Competition and regulation in banking*

Elena Carletti
Center for Financial Studies
at the University of Frankfurt
Merton Str. 17-21
60325 Frankfurt
Germany
carletti@ifk-cfs.de

August 11, 2007

*I would like to thank Franklin Allen, Hans Degryse, Martin Hellwig, Steven Ongena and Xavier Vives (the editor) for valuable comments and discussions.

1 Introduction

It is well known that banks are special in that they are vulnerable to instability. As the numerous episodes of crises show, banks are fragile and are prone to take excessive risks. Their function as intermediaries between firms and borrowers and the maturity transformation they operate in their asset-liability management make banks play an important role as providers of liquidity to depositors, but also expose them to runs and systemic crises. The great reliance on deposits as source of funds creates a severe agency problem between banks and depositors in that, being subject to limited liability, banks do not bear the downside risk and have strong incentives to choose risks that are excessive from the viewpoint of depositors. The need of a stable banking sector, together with that of protecting consumers, provides the motivation for the introduction of deposit insurance schemes and lender of last resort facilities. These safety arrangements are effective in pursuing a stable system, but introduce several distortions and call for further regulatory measures, such as capital requirements.

Whereas the speciality of the banking system and the need of regulation have attracted much attention both in academic and policy debates, the issue of how competition affects the stability of the system and the effectiveness of regulation is not well understood yet. The desirability of competition in the banking sector has been questioned for a long time. Following the crises of the 1930s, competition was kept limited in an attempt to preserve stability. The process of deregulation in the last decades lifted many of the restrictions on competition, and opened up the possibility for banks to expand their investments in riskier activities and new locations. A new wave of failures followed in the 1980s and 1990. The increase in competition following the deregulation wave was regarded as the main reason behind this new instability. As found by Keeley (1990), the decline of banks' margins and charter values magnified the agency problem between banks and depositors (or deposit insurance fund), thus inducing banks to take excessive risks and increasing dramatically their failure probabilities.

The idea of a negative relationship between competition and stability has been pervasive in the literature since the 1990s, but more recent contributions indicate that the relationship

is much more complex. What are the trade-offs between competition and stability? How does competition affect the vulnerability of banks to runs and systemic crises, and their incentives to take risk? How does competition influence the effectiveness of the regulatory tools aiming at preserving stability? Can regulation "correct" the potential negative effects of competition on stability? This paper aims at providing insights to these questions by reviewing the literature on competition, stability and regulation in banking.

We start with looking at these issues separately. First, we briefly describe the reasons behind the risk of instability in the banking sector, and the need of regulation. Following what already mentioned above, we distinguish between sources of instability on the liability side (runs and systemic crises) and on the asset side (excessive risk taking), and discuss how regulation can help achieving a stable system. Then, we analyze how competition operates in this sector. The main conclusion is that, as often argued, the standard competitive paradigm is not appropriate for the banking industry. The presence of important markets failures changes dramatically the nature of competition and its outcome. Asymmetric information, switching costs and network externalities create entry barriers and allow banks to retain some market power in the form of informational rents or enhanced differentiation. Interestingly, this literature proceeds by taking the behavior of agents as exogenous. There is no concern for banks' incentives to take risk or depositors' desire to run prematurely. The only focus in on how the competitive mechanism operates in the presence of market failures. This is a very different approach from the stability literature, which instead focuses exclusively on the behavior of agents within the context of agency theory.

To understand better the link between competition and stability, we then review the literature addressing how competition affects the fragility and the risk taking problem, as well as the need of regulating the sector. Surprisingly, the issues have not been studied as extensively as one might expect. Despite a growing interest, the literature is still rather limited and inconclusive on many aspects of the trade-off between competition and stability. What emerges is that, whereas the literature on stability is centered around banks' vulnerability to bank runs and systemic crises, most of the contributions analyzing the impact of competition

on stability have instead addressed the impact of competition on banks' incentives to take risk, and the possibility of correcting its perverse impact through appropriate regulatory measures. Besides the limited focus, the literature is also still inconclusive. Whereas the prevailing view is that competition worsens the risk taking problem because lower margins and charter values increase the attractiveness of risky investments, some recent contributions have shown that competition may actually lead to the opposite result of improving the risk of banks' portfolios once specific features of the banking system, such as the relationship with borrowers or banks' monitoring function, are explicitly taken into account.

Regulation may help mitigating the trade-off between competition and stability, as long as such trade-off exists. But how to design regulation appropriately? Again, the literature is inconclusive. Despite there seems to be consensus on the negative effect of flat deposit insurance premia, the results are split for what concerns the effectiveness of capital regulation. While this seems effective in some contexts, it needs to be complemented by direct restrictions of competition such as interest rate ceilings in others. Overall, what emerges is again the need of further attention and research on the impact of competition on stability as well as on the appropriate design of regulation.

The rest of the paper proceeds as follows. Section 2 introduces the issue of bank stability, distinguishing between vulnerability to runs and systemic crises, and excessive risk taking. Section 3 reviews the contributions on the functioning of competition in the presence of asymmetric information, switching costs and network externalities. Section 4 analyzes more deeply the link between competition and stability, and in particular the link between market structure and financial fragility, and between market structure and excessive risk taking. Section 5 looks at the impact of regulatory tools on the trade-off between competition and stability. Section 6 concludes.

2 Bank instability and the need of regulation

It is well known that banks are special because they are more vulnerable to instability than firms in other sectors, and also because less wealthy people may hold some non-negligible

share of their wealth in various forms of bank deposits. The potential instability of the banking system and the need of consumer protection are the fundamental rationales behind the introduction and development of regulation.

The course of events and in particular the US experience suggest two possible connotations of the term "instability": the crises occurred in the 1930s show that the banking system is fragile since it is vulnerable to runs and panics; the massive distress which came to light in the 1980s and 1990s demonstrates that intermediaries may have strong incentives to assume excessive risk and that, as a result, the system has a high probability of failure.

Bank fragility: individual runs and systemic crises[1]

Intermediaries emerge as a response to the imperfection and incompleteness of financial markets. In an economy characterized by asymmetric information and uncertainty, intermediaries are valuable because they have economies of scale in producing information and provide insurance to depositors who are uncertain in their timing of consumption. Information production and insurance provision are the two main characteristics of bank specificity, but are also the sources of their fragility. The informational asymmetries existing between banks, borrowers and depositors, and the maturity transformation that banks operate by investing short-term deposits in long-term assets expose banks to the possibility of runs. Banks offer depositors demandable contracts which allow depositors to withdraw a fixed amount on demand. If the total value of the early withdrawals exceeds the amount available from short-term investments, a run originates and the bank has to sell its illiquid assets. This illiquidity problem may turn into insolvency and force the premature liquidation of the bank if no assets are left after satisfying the early withdrawals.

To illustrate the basic mechanism triggering a run, consider a three-date economy, with one bank operating under perfect competition and raising funds from a continuum of depositors of measure one. The bank invests a fraction M in a short-term asset and a fraction $1 - M$ in a long-term asset. The former simply transfers the unit invested from date 0 to

[1]The literature on invidual runs and systemic crises is vast. We describe here only a few contributions. Excellent broader reviews are contained in Gorton and Winton (2003) and De Bandt and Hartmann (2002).

date 1, while the latter yields $R > 1$ at date 2 and $\ell < R$ if interrupted prematurely at date 1. Depositors are all ex ante identical, but face a preference shock at date 1. A fraction t of them becomes of type 1 (early type) and wishes to consume at date 1, while the remaining fraction $1 - t$ turns to be of type 2 (late type) and prefers consuming at date 2. Depending on the specific assumptions on the return of the long-term investment and on the structure of the preference shocks, the rationale behind depositors' withdrawal differs and runs can be either irrational or information-induced events.

Following Diamond and Dybvig (1983), suppose initially that the return R and the fraction t of early depositors are deterministic, and that the liquidation value of the long-term investment is $\ell = 1$. Then, the bank offers a deposit contract to depositors so as to maximize

$$U^* = \max_{\{c_{ij}\}} tU^1(c_{11}) + (1-t)U^2(c_{22}) \tag{1}$$

subject to:

$$tc_{11} \leq M \tag{2}$$

$$(1-t)c_{22} \leq R(1-M) \tag{3}$$

$$U^1(c_{11}) \geq U^1(c_{22}) \tag{4}$$

$$U^2(c_{22}) \geq U^2(c_{11}), \tag{5}$$

where expression (1) is depositors' expected utility with c_{ij} being the consumption of type j at date i; constraints (2) and (3) represent the resource balance constraints at dates 1 and 2 respectively; and conditions (4) and (5) are the incentive compatibility constraints stating that the deposit contract should be designed so that each type of depositors prefers its own withdrawal profile. If depositors are risk averse with $RRA > 1$, i.e., $-u''(c)/u'(c) > 1$, the optimal deposit contract satisfies

$$1 < c_{11}^* < c_{22}^* < R.$$

This result shows that the deposit contract offers insurance to depositors and is Pareto improving relative to the autarkic situation, where individuals invest directly. The insurance provision, however, makes the bank vulnerable to runs. There is a good equilibrium which

realizes optimal risk sharing when depositors choose the withdrawal decisions embedded in the deposit contract; but there is also a bad equilibrium in which all depositors withdraw their funds prematurely and the bank collapses. The condition $c_{11}^* > 1$ implies that depositors find it optimal to withdraw if they simply fear that others will withdraw first. There is no rational motivation behind such a panic run other than a coordination failure due to sunspots. The possibility of a run is intrinsic in the provision of insurance. If $c_{11}^* \leq 1$, no run would occur.

