in countries like Germany where bankers sit on boards and control a significant number of proxy votes. Finally, banks can help overcome asymmetric information problems by forming long-lived relationships with firms.

There a number of other roles that we have not covered as they are the subjects of other chapters of the book. These include the role of banks in underwriting securities, covered in Chapter 7 and the role of banks in payments systems, covered in Chapter 28. There remain other roles that are important that are less well understood. Many of these involve the interaction of banks with financial markets of various kinds. The recent crisis has illustrated that securitization can lead to significant problems because bank incentives are fundamentally different when loans are sold rather than retained. The role that banks play in derivative markets is also not fully understood. If there is a chain of counterparties how can that risk be fully assessed if the chain is opaque as it usually is? Finally, how can banks be prevented from taking risks if they retain the profits when there are good outcomes but are bailed out by the government in times of crisis? These are all important issues for future research.

References

Allen, F. and E. Carletti (2006). 'Credit Risk Transfer and Contagion', *Journal of Monetary Economics*, 53: 89–111.

——— ——— (2008). 'Mark-to-Market Accounting and Liquidity Pricing', *Journal of Accounting and Economics*, 45: 358–78.

——M. Chui, and A. Maddaloni (2004). 'Financial Systems in Europe, the USA, and Asia', *Oxford Review of Economic Policy,* 20: 490–508.
——and D. Gale (1997). 'Financial Markets, Intermediaries, and Intertemporal Smoothing', *Journal of Political Economy,* 105: 523–46.
————(1998). 'Optimal Financial Crises', *Journal of Finance,* 53: 1245–84.
————(1999). 'Diversity of Opinion and the Financing of New Technologies', *Journal of Financial Intermediation,* 8: 68–89.
————(2000a). *Comparing Financial Systems*, Cambridge, MA: MIT Press.
————(2000b). 'Financial Contagion', *Journal of Political Economy,* 108: 1–33.
————(2000c). 'Bubbles and Crises', *Economic Journal,* 110: 236–55.
————(2004). 'Financial Intermediaries and Markets', *Econometrica,* 72: 1023–61.
Aoki M. and H. Patrick (eds.) (1994). *The Japanese Main Bank System: Its Relevancy for Developing and Transforming Economies*, New York: Oxford University Press, 592–633.
Babus, A. (2007). 'The Formation of Financial Networks', Discussion Paper 06–093, Tinbergen Institute.
Bernanke, B. (2008). 'Reducing Systemic Risk', Opening Speech at the 2008 Jackson Hole Symposium, Federal Reserve Bank of Kansas City.
Bhattacharya, S. and G. Chiesa (1995). 'Financial Intermediation with Proprietary Information', *Journal of Financial Intermediation,* 4: 328–57.
Bolton, P. and D. Scharfstein (1996). 'Optimal Debt Structure and the Number of Creditors', *Journal of Political Economy,* 104: 1–25.
Boot, A. (2000). 'Relationship Banking: What Do We Know?, *Journal of Financial Intermediation,* 9: 7–25.
——and A. Thakor (1997). 'Financial System Architecture', *Review of Financial Studies,* 10: 693–733.
Boss, M., H. Elsinger, S. Thurner, and M. Summer (2004). 'Network Topology of the Interbank Market', *Quantitative Finance,* 4: 1.8.
Brusco S. and F. Castiglionesi (2007). 'Liquidity Coinsurance, Moral Hazard and Financial Contagion', *Journal of Finance,* 62: 2275–302.
Bryant, J. (1980). 'A Model of Reserves, Bank Runs, and Deposit Insurance', *Journal of Banking & Finance,* 4: 335–44.
Carletti, E. (2004). 'The Structure of Bank Relationships, Endogenous Monitoring and Loan Rates', *Journal of Financial Intermediation,* 13: 58–86.
——V. Cerasi, and S. Daltung (2007). 'Multiple Bank Lending: Diversification and Free-Riding in Monitoring', *Journal of Financial Intermediation,* 16: 425–51.
Carlsson, H. and E. van Damme (1993). 'Global Games and Equilibrium Selection', *Econometrica,* 61: 989–1018.
Castiglionesi, F. and N. Navarro (2007). 'Optimal Fragile Financial Networks', Working Paper, Tilburg University.
Chari, V. and R. Jagannathan (1988). 'Banking Panics, Information, and Rational Expectations Equilibrium', *Journal of Finance,* 43: 749–60.
Chen, Q., I. Goldstein, and W. Jiang (2007). 'Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Flows', Working Paper, University of Pennsylvania.
Cifuentes, R., G. Ferrucci, and H. Shin (2005). 'Liquidity Risk and Contagion', *Journal of European Economic Association,* 3.
Cocco, J., F. Gomes, and N. Martins (2005). 'Lending Relationships in the Interbank Market', Working Paper, London Business School.

Dasgupta, A. (2004). 'Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics', *Journal of the European Economic Association,* 6: 1049–84.

Degryse, H. and G. Nguyen (2007), 'Interbank Exposures: An Empirical Examination of Systemic Risk in the Belgian Banking System', *International Journal of Central Banking,* 3: 123–71.

—— and S. Ongena (2008), 'Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents', in A. Thakor and A. Boot (eds.), *Handbook of Financial Intermediation and Banking,* Amsterdam: Elsevier.

Demirgüç-Kunt, A. and V. Maksimovic (1998). 'Law, Finance, and Firm Growth', *Journal of Finance,* 53: 2107–37.