An alternative explanation for the occurrence of bank runs is that they are linked to changes in fundamental variables and are therefore information-based (or fundamental based). If the return on the long-term investment is stochastic, the perspective of a negative shock increases the probability that the bank is unable to meet its future commitments. If depositors anticipate this, they withdraw their funds and force the premature closure of the bank. Jacklin and Bhattacharya (1988) formalize this mechanism in a context where the assumptions of an illiquid and risky long-term asset and depositors with RRA<1 lead to an optimal contract satisfying

$$c_{11}^* < 1 < c_{22}^*.$$

Whereas this solution excludes the possibility of irrational runs, it still leaves room for information-based runs. After the contract is signed, some type 2 depositors receive a partial signal s describing the posterior distribution of the success probability of the long-term asset. Thus, they update their priors p and choose to withdraw prematurely whenever

$$\widehat{E}\left[U^2(c_{11})\right] > \widehat{E}\left[U^2(c_{22})\right], \tag{18}$$

that is whenever the expected utility from withdrawing prematurely calculated using the posterior beliefs on the success probability p, $\widehat{E}\left[U^2(c_{11})\right]$, exceeds the expected utility from waiting and receiving the consumption profile initially designed for them, $\widehat{E}\left[U^2(c_{22})\right]$. This triggers a run. Given the total illiquidity of the long-term asset, the bank does not have enough funds to satisfy the withdrawal demands at date 1, and has to close down. The origin of the run is now the rational response of depositors to the arrival of sufficiently negative

information on the future solvency of the bank. Therefore, the run is information-based and is efficient, as long as it leads to the liquidation of an impending insolvent bank.

Panic and information-based runs can also be related, as shown by Chari and Jagannathan (1988). The analysis focuses on the signal-extraction problem faced by uninformed depositors in their withdrawal decisions in a framework characterized by shocks to asset returns and to the proportions of early depositors and informed depositors. Late type depositors who remain uninformed know that other depositors may be informed on the future return of bank asset and try to infer such information from the size of the withdrawal queue at date 1. However, since the proportion of early depositors is stochastic and unobservable, uninformed depositors may not be able to infer bank's future performance correctly. In particular, they may not be able to distinguish whether a long queue is formed by the informed depositors receiving a negative signal, or simply by a large proportion of early depositors wishing to consume early. A pure panic run generates from uninformed depositors' confusion between insolvency and high liquidity shocks. It occurs when uninformed depositors withdraw prematurely for *fear* that some depositors have received a bad signal on the bank's future performance in cases where no one is informed about it.

A common feature in this strand of literature is the presence of multiple equilibria, in one of which a bank run occurs. A potential problem with this approach is that individuals may not want to deposit in the first place since they cannot calculate the probability of a run occurring. Consequently, runs should not be observed in equilibrium as no one would deposit anticipating a run. This leaves open the important question of whether the emergence of banks as liquidity providers is desirable from an ex-ante perspective. One way around the multiplicity of equilibria, as suggested by Postlewaite and Vives (1987), is to associate a bank with a sort of Prisoner's Dilemma-type situation in which agents withdraw their deposits for self-interest reasons rather than for consumption reasons. In this context, agents do not condition their behavior on any exogenous event, and there is only a unique equilibrium involving a positive probability of a bank run. A bank run may occur because depositors have incomplete information about the liquidity shocks they face.

Another way around the multiplicity of equilibria is suggested by Rochet and Vives (2004) and Goldstein and Pauzner (2005). They analyze a modification of the Diamond and Dybvig model, in which the fundamentals of the economy uniquely determine whether a bank run occurs. The key features of the analysis are the assumptions that fundamentals are stochastic and investors obtain noisy private signals on the realization of the fundamentals. This leads to a unique equilibrium in which a bank run occurs when the fundamentals are below some critical level. Importantly, despite being determined by fundamentals, runs can be also driven by bad expectations. Depositors tend to withdraw prematurely for fear that others will do so. Thus a run may occur even when the economy environment is sufficiently good that a run would not occur if depositors had not had bad expectations on other depositors' actions. In this respect the model reconciles the view of bank runs as panics due to coordination failure and the one of runs as being linked to fundamentals. The uniqueness of the equilibrium allows also the determination of the ex ante probability of a bank run. This is increasing in the short-term payment the bank offers and therefore in the risk-sharing embodied in the banking contract. When the short-term asset is set at the autarkic level, only efficient runs occur. Depositors withdraw prematurely only if the long-term return of the bank's asset is lower than its liquidation value. In contrast, when the short-term asset is above the autarkic level, inefficient runs occur when a bank is forced to liquidate the long-term asset even though it has a high expected return. Given this inefficiency, having banks offering short-term payments above the autarkic level is viable and desirable provided that maintaining the long-term investments till maturity is generally efficient.

To sum up, bank runs result from either(both) a coordination failure among depositors or(and) an expectation of poor performance of the bank. Runs may be costly, because they force the interruption of a production process and the premature liquidation of assets. Moreover, runs may trigger a systemic crisis, if they propagate through the economy. A systemic crisis has a narrow and a broad interpretation (De Bandt and Hartmann, 2002). A crisis in the narrow sense refers to a situation in which the failure of one bank, or even only the release of bad news about its state of solvency, leads in a sequential fashion to the failure of numerous other banks or of the system as a whole. A crisis in the broad sense

includes also the simultaneous failure of many banks or of the whole system as result of a generalized adverse shock. The sequential spread out of failures in a narrow crisis implies a strong spillover effect, defined as contagion, which can take place through contagious runs or domino effects (Schoenmaker, 1996). The former refers to the propagation of a run from a single bank to other banks. As for individual runs, such propagation can be due to sunspots or be information-based. The domino effect refers to the mechanism through which difficulties faced by a single bank spread to others through the payment system and/or the interbank market. If relationships among banks are neither collateralized nor insured against, the distress of one bank may trigger a chain of subsequent failures. Other banks may incur a liquidity or an insolvency problem depending on the intensity of the linkages with the distressed bank, and on the correlation of shocks in the system. The channels of contagious runs and domino effects can work in conjunction as well as independently. In most cases, however, a systemic crisis is the result of the propagation of an individual failure through both of them.

Most of the interrelations among banks occur through the payment system. Their internal arrangement determines how individual shocks propagate, and thus the severity of the contagion risk. Depending on the timing and the methodology of settlement, payment systems can be classified in net settlement systems (only net balances are settled and at a certain point in time), pure gross systems (payments between members are settled without netting and a certain point in time), real-time gross systems (payments between members are settled without netting and immediately after every transaction) and correspondent banking (payments are settled bilaterally between a correspondent bank and members of a group of small or foreign banks). Net systems economize on liquidity, but expose banks to contagion because they involve the transfer of asset claims from one location to another. By contrast, gross systems entail high liquidity costs, but do not face any risk of contagion (Freixas and Parigi, 1998).

Surprisingly, the academic literature has devoted attention to the issue of contagion and systemic risk only very recently. The former models of individual runs can be read in terms

of generalized systemic crises, but they are not suited for the analysis of the propagation mechanism of individual failures. The analysis of such mechanism requires models with multiple banks.

Rochet and Tirole (1996) examine the domino effect in a model of interbank lending with heterogenous banks. Some banks are good at collecting deposits but have poor investment opportunities, while others have plenty of investment opportunities but needs funds. This leaves room for interbank lending, although it exposes banks to the risk of contagion. If a borrowing bank is hit by a liquidity shock, the lending bank may be negatively affected and be forced to shut down. The survival of the lending bank depends on the severity of the shock affecting the borrowing bank and on the revenues (or losses) of the interbank loan. Clearly, the greater the liquidity shock faced by the borrowing bank, the more likely is the closure of the lending bank.

The occurrence of contagious runs and domino effects is analyzed by Allen and Gale (2000a) in an economy where banks hold inter-regional deposits on other banks to insure against liquidity preference shocks. The economy works well and achieves optimal risk sharing when there is no aggregate uncertainty; but it may lead to a systemic crisis when there is an excess aggregate demand for liquidity. In such a case, each bank starts to withdraw deposits from banks in other regions in an attempt to satisfy depositors' withdrawal demands and avoid liquidating the long-term assets. This mutual liquidation denies liquidity to the troubled bank, which then experiences a run. Depending on the structure of the interbank market, the individual run propagates to other banks and leads to a systemic crisis. If regions are well connected (complete interbank market), contagion is avoided. If connections among regions are limited (incomplete interbank market) and liquidity shocks are strong enough, contagion arises.

In a similar spirit, Freixas et al. (2000) analyze the risk of contagious runs through the payment system when banks are located in different regions and face both liquidity and solvency shocks. The former originate from depositors' geographical consumption preferences, while the latter from shocks to the return of bank assets. Depositors have two ways

to satisfy their wish of consuming in a different location from where they have deposited initially. They can withdraw their funds and transfer cash to the other region, or they can transfer deposits from one bank to another through the payment system. When banks are subject only to liquidity shocks, the economy shows multiple equilibria. Either depositors do not run and the payment system is efficient in reducing the opportunity costs of holding liquid assets; or depositors run and banks have to liquidate the long-term assets (speculative gridlock equilibrium). This latter equilibrium resembles the sunspot equilibrium in Diamond and Dybvig. When banks face also (idiosyncratic) solvency shocks, the stability of the system depends on the architecture of the payment system. As in Allen and Gale (2000a), the closure of an insolvent institution is less likely to generate contagious runs when payment systems are well diversified.

Excessive risk taking

A second source of instability of the banking system relates to risk-taking on the asset side. As it is well known from agency theory, in a principal-agency relationship the objectives of the involved parties are not perfectly aligned so that the agent does not always act in the best interest of the principal. The problem can be limited by designing appropriate incentive schemes for the agent or by controlling his decisions through costly monitoring. In general though, the divergence of interests will not be completely resolved, at least not at zero cost. Applying these arguments to corporate finance, it is easy to see that there is a misalignment in the objectives of debtholders and firm managers. Even if all parties are utility maximizers, their attitude toward risks diverges. Whereas debtholders bear the downside risk, the manager pursuing shareholders' interests benefit from upside potential. Thus, the manager has strong incentives to engage in activities that have very high payoffs but very low success probabilities (Jensen and Meckling, 1976).