Detragiache, E., P. Garella, and L. Guiso (2000). 'Multiple vs. Single Banking Relationships: Theory and Evidence', *Journal of Finance,* 55: 1133–61.

Dewatripont, M. and E. Maskin (1995). 'Credit and Efficiency in Centralized and Decentralized Economies', *Review of Economic Studies,* 62: 541–55.

de Vries, C. (2005). 'The Simple Economics of Bank Fragility', *Journal of Banking & Finance,* 29: 803–25.

Diamond, D. (1984). 'Financial Intermediation and Delegated Monitoring', *Review of Economic Studies,* 51: 393–414.

—— and P. Dybvig (1983). 'Bank Runs, Deposit Insurance, and Liquidity', *Journal of Political Economy,* 91: 401–19.

Edwards, J. and K. Fischer (1994). *Banks, Finance and Investment in Germany*, Cambridge, U.K.: Cambridge University Press.

Freixas, X., B. Parigi, and J. Rochet (2000). 'Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank', *Journal of Money, Credit and Banking,* 32: 611–38.

Furfine, C. (2003). 'The Interbank Market During a Crisis', *Journal of Money, Credit and Banking,* 35: 111–28.

Gale, D. (1991). 'Optimal Risk Sharing through Renegotiation of Simple Contracts', *Journal of Financial Intermediation,* 1: 283–306.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.

Goldsmith, R. (1969). *Financial Structure and Development*, New Haven, CT: Yale University Press.

Goldstein, I. and A. Pauzner (2005). 'Demand-Deposit Contracts and the Probability of Bank Runs', *Journal of Finance,* 60: 1293–327.

Gorton, G. (1988). 'Banking Panics and Business Cycles', *Oxford Economic Papers,* 40: 751–81.

—— and F. Schmid (2000). 'Universal Banking and the Performance of German Firms', *Journal of Financial Economics,* 58: 29–80.

—— and A. Winton (2003). 'Financial Intermediation', in G. Constantinides, M. Harris, and R. Stulz (eds.), *Handbook of the Economics of Finance*, Volume 1A, Chapter 8, Amsterdam: North Holland, 431–552.

Hart, O. and J. Moore (1988). 'Incomplete Contracts and Renegotiation', *Econometrica,* 56: 755–85.

Hayashi, F. (2000). 'The Main Bank System and Corporate Investment: An Empirical Reassessment', in M. Aoki and G. Saxonhouse (eds.), *Finance, Governance, and Competitiveness in Japan*, Oxford and New York: Oxford University Press, 81–97.

Hellwig, M. (1991). 'Banking, Financial Intermediation and Corporate Finance', in A. Giovannini and C. Mayer (eds.), *European Financial Integration*, New York: Cambridge University Press, 35–63.

——(1994). 'Liquidity Provision, Banking, and the Allocation of Interest Rate Risk', *European Economic Review,* 38: 1363–89.

Hoshi, T., A. Kashyap, and D. Scharfstein (1990). 'The Role of Banks in Reducing the Costs of Financial Distress in Japan', *Journal of Financial Economics,* 27: 67–68.

Kindleberger, C. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*, New York: Basic Books.

Huberman, G. and C. Kahn (1988). 'Limited Contract Enforcement and Strategic Renegotiation', *American Economic Review,* 78: 471–84.

Iyer, R. and J. L. Peydro-Alcalde (2007). 'The Achilles Heel of Interbank Markets: Financial Contagion Due to Interbank Linkages', Working Paper, University of Amsterdam.

Lagunoff, R. and S. Schreft (2001). 'A Model of Financial Fragility', *Journal of Economic Theory,* 99: 220–64.

Levine, R. (2002). 'Bank-Based or Market-Based Financial Systems: Which is Better?', *Journal of Financial Intermediation,* 11: 398–428.

——and S. Zervos (1998). 'Stock Markets, Banks and Economic Growth', *American Economic Review,* 88: 537–58.

Mitchell, W. (1941). *Business Cycles and Their Causes*, Berkeley: University of California Press.

Morris, S. and H. Shin (1998). 'Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks', *American Economic Review,* 88: 587–97.

Padilla, A. J. and M. Pagano (1997). 'Endogenous Communication Among Lenders and Entrepreneurial Incentives', *Review of Financial Studies,* 10: 205–36.

Rajan, R. (1992). 'Insiders and Outsiders: The Choice Between Informed and Arm's-Length Debt', *Journal of Finance,* 47: 1367–400.

Rochet, J. and X. Vives (2004). 'Coordination Failures and the Lender of Last Resort: Was Bagehot Right after All?', *Journal of the European Economic Association,* 2: 1116–47.

Sharpe, S. (1990). 'Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships', *Journal of Finance,* 45: 1069–87.

Tadassee, S. (2002). 'Financial Architecture and Economic Performance: International Evidence', *Journal of Financial Intermediation*, 11: 429–54.

Upper, C. (2006). 'Contagion Due to Interbank Credit Exposures: What Do We Know, Why Do We Know It, and What Should We Know?', Working Paper, Bank for International Settlements.

——and A. Worms (2004). 'Estimating Bilateral Exposures in the German Interbank Market: Is There a Danger of Contagion?', *European Economic Review,* 48: 827–49.

Von Thadden, E. (1995). 'Long-Term Contracts, Short-Term Investment and Monitoring', *Review of Economic Studies,* 62.

Yosha, O. (1995). 'Information Disclosure Costs and the Choice of Financing Source', *Journal of Financial Intermediation,* 4: 3–20.