While this agency problem is present in all leveraged firms, two features of the banking system makes it more severe among banks. First, the opacity and the long maturity of banks' assets make it easier to cover any misallocation of resources, at least in the short run. Second, the wide dispersion of bank debt among small, uninformed (and often fully

insured) investors prevents any effective discipline on banks. Thus, because banks can behave less prudently without being easily detected or paying additional funding costs, they have stronger incentives to take risk than firms in other industries.

To illustrate the agency problem between banks and depositors, we use a simple model adapted from Holmstrom and Tirole (1997), Cerasi and Daltung (2000) and Carletti (2004). Consider a two-date economy ($T = 0, 1$), in which at date 0 a bank invests in a project, which yields a return R if successful and 0 if unsuccessful. The success probability of the project depends on the monitoring effort $m \in [0, 1]$ that the bank exerts. It is p_H if the bank monitors, and p_L if it does not, with $p_H > p_L$, $\Delta p = p_H - p_L$ and $p_H R > 1 > p_L R$. Monitoring is costly; an effort m entails a private cost $C(m) = \frac{c}{2}m^2$. The choice of the monitoring effort depends crucially on the financing structure of the bank. If it is self-financed, it chooses m so to maximize its expected profit

$$\Pi = mp_H R + (1 - m)p_L R - y - \frac{c}{2}m^2,$$

where y represents the return on an alternative safe investment. In this case the first order condition gives

$$m = \frac{\Delta p R}{c}.$$

By contrast, if the bank raises external funds in the form of debt with promised (gross) return r_D, it chooses m so as to maximize

$$\Pi = mp_H(R - r_D) + (1 - m)p_L(R - r_D) - \frac{c}{2}m^2.$$

The first order condition is then given by

$$m = \frac{\Delta p(R - r_D)}{c}.$$

Clearly, raising deposits reduces the equilibrium monitoring effort. The reason is that the bank has now to share the benefit of greater monitoring with depositors. If the deposit rate is set before m is chosen, increasing monitoring simply raises the probability of repaying depositors without reducing the funding costs. This worsens the bank's incentive and leads to a lower equilibrium effort.

The need of regulation

The vulnerability of banks to runs and systemic crises and the consequent concern for consumers' wealth are the main factors justifying the need of regulation and safety net arrangements in the form of deposit insurance and lender of last resort. For example, as shown by Diamond and Dybvig (1983), deposit insurance prevents the occurrence of panic (sunspot) runs without reducing banks' ability to transform short-term liabilities into long-term assets. A demand deposit contract with government deposit insurance achieves optimal risk sharing among depositors as unique Nash equilibrium. Government's ability to levy non-distortionary taxes and deposit insurance guarantees induce depositors not to withdraw prematurely. Consequently, bank liquidation policy is independent of the volume of withdrawals, no strategic issues of confidence arise and no bank runs take place.

The underlying idea behind the introduction of regulation and safety net arrangements is that runs and systemic crises are inefficient and therefore have to be prevented. Whereas this is always true for panic runs, it may not be the case for information-based runs. These are efficient whenever the liquidation value of the long-term asset is higher than its long-term expected return. Given this distinction, it is important to understand why bank runs occur and eventually how to deal with them. Allen and Gale (1998) analyze the potential costs of bank runs and the need for central bank intervention. In their model, bank runs are information-based events which play the important role of sharing risk among depositors. Their welfare properties depend on the potential costs of early withdrawals. When withdrawing early involves no costs, runs are efficient since they occur only when banks' long-term asset returns are low. The optimal deposit contract reaches the first-best solution in terms of both risk sharing and portfolio choice, and regulation is not needed. In contrast, when there are real costs associated with early withdrawals (e.g., because the return of the safe asset is higher within the banking system than outside), bank runs reduce the consumption available to depositors. Then laissez-faire does not achieve the first-best allocation any longer and there is scope for central bank intervention in the form of money injection. If the central bank grants an interest-free loan to banks when runs occur, banks

can avoid liquidating the safe assets prematurely and depositors receive higher consumption levels. The first-best allocation can then be achieved again by a combination of standard deposit contracts, runs and policy intervention. In a similar spirit, central bank intervention is needed when the long-term asset can be liquidated and traded on market. Bank runs are again costly, and the premature liquidation of long-term assets forces down the price in the market and makes crises worse. The intervention of the central bank is needed to prevent the collapse of asset prices.

The issue of the optimal form of central bank intervention have long been debated in the academic literature. According to the "classical" view of Bagehot (1873), central banks should lend freely at a penalty rate and against good collateral. This should prevent banks from using central bank lending to fund current operations and should guarantee that emergency liquidity loans are extended only to illiquid but solvent banks. This view has been criticized in various ways. First, according to Goodhart (1987), it is virtually impossible, even for central banks, to distinguish illiquidity from insolvency at the time the lender of last resort (LOLR) should act. Banks demanding such assistance are under a suspicion of insolvency since they could otherwise raise funds from the market. Second, it has been argued for example by Goodfriend and King (1988) that there is no need for central bank's loans to individual banks since open market operations are sufficient to deal with systemic liquidity crises. In other words, LOLR should intervene at the macroeconomic level but not at the microeconomic level.

This debate is also relevant with respect to the possible consequences that the safety net arrangements can create. If on the one hand both deposit insurance and LOLR may suffice to prevent runs and systemic crises, on the other hand they have side effects and bring in new inefficiencies. For example, they worsen the problem of excessive risk taking and call for further regulatory measures. Both deposit insurance and a systematic use of the lender of last resort induce banks to undertake greater risks since depositors do not have incentives to monitor their banks' asset values and can rely on future bailout in case of distress (e.g., Merton, 1977; Boot and Greenbaum, 1993).

Some of these issues have been recently addressed in formal theoretical models. Rochet and Vives (2004) provide a possible theoretical foundation of Bagehot's view using the "global game" approach. Their analysis builds on a model of banks' liquidity crises with a unique Bayesian equilibrium. At this equilibrium there is an intermediate range of values of the bank's asset in which, due to a coordination failure, depositors may run despite the bank being solvent. Thus, as argued by Bagehot, a solvent bank may face a liquidity problem and be in need of assistance. The likelihood of this happening decreases with the ex ante strength of the fundamentals. The optimal policy consists of prudential measures and ex post emergency loans. Liquidity and solvency regulation can solve depositors' coordination problem and avoid the failure of solvent banks, but may be too costly in terms of foregone returns. Thus they need to be complemented by emergency discount-window lending. The optimal policy is richer when bank managers can exert an effort and influence the risk of asset returns since it has to account for the effect it has on bank managers' incentives. Depending on the value of the fundamentals, the optimal policy may comprise early closure of solvent banks to prevent moral hazard and emergency liquidity assistance.

The relationship between banks' moral hazard and optimal central bank intervention is also analyzed by Freixas et al. (2004) in a model in which banks are subject to both liquidity and solvency shocks and can operate under moral hazard either in screening loan applicants or in monitoring borrowers. Given the difficulty to discern banks' solvency state, insolvent banks may be able to borrow from the interbank market or from the central bank and "gamble for resurrection" (i.e., invest in projects with negative net present value). The optimal policy depends on the nature of banks' incentive problem. If banks face moral hazard in monitoring borrowers, there is no need for central bank intervention. A secured interbank market suffices to implement the first-best allocation. In contrast, if banks face moral hazard mainly in screening loan applicants, the central bank should provide emergency liquidity assistance but at a penalty rate to discourage insolvent banks from borrowing.

Gale and Vives (2002) builds on the time-inconsistency embodied in central bank bailout policy to characterize the optimality of dollarization as a way for devaluing depositors' claims

and avoiding bank failure. The idea is that competitive banking systems lead to excessive liquidation when banks face moral hazard problems. By using dollarization as a credible commitment not to bail out banks, the central bank can then implement the incentive-efficient solution and avoid failures.

To sum up, the debate around the optimal central bank intervention centres on the trade-off between the benefits (prevention of crises) and the costs (distortion of incentives and moral hazard problem) of bailing out distressed banks. This trade-off may call for other regulatory measures, such as capital regulation, rate regulation and entry restrictions. These too, though, have been heavily criticized as not being effective or inducing other negative distortions, such as a reduction in competition. The sides effects of regulation are therefore crucial for understanding the role and the importance of competition in the banking sector.

3 Competition in banking

Analyzing how competition works in the banking sector and whether it is beneficial is a difficult task. On the one hand, the general argument in favor of competition in terms of cost minimization and allocative efficiency apply to the banking industry. On the other hand, however, the presence of various market failures distorts the functioning of competition and makes the standard competitive paradigms not appropriate for the banking sector. The presence of asymmetric information in corporate relationships and of switching costs and networks in retail banking alters the market mechanism. This creates significant entry barriers, which affect the industry structure and lead to an ambiguous relation between the number of banks and the competitive outcome. We analyze these effects more in detail below.

It is worth noting though that also other aspects of the role and specificity of banks affect the working of competition in this sector. For example, the simple fact that banks compete on both sides of the balance sheet may lead to departures from the competitive outcome. Stahl (1988) and Yannelle (1988, 1997) show that when banks compete for both loans and deposits, they may want to corner one market in an attempt to achieve a monopoly on the

other. Furthermore, the role of banks as financiers of industrial loans may create endogenous entry barriers in both the banking sector and the borrowing industries, thus leading to a natural monopoly in both sectors (González-Maestre and Granero, 2003).

Competition under asymmetric information

As already mentioned, banks emerge as intermediaries between depositors and borrowers. Thus, their two main functions are to provide insurance to depositors, and to screen and monitor investment projects. The former creates the risk of instability; and the latter creates important informational asymmetries among banks and potential borrowers and among banks themselves, which may distort the competitive mechanism significantly.

Broecker (1990) analyzes how competition in the credit market affects the screening problem banks face in the choice of granting loans. The set up is such that firms applying for credit differ in their ability to repay loans, i.e., in their credit-worthiness, and banks perform independent and imperfect screening tests to discern firms' quality and decide whether to grant loans. Conditional on their own test results, banks compete with each other by setting a loan rate. Given, however, that screening tests are imperfect, the competitive market mechanism does not work properly in that it leads to a negative externality among banks. Increasing the loan rate above that of the competitor has two opposite effects on the profit of the deviating bank. On the one hand, it increases its profit through the usual price effect. On the other hand, it worsens the quality of firms accepting the loan, thus reducing its profit. A firm will indeed accept the least favorable loan rate only after being rejected by all other banks setting more favorable rates; but this implies that the firm has a low credit-worthiness on average. Because of this "winner's curse" problem, increasing the number of banks performing screening tests decreases the average credit-worthiness of firms, and increases the probability that a bank does not grant any loan. In the limit, the number of active banks is positive and the equilibrium maintain some degree of oligopolistic competition.

Similar conclusions are reached by Riordan (1993). Using the theory of common value auctions, he shows that a higher number of competing banks worsens the informativeness

of the signal that banks receive on firms' loan quality and make them more conservative in granting loans. Both of these two effects are detrimental for social welfare, since they reduce the quality of banks' portfolios and lead to the financing of less efficient investment projects.

The relationship between the degree of market competition (or integration) and banks' screening incentives is also analyzed by Gehrig (1998). In a context where banks use imperfect credit-worthiness tests to discriminate between good and bad projects, he shows that screening incentives increase with the profitability of loans. Thus, more intense competition due to the entry of outside banks worsens the quality of banks' portfolios, since it reduces the investment that banks make to improve the precision of their screening tests.

Besides acquiring information on borrowers through screening, banks monitor them also in the course of the relationship, thus obtaining further information on their quality. This creates an informational asymmetry among banks. If a borrower needs a renewal of the loan, the incumbent bank has better information about his quality relative to outside banks. This gives the incumbent bank an informational monopoly over its borrowers, which reduces competition from outside banks, and allows the incumbent bank to "hold-up" its borrowers and extract monopoly rents. Such expropriation disincentives the borrower from exerting more effort, thus reducing the expected return of investment projects (Rajan, 1992) and leading to an inefficient allocation of capital toward lower quality firms (Sharpe, 1990).

The heterogeneity of borrowers and the consequent informational advantages of incumbent banks affect the competitive market mechanism in several ways. As already mentioned, an increase in the number of competing banks reduces the screening ability of each of them. Consequently, more low-quality borrowers obtain financing, and banks may have to increase loan rates to compensate for the higher portfolio risk, thus leading to an inverse relationship between competition and level of loan rates (Marquez, 2002). This results may not obtain any longer, however, when information acquisition is endogenous. In such a context, competition lowers loan rates, in the usual way. Hauswald and Marquez (2005) show that when banks acquire information to soften competition and increase market shares, a higher number of banks reduces the winner's curse problem originating from competitors' superior

information, thus leading to lower loan rates. In other words, an increase in the number of competing banks reduces the degree of product differentiation among banks, and thus loan rates.

The presence of adverse selection affects also the structure of the industry. The informational advantage of the incumbent banks allows them to reject the riskier borrowers in need of refinancing. Because outside banks cannot distinguish between new borrowers and old riskier borrowers rejected by their previous incumbent banks, they face an adverse selection problem which may keep them out of the market. An equilibrium of blockaded entry may then emerge, where only two banks are active and make positive profits even under pure Bertrand price competition (Dell'Ariccia et al., 1999); or, more generally, the equilibrium is characterized by a finite number of banks even in the absence of exogenous fixed costs (Dell'Ariccia, 2001). The general idea is that the heterogeneity of borrowers and the acquisition of information gathered through lending generate endogenous fixed costs, which limit the number of active competitors.

To sum up, focusing mostly on an adverse selection problem (i.e., heterogeneity of borrowers), the literature on competition with asymmetric information discusses the possibility for lenders to exercise market power, the imperfect functioning of competitive markets, and the endogenous entry barriers which the informational advantage of incumbent banks can generate. Despite not addressing directly the consequences for stability, this literature provides some intuitions for the effect that competition may have on banks' solvency. Because banks' screening abilities worsen with the number of competing banks, tougher competition leads to riskier banks' portfolios and high failure probabilities. The mechanism behind the negative relationship between competition and stability derives exclusively from the heterogeneity of borrowers. This contrasts sharply with the mechanism in the literature on competition and stability, where the focus in on how competition modifies the behavior of either borrowers or banks. We will come back to this issue in Section 4.

Competition and switching costs

Switching costs are an important source of market power in retail banking (e.g., Diamond, 1971). In moving from one bank to another, consumers may incur costs associated with the physical change of accounts, bill payments or lack of information (Vives, 2001). The presence of switching costs produces in general two opposing effects on the degree of competition. On the one hand, they may lead to collusive behavior once banks have established a customer base which remains locked in. On the other hand, they induce fierce competition to enlarge the customer base. Thus, switching costs may lead banks to offer high rates initially to attract customers and to reduce them subsequently, when consumers are locked in.

A different result may be obtained when switching costs are combined with asymmetric information about borrowers' credit-worthiness. Bouckaert and Degryse (2004) analyze a two-period model where heterogenous borrowers face switching costs of changing banks, and banks face an adverse selection problem. In such a context, banks find it convenient to disclose their private information about borrowers' credit-worthiness and induce them to switch bank in order to soften overall competition. Disclosure of borrowers' quality removes the information disadvantage of rival banks in the interim period, thus allowing them to poach only good borrowers and have positive second-period profits. This relaxes the initial competition for enlarging the customer base, and it increases banks' overall profits. Thus, the removal of future informational entry barriers may emerge for strategic reasons as it softens overall competition.

The presence of switching costs can also affect significantly the link between number of banks and degree of market competition. Allen and Gale (2000b, 2004) show that a small fixed cost of switching banks may imply higher rates in a system with many small independent banks (unitary system) than in a system with two large banks having extensive nationwide branching networks (branching system). This result is obtained in a model characterized by fixed costs of switching banks, customers' initial limited information about the future offer of banks' services and prices, and product diversity in the services that banks provide at different locations. Consumers are allocated randomly at each location every period and have to choose which bank to patronize. In a unitary system, each bank consists of one

branch in one location. Thus, each bank can raise its initial rate by a small amount without losing its customers, because of the fixed cost of switching. The only equilibrium is when all banks charge the monopoly rate. In a branching system, there are only two banks with one branch in each of the locations. Although consumers change location in each period, they can stay with the same bank if they wish. This possibility increases the costs for each bank of deviating from the equilibrium strategy and losing customers. As a result, branching banking supports more efficient equilibria, where the two banks may charge a rate close to the perfectly competitive level.

Competition and networks

A final important element affecting the nature of competition in retail banking is the presence of networks. This introduces elements of non-price competition in the interaction between banks, thus affecting the pricing of banking products and the structure of the industry.

The possibility for banks of sharing Automatic Teller Machine (ATM) networks can be used as strategic variable to affect price competition. Matutes and Padilla (1994) analyze this issue in a two-period model, where banks choose first whether to build compatible ATM networks and then compete imperfectly on the deposit market. A large ATM network has two opposite effects. On the one hand, it allows banks to offer lower deposit rates, because depositors benefit from an easier access to their deposits when they need cash unexpectedly (network effect). On the other hand, a large ATM network increases price rivalry, because it makes banks more substitutable. Depositors benefit from the location of a bank ATM and the high rates offered by a rival bank sharing the same network (substitution effect). Banks choose to share ATM networks when the network effect dominates, i.e., when depositors do only a small number of transactions through ATMs. The equilibrium is characterized by either partial sharing of ATM networks or no sharing. The former emerges when the network effect dominates; the latter when the substitution effect prevails. Full sharing does not occur in equilibrium, since banks prefer to maintain some differentiation and face softer competition in the deposit market. However, if future entry is possible, banks can use sharing

agreements to exclude rivals from the market when the network effect is sufficiently high. Then, the threat of entry may lead all incumbents banks to share their network, since this allows them to credibly commit to fierce post-entry competition and foreclose any potential entrant.

Sharing of networks is used to limit competition also in McAndrews and Rob (1996) in a two-period model where banks choose whether to own jointly the switches in ATM networks and then compete on the pricing of ATM services. Given the presence of fixed costs in operating a switch and network effects in the demand of ATM services, banks prefer to join switches as a way to achieve a more concentrated structure in the switches industry and monopoly prices in the sale of ATM services to consumers. The implications in terms of welfare are ambiguous. Whereas the joint ownership is inefficient because it leads to the extraction of monopoly rents from final consumers, it is beneficial in that it saves the fixed costs of setting up a switch and gives consumers the possibility of benefitting from a larger network.

Similar results are obtained in a context where banks have to decide first whether to offer remote access to their customers, such as postal or telephone services, and then compete for deposits (Degryse, 1996). Depositors differ in terms of taste over location and quality (remote access). Thus, as in Matutes and Padilla (1994), the decision of a bank to offer remote access has the double effect of introducing vertical differentiation between banks and reducing the degree of horizontal differentiation. Consumers with a higher taste for remote access have lower transportation costs if this access is available. Thus, introducing remote access produces two opposite effects. It steals depositors from the rival bank (stealing effect), but it also increases the substitutability between banks (substitution effect). The equilibrium depends on which of these two effects prevails. For low and high values of the ratio quality difference to transportation cost, only one bank offers remote access and offers lower deposit rates.

The impact of networks on the structure of the industry and the possibility of entry is analyzed by Gehrig (1996) in a model that also applies to the banking sector, despite being

developed for the brokerage industry. Intermediaries reduce search frictions by facilitating the matching between buyers. Similarly to the models described above, they first choose the size of their network and then compete in prices. Setting up a large network involves costs, but it also increases the probability with which an intermediary is able to match buyers. Thus, the size of the network differentiates the quality of matching services, and intermediaries may relax price competition by choosing networks of different sizes and offering products of different quality. Intermediaries with large networks gain market power and can command higher prices than rivals with smaller networks. Vertical differentiation and fixed costs of establishing a network imply that the industry converges to an oligopolistic structure with a few large intermediaries having large networks and a number of smaller competitive intermediaries active in 'niche' markets. The different size of intermediaries allows them to differentiate and relax price competition. Still, as the size of the market grows relatively to the cost of establishing a network, the importance of smaller intermediaries vanishes as many more can enter the market and the degree of vertical differentiation among the large intermediaries disappears. Competition becomes tighter and equilibrium prices approximate perfectly competitive levels. As in McAndrews and Rob (1996), there is a trade-off between competition and fixed costs of setting up a network. An increase in the number of active players leads to more competitive outcomes but also to higher fixed costs, thus having an ambiguous effect on welfare.

One important final note is that competition in networks can also be analyzed in two–sided markets. Rochet and Tirole (2002) analyze this issue using the context of credit card associations. They develop a model in which customers' banks and merchants have market power, and consumers and merchants decide rationally whether to buy or accept credit cards. The focus is on the factors affecting merchants' resistance to accept credit cards and on the collusive determination of interchange fees, i.e., the fees that merchants' banks (the acquirers) pay to consumers' banks (the issuers). Merchants' decisions depend on their technological benefit of accepting cards, but also on the effects that card acceptance has on consumers and price competition. As in the ATM literature, merchants can then use card acceptance to increase customer base and relax price competition. Differently from the

ATM literature, however, the system has to attract two sides of the market, i.e., issuers and acquires, merchants and consumers. Thus, changes in interchange fees and prices affect the relative price structure of the two sides with important consequences on the equilibrium outcomes.

4 Competition and stability: a positive or a negative link?

In the previous sections we have described the specificity of banks and the potential sources of their instability. We have then discussed how the presence of market imperfections such as asymmetric information, switching costs and network externalities, affect the competitive mechanism in the banking sector and its outcome. Interestingly, these two strands of literature do not have much in common. In other words, the literature on competition in the presence of market failures does not say much on how competition affects the stability of the sector in the sense of either fragility or excessive risk taking. The one implication which can be drawn is that when borrowers are heterogenous and banks perform screening tests to sort out borrowers, an increase in the number of competing banks worsens the quality of the tests. This aggravates the information problem that banks face, thus increasing the riskiness of their portfolios. But this result depends entirely on the adverse selection problem that banks face vis a vis borrowers, a problem that the stability literature does not address directly.

We now turn more directly to the relationship between competition and stability. We structure the analysis according to the effects that competition, either in the deposit or in the loan market, has on the two sources of bank instability that we have outlined in Section 2. We start with discussing the effects of competition on banks' vulnerability to individual runs and systemic risk, and then we move to the effects that competition has on excessive risk-taking. It is worth pointing out though that only few of the papers we will discuss endogenize aspects of industrial organization in their analysis. The majority of them just compare the equilibria achievable in different market settings without taking account of any strategic interaction among intermediaries.

4.1 Market structure and financial fragility

The relationship between competition and financial fragility has been largely ignored in the banking literature. Most of the contributions on bank runs and systemic risk reviewed in Section 2 pay very little attention to the strategic interaction between banks, simply assuming that they operate in a perfectly competitive environment. Runs and systemic crises occur either as a consequence of a coordination failure among depositors or as their rational response to the impending banks' solvency problems. These models do not provide any insights concerning which market structure is more fragile.

A few models address directly the relationship between competition and liability risk. Smith (1984) analyses this issue in a framework á la Diamond and Dybvig (1983) where banks compete to attract depositors that have different probability distributions over the dates of withdrawal. In the case when an adverse selection problem is present, that is when depositors only know their own probability of withdrawals, there may not exist any Nash equilibrium. The equilibrium contract, either pooling or separating, is destroyed by the possibility for banks to offer positive profit contracts to a specific segment of depositors. The banking system is then not viable or "unstable". The problem can be resolved by appropriate regulatory measures, such as ceilings on deposit rates.

A similar positive relationship between competition and fragility emerges also from the works by Rochet and Vives (2004) and Goldstein and Pauzner (2005), where higher deposit rates lead to more coordination failures and bank runs.[2]

Allen and Gale (2004) analyze the link between competition and stability in the interbank market. Banks do not have incentives to provide liquidity to a troubled bank when the interbank market is perfectly competitive, because each of them acts as price taker and assumes that its action does not affect the equilibrium. In contrast, when the interbank

[2]The question of whether competition increases bank fragility can be posed in terms of whether cooperatives distributing profits to their members are less or more fragile than profit-maximizing institutions. Rey and Tirole (2000) suggest that cooperatives are more fragile than institutions making positive profits, because the the lack of a buffer and the sharing of fixed costs among members create a network externality, which exposes cooperative to runs. Differently, institutions making positive profits are less prone to runs as they can use their buffer to bear the risk of members' exit and avoid the consequent negative externality on the remaining members.

market is imperfectly competitive, banks may want to help the troubled bank in order to prevent contagion. This underlines a sort of cooperative equilibrium. Since the provision of liquidity resembles the provision of a public good, when the number of banks is not too large each bank may have an incentive to coordinate with the others and provide liquidity, as long as avoiding contagion makes everybody better off. However, a coordination failure may still arise in that each bank may find optimal not to provide any liquidity, if it thinks the others will not contribute anything.[3]

Importantly, this coordination failure may occur independently of the degree of competition in the market, and thus occur in any market structure. This is formally shown by Matutes and Vives (1996) in a model á la Diamond (1984) enriched with elements of product differentiation, network externalities and possibility of bank failures. Consider two banks, $i = a, b$, each located at the opposite extremes of a unit segment [0,1], and competing for depositors with a reservation value of v and linear transportation costs $c \geq 0$. Banks offer depositors a standard debt contract with a (gross) return r_i^D and nonpecuniary bankruptcy penalties, and obtain a deposit market share of d_i. Then, they invest in risky projects, which require a minimum initial investment I and yield a (random) return $\widetilde{R}_i$. The distribution function of $\widetilde{R}_i$ depends on the market share d_i each bank obtains in the deposit market. In particular, the larger a bank the more it can diversify away risk and decrease the probability of going bankrupt. Depositors do not observe the returns of banks' investments, but are endowed with homogenous prior beliefs (p_a, p_b) about their success probabilities. Given these beliefs, banks set r_i^D and depositors choose which bank to patronize. The market share and the aggressiveness of each bank i depend crucially on depositors' perceptions of its success probability, and in particular on whether the difference $p_i r_i - p_j r_j$ is in the interval $[-c, c]$ and $p_i r_i - c d_i$ is above v. Both banks are active and have market share $d_i = \frac{1}{2} + \frac{(p_i r_i - p_j r_j)}{2c}$ if both $p_i r_i - p_j r_j \in [-c, c]$ and $p_i r_i - c d_i \geq v$; banks enjoy local monopolies if $p_i r_i - c d_i < v$; and only the bank with higher expected return is active if $p_i r_i - p_j r_j$ is not in the interval $[-c, c]$. Depositors' perceptions of p_a and p_b differentiate the model from the standard Hotelling game, and introduce vertical product differentiation in banking competition. In

[3] See also Sáez and Shi (2004).

equilibrium, the success probabilities p_a and p_b are endogenously determined by depositors' expectations, which are self-fulfilling given the use of standard deposit contracts and the presence of economies of scale. A bank perceived to be safer commands a higher margin and a larger market share, which in turn makes it safer because of better diversification. The self-fulfilling character of depositors' expectations implies multiple equilibria. Possible equilibria include corner solutions, where only one bank is active, and even equilibria where no banks are active. These are due to a coordination problem among depositors, which arises for reasons similar to those encountered in the network literature. A bank is a large network which requires a minimum size to be viable and in which each customer benefits from a larger number of members. In this view, the nonbanking equilibrium is reminiscent of the bad equilibrium in Diamond and Dybvig (1983) and can be interpreted as a 'systemic confidence crisis'. The coordination problem among depositors occurs irrespectively of the degree of competition in the deposit market. A monopoly bank can suffer from fragility in the same way as a competitive bank. Deposit insurance can prevent depositors' coordination failure and thus bank failures. Since depositors are repaid with certainty, they view $p_i = 1$, $i = A, B$, and the model converges to a Hotelling game with no vertical product differentiation and no more multiple equilibria. As in Diamond and Dybvig (1983), deposit insurance eliminates the nonbanking equilibrium and stabilizes the system, but it is not always welfare-enhancing. On the one hand, by ensuring that all banks remain active, deposit insurance may preclude the realization of desirable diversification economies. Also, deposit insurance induces fiercer competition for deposits, which in turn increases the deadweight loss in case of failure and decreases the success probability of banks. On the other hand, however, deposit insurance has the positive effect of extending the market. This leads potentially to transform a market where banks have local monopoly power to one where banks compete. Thus, the net welfare effects of deposit insurance are ambiguous, and cannot be assessed independently of the market structure even in the absence of moral hazard considerations.

An ambiguous relationship between competition and stability is also found by Carletti et al. (2003) in a model analyzing the effects of bank mergers on loan competition, reserve management and banking system liquidity. Banks compete for loans and engage in interbank

lending in order to deal with stochastic liquidity shocks á la Diamond and Dybvig (1983). A merger creates an internal money market where the merged banks can reshuffle liquidity. This affects their reserve management, pushing towards higher reserves when borrowing on the interbank market is not so costly relative to raising deposits and keeping more reserves initially, and towards lower reserves otherwise. The merger modifies also loan rates and banks' market shares. The overall effect on loan rates depend on how strong is the increase in market power relative to potential efficiency gains. More importantly, the change in banks' size affects aggregate liquidity. Greater heterogeneity among banks increases the variance of the aggregate liquidity demand and leads, ceteris paribus, to higher aggregate liquidity needs. This asymmetry channel, together with the change in reserve holdings, determines the effect of bank consolidation on the aggregate liquidity. Thus, the model suggests that imperfect loan market competition increases the volatility of the aggregate demand in the interbank market and may negatively affect the working of the interbank market. This would not occur if banks were perfectly competitive.

The relationship between crises and market structure is also analyzed by Boyd et al. (2004) in a monetary, general equilibrium economy in which banks provide inter-temporal insurance to risk-averse depositors and a monetary authority controls the rate of inflation. The model generates two different types of crises. In the first one ("banking crisis") banks exhaust their reserve assets but do not liquidate the long-term asset. In the second one ("costly banking crisis") banks liquidate the long-term asset, thus creating a real cost for the economy. The level of inflation rate determines the relationship between crises and market structure of the banking system. A monopolistic banking system faces a higher probability of banking crises when the inflation rate is below some threshold, while a competitive system is more fragile otherwise. This result is driven by a trade-off implicit in banks' behavior. A monopolistic bank limits reserve holdings and offers lower deposit rates relative to a competitive bank. When the inflation rate is low, the first effect dominates, thus increasing the probability of banking crises in monopolistic banking systems. Concerning resource losses, costly banking crises are always more likely to occur under competition than under monopoly because the latter provides poorer inter-temporal insurance to depositors. Thus,

despite being more stable, a monopolistic banking system is not necessary welfare-enhancing.

4.2 Market structure and risk taking

Most of the literature on the relationship between competition and stability analyzes the impact that competition has on banks' incentives to take risk. This focus originates from some empirical studies finding a negative effect of higher charter values on risk taking. For example, Keeley (1990) finds that the surge of bank failures in the US during the 1980s derived mostly from various deregulation measures and market factors that reduced banks' monopoly rents and increased the value of their put option on the deposit insurance fund. Similarly, Edwards and Mishkin (1995) argue that the excessive risk-taking observed in the 1980s in the US was banks' obvious response to the erosion of profits due to competition from financial markets. This decreased banks' cost advantages in the acquisition of funds and undermined their position in the loan market.

Following these empirical findings, the theoretical literature has initially stressed how competition worsens banks' incentives to take risk (e.g., Besanko and Thakor, 1993, Boot and Greenbaum, 1993, Allen and Gale, 2000b and 2004) and how regulation can help in mitigating this perverse link (e.g., Hellmann et al, 2000, Perotti and Suarez, 2002, Repullo, 2004). Despite this still being the prevailing view, more recent studies have suggested that the detrimental relationship between competition and bank risk taking is not robust. In particular, a higher degree of competition may induce banks to become more prudent once particular aspects of bank-firm relationships (e.g., entrepreneurs' effort) or important banks' functions (e.g., monitoring) are taken into account. In the following we describe the most important contributions on the link between competition and risk taking, postponing to the next section the discussion on the importance and role of bank regulation in mitigating the negative effects that competition may have on stability.

The perverse link between competition and bank risk taking has been shown in several different frameworks. The general idea is that greater competition reduces banks' charter values (or rents available to shareholders and/or managers). This increases the attractiveness

of the gains from taking risks, and therefore the incentives to exploit the non-convexity in banks' payoff functions. Besanko and Thakor (1993) uses this idea in a framework of relationship banking, where banks acquire private information over their borrowers. This gives banks an informational monopoly and generates informational rents. As long as banks appropriate at least part of these rents, they have an incentive to limit their risk exposure and enjoy the value of the relationship. However, as soon as the banking industry becomes more competitive, relationship banking decreases in value and banks take more risk, particularly when deposits are backed by a risk insensitive insurance scheme. Boot and Greenbaum (1993) obtain similar results in a two-period model in which banks can acquire funding-related reputational benefits and improve their rents through costly monitoring.

To see how competition may exacerbate bank risk shifting problem, we consider the simple model by Allen and Gale (2000b and 2004). Consider n banks choosing a portfolio consisting of perfectly correlated risks and competing á la Cournot on the deposit market. Each bank i receives a per unit return $R_i \in [0, \overline{R}]$ with probability $p(R_i)$ and 0 with probability $(1-p(R_i))$, with $p(R_i)$ satisfying $p(0) = 1, p(\overline{R}) = 0$, $p'(R_i) < 0$ and $p''(R_i) < 0$. Each bank raises an amount d_i of deposits and faces an upward sloping supply of funds. Given a total demand for deposits equal to $D = \sum_i d_i$, the opportunity costs of funds is $r_D(D)$, which satisfies $r'_D(D) > 0$, $r''_D(D) > 0$, $r_D(0) = 0$ and $r_D(\infty) = \infty$. Depositors are insured and the supply of funds is independent of banks' portfolio risk. The payoff to bank i is then given by

$$\Pi_i(R, d) = p(R_i)[R_i - r_D(D)]d_i,$$

where $R = (R_1, ..., R_n)$ and $d = (d_1, ..., d_n)$. A Nash-Cournot equilibrium where each bank i chooses a strictly positive pair (R_i, d_i) has then to satisfy

$$p(R_i)[R_i - r_D(D) - r'_D(D)d_i] = 0$$

$$p'(R_i)[R_i - r_D(D)]d_i + p(R_i)d_i = 0.$$

In a symmetric equilibrium these conditions reduce to

$$R - r_D(nd) - r'_D(nd)d = 0$$

$$p'(R)[R - r(nd)] + p(R) = 0,$$

which imply

$$-\frac{p(R)}{p'(R)} = R - r_D(nd) = r'_D(nd)d.$$

This condition characterizes a symmetric equilibrium where each bank chooses the riskiness and size of the portfolio equal to (R^*, d^*). The equilibrium depends on the number of banks n. As $n \to \infty$, $d = \frac{D}{n} \to 0$, since D must be bounded above when n increases if $r_D(\infty) \to \infty$. This implies $r_D(nd)d \to 0$, and thus $R - r_D(nd) \to 0$ and $p(R) = 0$. An increase in competition has then a negative effect on bank riskiness. As $n \to \infty$, banks choose the maximum level of risk, i.e., $R \to \overline{R}$. The reason is that banks become smaller and behave more like perfect competitors, thus increasing their size as long as profits are positive. In equilibrium, they make zero profits and have extreme incentives for risk taking.

The result of a positive perverse relationship between competition and risk taking problem extends to richer frameworks. Banks have an incentive to engage in risk shifting as long as their objective function is convex. The problem is particularly acute when banks are close to bankrupt, and it is worsened by competition. The property of a convex objective function holds for example in the presence of increasing returns to scale, and in dynamic stationary environments in which banks compete for market shares and play the short-run strategy at each date (Allen and Gale, 2000b and 2004).

One crucial element is how banks operate on the asset side. A common assumption of the models showing a positive relationship between competition and risk taking is that banks have complete control over the risk of their portfolios. Each bank invests in assets with given risk characteristics and determines the riskiness of its portfolio. As competition for deposits becomes tougher, profits decline and banks' preference for risk increases. The assumption of banks determining exclusively their portfolio risk covers however an important risk-incentive mechanism on the asset side because it ignores the bank-borrower relationship. Once this is taken into account, the result can change dramatically. When both banks and entrepreneurs can influence the risk of investment projects, the relationship between competition and risk

taking becomes ambiguous. This is formally shown by Caminal and Matutes (2002) in a model in which banks compete for loans and can use monitoring or credit rationing to deal with an entrepreneurial moral hazard problem. Given limited liability and non-verifiable actions, entrepreneurs have distorted incentives to allocate funds among alternative projects, which have different levels of risk and are subjective to multiplicative aggregate shocks. Monitoring and credit rationing help in reducing entrepreneurial moral hazard, but are imperfect substitute tools for the bank. The former requires the use of costly resources, while the latter reduces the potential gain from trade. If the bank does not monitor, credit has to be restricted in order to increase the marginal return of the funds invested and induce entrepreneurs to choose appropriate projects. The choice between monitoring and credit rationing depends on banks' profits and hence on the degree of competition. A monopoly bank uses more monitoring and less credit rationing. This may induce a monopoly bank to grant larger loans and thus have a higher failure probability than a competitive bank, since projects are subject to multiplicative shocks. As a consequence, the relationship between market power and failure probability is ambiguous.[4]

The discussion above also suggests that when entrepreneurs choose exclusively the risk of the investment projects, banks may become riskier as competition decreases. Greater competition in the loan market reduces the loan rates that entrepreneurs pay, thus increasing their profits and reducing their incentives to take risks. To show this mechanism, Boyd and De Nicoló extend the model of Allen and Gale (2000b and 2004) described above. In particular, they introduce many entrepreneurs who borrow from banks at a rate r_L and invest in investment projects yielding a per unit return $R \in [0, \overline{R}]$ with probability $p(R)$ and 0 with probability $(1 - p(R))$. Given a total amount of deposits and thus of loans equal to D in the economy, the loan rate is $r_L(D)$ and satisfies $r_L(0) > 0$, $r'_L < 0$, $r''_L \leq 0$, and $r_L(0) > r_D(0)$. The return R is still a measure of project risk, but it is now chosen by the

[4]Koskela and Stenbacka (2000) find an unambiguous positive relationship between competition and stability but their framework is somewhat different. Banks compete in the loan market but, absent any moral hazard problem, stability refers to entrepreneurs' bankruptcy risk. Under the assumption of a mean-shifting investment technology, a monopoly bank charges higher lending rates than competitive banks, which leads to lower investments and thus to a higher probability of bankruptcy.

entrepreneurs to maximize

$$p(R)[R - r_L].$$

Thus, R satisfies

$$r_L = R + \frac{p(R)}{p'(R)}, \tag{6}$$

and it increases with r_L. Each bank chooses the amount of deposits d_i to maximize

$$p(R)[r_L(D) - r_D(D)]d_i$$

where r_L satisfies (6). The equilibrium depends again on the number of banks n. Differently from Allen and Gale though, now the level of R is strictly decreasing in n and prices converge to the competitive outcome, i.e., $r_L(D) - r_D(D) = 0$ as $n \to \infty$. This occurs because banks take into account the risk incentive mechanism of the entrepreneurs when setting loan rates, thus anticipating that the risk of their portfolio increases with loan rates.

5 Competition and regulation

As discussed above, market power is often thought to be associated with a lower probability of bank failure both in static and dynamic contexts. High margins act as buffers against expected losses; high future expected profits increase the opportunity cost of going bankrupt, thus reducing banks' incentives to take excessive risk. The implication of this (still prevailing) view is that the banking system needs to be regulated to limit the adverse consequences of intense competition and achieve stability. But how to design regulation appropriately? One possibility is to limit competition directly. Ceilings on interest rates or limited entry are examples of how to reduce competition and induce banks to behave more prudently. Another possibility is to design regulation in a way to "correct" the negative effects of competition. For example, risk-adjusted deposit insurance premia or appropriate capital requirements may be effective ways to control risk taking even in the presence of intense competition. What is important is that the design of regulation has to take account of the effects that different market structures have on banks' incentives to take risk. As argued by Boyd and Gertler (1993), the poor performance of the US banking system in the 1980s resulted from enhanced competition and an inadequate regulatory policy that encouraged excessive risk

taking. The main source of problems was in fact the great risk taken by large banks, which faced more intense competition while being implicitly insured through the "too-big-to-fail" policy. Similarly, Edwards and Mishkin (1993) argue that the decline of bank profitability induced by enhanced competition entails a risk to the financial system only if regulators fail to adapt their policies to the changing financial environment.

A growing literature analyzes how regulation affects the relationship between competition and stability, in particular risk taking. Starting from the standard paradigm that competition leads to higher risk, most contributions focus on the effectiveness of regulation in reducing the negative consequences of competition. Results are ambivalent and sensitive to the specific framework of analysis. Regulation such as risk-adjusted deposit insurance or capital requirement is sufficient to eliminate the negative impact of competition on risk taking in some cases; but specific restrictions on competition such as interest rate ceilings or entry restrictions are needed in other cases.

Following the mainstream, we will describe how regulation can remove the perverse effects of competition on risk taking. It is worth pointing out though that, besides removing such effects, regulation may also directly influence the "sign" of the relationship between competition and stability. For example, the effect of high charter values on banks' incentives may depend crucially on how they are generated. Nagarajan and Sealey (1995) show that high margins may not be effective in improving banks' incentives when they result from a forbearance policy extending the expiration date of equity holders' call option. More precisely, high charter values induce banks to choose high asset quality only if they are generated by an optimal forbearance policy, which takes account of both the performance of individual banks and the overall market.

Matutes and Vives (2000) examine the impact of deposit insurance on bank competition and risk taking incentives in a context where banks are subject to limited liability and their failure implies social costs. In line with the charter value literature, banks choose the risk level of their portfolios $\gamma_i \in [\underline{\gamma}, \overline{\gamma}]$, and have a per unit return $\widetilde{R}_i \in [\underline{R}, \overline{R}]$ with density function $g_i = (R_i, \gamma_i)$ and distribution function $G_i(\cdot)$. The choice of γ_i is not observable,[5]

[5]Matutes and Vives (2000) also analyze the case with disclosure requirements when γ_i is observable. In

and higher levels of risk (higher γ_i) are associated with mean preserving spreads over G_i so that $E(R_i)$ is the same for all G_i but the variance increases with γ_i. Banks raise funds from investors in the form of standard deposit contracts promising a (gross) return equal to r_i^D. This implies that depositors are repaid fully only when the bank does not go bankrupt, and they get whatever is left otherwise. The supply of deposits is elastic and equal to

$$d_i = a + b\Phi_i^e(r_i^D) - c\Phi_j^e(r_j^D),$$

where $\Phi_i^e(r_i) = \Phi_i^e(r_i^D, \gamma_i^e)$ is depositors' assessment of the expected return of one unit invested in bank i, and γ_i^e is their assessment of bank risk. Given depositors' priors $\Phi_i^e(r_i)$, banks set r_i^D and depositors choose how much to supply. Banks retain some market power and have positive profits only if $R_i > r_i^D$. Otherwise, they fail and impose a social cost K on the economy. Then banks choose γ_i to maximize their expected profits

$$\Pi_i = d_i \int_{r_i^D}^{\overline{R}} (R_i - r_i^D) g_i(R_i, \gamma_i) dR_i.$$

The non-observability of γ_i together with the limited liability imply that banks always choose the maximum level of risk, i.e., $\gamma_i = \overline{\gamma}$. This is the only credible level consistent with depositors' priors γ_i^e as banks' expected profits are increasing in γ_i. As a consequence, banks behave aggressively on the deposit market to increase their deposit base. The equilibrium is inefficient as it involves a high risk of bank failure and high social failure costs. The inefficiency may be ameliorated with the introduction of a deposit insurance scheme, but its effectiveness crucially depends on how the scheme is designed. When deposits are insured through a flat rate scheme, banks have still the incentives to take the maximum level of risk, and both deposit rate regulation and asset restrictions are necessary to improve welfare. In contrast, when deposit insurance premia are risk-based, deposit insurance may be sufficient to improve welfare. If the regulator observes γ_i, banks pay a premium contingent on their asset risk and deposit rates equal to

$$\tau_i(r_i^D, \gamma_i) = 1 - \frac{(E(R_i) - r_i^D)}{\int_{r_i^D}^{\overline{R}} (R_i - r_i^D) g_i(R_i, \gamma_i) dR_i},$$

this case regulation is a sufficient instrument to increase welfare.

and have expected profits equal to

$$\begin{aligned}\Pi_i &= (1-\tau_i)d_i \int_{r_i^D}^{\overline{R}} (R_i - r_i^D) g_i(R_i, \gamma_i) dR_i \\ &= (E(R_i) - r_i^D) d_i.\end{aligned}$$

Banks' expected profits are independent of the level of risk because the positive effect of higher risk on expected margins is offset by an increase in the premium banks have to pay. Thus, banks behave less aggressively on the deposit market as expanding their deposit base is no longer profitable. Concerning the choice of risk, any level of risk is consistent with the equilibrium when the deposit insurance premia are set simultaneously to the choice of γ_i, whereas the maximal risk is still chosen when the premia are set before γ_i is determined. In this case, deposit rate ceiling and asset restrictions are again necessary to improve welfare.

Similar results are obtained by Cordella and Yeyati (2002) in a framework which extends Matutes and Vives (2000) to explicitly endogenize competition for deposits. As before, banks have limited liability and choose privately the level of risk of their portfolios. Each bank chooses a monitoring effort m_i, which determines the success probability of its portfolio at a cost m_i^2. Banks compete à la Salop (1979) on the deposit market and incur a fixed entry cost F. The deposit supply function is

$$d_i(r_i^D, r^D, m_i, m_i^e, n) = \frac{1}{n} + \frac{[a + (1-a)m_i^e]r_i^D - [a + (1-a)m^e]r^D}{c},$$

where n is the number of banks, $a \in [0,1]$ is the fraction of insured deposits, r_i^D and r^D are the deposit rates set by bank i and by all other banks respectively, m_i^e and m^e are depositors' assessments of the success probability of bank i and of all other banks, and c is depositors' transportation cost. Banks' expected profits are $\Pi_i - F$, where

$$\Pi_i = d_i[m_i(R_i - r_i^D) - m_i^2 - m_i \tau r_i]$$

with $\tau = \frac{a(1-m^e)}{m^e}$ representing the premium that each bank pays on the liabilities $d_i r_i$. Since both deposit rates and insurance premia are set as functions of expected rather than actual risk and deposits, banks have incentives to choose lower monitoring efforts. The equilibrium of the benchmark case can be improved by disclosing the level of monitoring m_i either to

depositors (scenario D) or to a deposit insurance agency which can then charge risk-based premia (scenario R). In both cases, banks choose higher monitoring efforts than under the benchmark scenario. The economy converges to the same equilibrium in terms of risk and expected returns on deposits in the limiting cases when there is no insurance ($a = 0$) in scenario D and full insurance ($a = 0$) in scenario R. The reason is that the disciplining effect in each scenario depends on the fraction of deposits which are priced correctly, namely $(1-a)$ in scenario D and a in scenario R. Thus, the two systems converge at the opposite extremes of the deposit insurance coverage and welfare is always higher than in the benchmark case.

An alternative way to restore prudent behavior is to introduce capital requirements. Hellman et al. (2000) analyze the relationship between competition for deposits, risk taking and capital regulation in a dynamic framework where banks choose privately their asset risk and compete for deposits. Banks operate for T periods, and can invest either in a prudent or in a risky asset each period. The former yields a safe return S, while the latter yields $\overline{R}$ with probability p and $\underline{R}$ with probability $(1-p)$. The risky asset has a higher return in case of success ($\overline{R} > S$) but a lower one in expectation ($S > p\overline{R} + (1-p)\underline{R}$). Each bank i competes for insured deposits by offering a deposit rate r_i^D and raises an amount $d_i(r_i^D, r_{-i}^D)$, where r_{-i}^D is the rate offered by the other banks. Each bank raises also an amount of capital k even if costly ($\rho > S$), and it invests a total amount of $(1+k)d_i(r_i^D, r_{-i}^D)$. Prudential regulation requires that at the end of each period a regulator inspects the amount of capital each bank has and closes it down if such an amount is negative. The per-period profit of the bank is

$$\Pi_P(r_i^D, r_{-i}^D, k) = (S(1+k) - r_i^D - \rho k)d_i(r_i^D, r_{-i}^D),$$

when it invests in the prudent asset, and

$$\Pi_R(r_i^D, r_{-i}^D, k) = [p(\overline{R}(1+k) - r_i^D) - \rho k]d_i(r_i^D, r_{-i}^D),$$

when it gambles and invests in the risky asset. In this case, the bank has a positive return if the asset succeeds, while it is closed down if it fails. After raising capital and attracting depositors, banks choose the asset portfolio to maximize their expected discounted profits $V = \sum_{t=0}^{T} \delta^t \Pi_t$. As $T \to \infty$, banks play an infinitely repeated static Nash equilibrium so

that the game has a static structure within each time period. Banks choose to gamble and invest in the risky asset whenever

$$V_R(r_i^D, r_{-i}^D, k) = \frac{\Pi_R(r_i^D, r_{-i}^D, k)}{(1-\delta p)} > V_P(r_i^D, r_{-i}^D, k) = \frac{\Pi_P(r_i^D, r_{-i}^D, k)}{(1-\delta)},$$

that is whenever

$$\Pi_R(r_i^D, r_{-i}^D, k) - \Pi_P(r_i^D, r_{-i}^D, k) V_R(r_i^D, r_{-i}^D, k) > (1-p)\delta V_P(r_i^D, r_{-i}^D, k). \tag{7}$$

Condition (7) defines a critical level of the deposit rate, $\widehat{r}^D$, such that banks gamble when $r^D(k) > \widehat{r}^D(k)$, and behave prudently otherwise. This implies that banks choose to gamble for sufficiently competitive deposit markets (i.e., when the supply of deposit is sufficiently elastic), since then their per-period gains from this strategy ($\Pi_R - \Pi_P$) exceed the franchise value (δV_p) that banks lose when the risky asset fails (with probability $(1-p)$). The critical level of deposit rates, $\widehat{r}^D(k)$, increases with the level of capital k; but in an unregulated equilibrium banks prefer not to raise any capital since it is costly and decreases their franchise value (franchise-value effect). A possible way to restore prudent bank behavior is to introduce capital requirements. If banks hold sufficient capital, they internalize the negative consequences of gambling and choose to behave prudently (capital-at-risk effect). Capital regulation is however a Pareto-inefficient policy in a dynamic framework. When all competitors set a deposit rate consistent with prudent behavior, for a given amount of deposits a bank is indifferent between the prudent and the risky asset. The bank must then earn a higher expected margin from the risky asset than from the gambling asset, since it loses the franchise value if the risky asset fails. This implies that the bank has an incentive to offer a slightly higher deposit rate than its competitors so to "steal" depositors, and invests in the risky asset (market-stealing effect). Because each bank has an incentive to do so, capital requirements coupled with freely determined deposit rates do not achieve Pareto-efficient equilibria. Capital requirements become effective only when they raise banks' costs to the level that banks are no longer willing to pay out higher deposit rates. But then other forms of regulation, such a deposit-rate controls, may achieve Pareto efficiency. By preventing the market stealing effects, deposit-rate controls increase banks' per-period profits and franchise values, and induce prudent behavior.

A different result on the effectiveness of capital regulation is obtained by Repullo (2004) in a similar dynamic model where, as before, banks can invest in either a prudent or risky asset, but compete à la Salop (1979) on the deposit market. Depositors are insured, and face a unit travelling cost of c. Thus, each of the n banks raises an amount $\frac{1}{n}$ of deposits every period and raises an amount of capital k at a cost of c. As in Hellmann et al. (2000), the equilibrium reached in an unregulated economy depends on the level of intermediation margins (equal to $\frac{c}{n}$) and thus on the degree of deposit competition. All banks choose the risky asset for low margins (i.e., when competition is intense), the prudent assets for high margins (i.e., when competition is not intense) and both types of assets in the intermediate cases. When banks choose to gamble in an unregulated economy, capital requirements are an efficient regulatory measure as they reduce deposit rates without affecting banks' franchise value. Hence, only the capital-at-risk effect is at work, and capital regulation is effective in ensuring the existence of the prudent equilibrium. The different results relative to Hellmann et al. (2000) depend on the more explicit analysis of deposit market competition and on the use of internal capital instead of outside capital. This particular assumption modifies the way in which capital regulation enters in banks' profit functions and affects franchise values. Given constant intermediation margins, the franchise values are equal to

$$V_P = \frac{c}{\rho n^2}$$

when banks behave prudently, and to

$$V_R = \frac{pc}{[\rho + (1-p)]n^2}$$

when they gamble. In both cases the franchise values do not depend on the capital requirement k because the negative effect of a higher level of capital is passed onto depositors in the form of lower deposit rates. Hence, capital requirements are effective in implementing prudent behavior, although they make depositors worse off. Deposit-rate ceilings do not do any better. Only risk-based capital requirements improve welfare, since they implement prudent behavior without reducing deposit rates.

Risk-based capital requirements are also effective in reducing banks' portfolio risk in Bolt and Tieman (2004) in a dynamic duopoly, where banks compete for borrowers by setting

acceptance criteria. In particular, banks offer differentiated loans and face a linear demand equal to

$$L_i(\alpha_i, \alpha_j) = L + a\alpha_i - b\alpha_j$$

where higher α_i and α_j represent lower acceptance criteria, and the parameter b is a measure of the degree of substitution between loans of banks i and j (with $i \neq j$). Easing the acceptance criteria increases the demand for loans and thus banks' per-period profits, but it also worsens the quality of their portfolios since riskier borrowers obtain financing. Thus, banks face a trade-off between increasing market shares in the short run and securing continuation in the long run. As standard in the charter value literature, competition (here intended as lower b) increases the attractiveness of larger market shares, thus inducing banks to ease acceptance criteria and increase risk. Prudential regulation can help in removing the negative effects of competition. Capital requirements (in particular if risk-based) lead to less risk taking because they improve banks' incentives to set tight acceptance criteria and reduce their failure probabilities.

An alternative regulatory instrument to create charter values and solve the trade-off between competition and stability is analyzed by Perotti and Suarez (2002) in a dynamic model where two banks compete on the deposit market and invest in either a prudent or a speculative asset. As in Hellman et al. (2000) and Repullo (2004), the choice of lending must trade-off the short-term gains from risk-taking against the risk of losing charter value. Intense competition increases risk-taking by enlarging short-term gains and reducing future charters. Differently from previous works, however, the degree of competition is endogenous, and is driven by banks' failures and regulatory policies on mergers and entry. When a bank fails as a result of an unsuccessful speculative lending, the regulator has to decide whether to merge it with the incumbent surviving bank (merger policy) and/or whether to allow entry (entry policy). A merger with the incumbent bank modifies the market structure to a monopoly until the entry of a new bank brings it back to a duopoly. The possibility of obtaining monopoly rents (albeit temporarily) gives banks an additional incentive beside the increase of charter value to behave prudently and remain solvent. Said it differently, banks' speculative

lending decisions become strategic substitutes in that the incentive of a bank to take risk decreases with the risk position of the competing bank. As a consequence, merger and entry policies imply a trade-off between competition and stability. Allowing a merger when a bank is insolvent involves prudent behavior but also monopoly inefficiencies. The optimal policy instrument is a combination of mergers following a failure and subsequent entry. This creates ex ante incentives for banks to remain solvent to acquire failing institutions while limiting the ex post market power that surviving banks get through the rescue.

6 Concluding remarks

This paper reviews the main literature on stability and competition in the banking industry. Each of these two issues has received a large amount of attention in the last decades but the two strands of literature remain somewhat disconnected. The stability literature proceeds typically under the assumption that banks operate in a perfect competitive system, thus disregarding the implications of different banking structures for the safety of the sector. In contrast, the competition literature analyzes the operation of the competitive mechanism in the presence of market failures disregarding the effects on depositors' and agents' behavior. Thus, whether greater competition enhances or worsens the stability of the system remains unclear.

Only very recently has the literature addressed more directly how competition affects stability. The general argument is that competition worsens stability. Higher deposit rates increase the probability of bank runs; lower margins worsen the problem of excessive risk taking. However, this view has been challenged by recent contributions that consider imperfect competition and endogenize important aspects of industrial organization. For example, it has been shown that coordination problems among depositors can emerge independently of competition; and that banks operating in monopolistic settings may face higher failure probabilities than those operating in competitive industries. Furthermore, the (few) contributions addressing the optimal regulation in models of imperfect competition suggest that, even if competition hurts stability, its negative effects can be ameliorated by designing financial regulation appropriately.

Despite the growing attention to the issue of competition and stability, additional research seems warranted in several directions. First, the link between market structure and bank fragility is worth further study. Models of runs and panics should be extended to situations of imperfect competition. Second, the effects of imperfect competition on bank risk taking should be examined in richer frameworks, which consider competition on both loan and deposit markets. Third, on the normative side, more research is needed for a better understanding on the effectiveness of regulation.

References

Allen, F. and D. Gale (1998), Optimal financial crises, *Journal of Finance*, 53, 1245-1284

Allen, F. and D. Gale (2000a), Financial contagion, *Journal of Political Economy*, 108(1), 1-29

Allen, F. and D. Gale (2000b), Comparing financial systems, Cambridge, MA: MIT Press

Allen, F. and D. Gale (2004), Competition and stability, *Journal of Money, Credit and Banking*, 36(3), 453-480

Allen, F., H. Gersbach, J.P. Krahnen and A. M. Santomero (2001), Competition among banks: Introduction and conference overview, *European Finance Review*, 5, 1-11

Bagehot, G. (1873), London Street: A description of the money market, *H.S. King*, London

Besanko, D. and A.V. Thakor (1993), Relationship banking, deposit insurance and bank portfolio, in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge, UK: Cambridge University Press, 292-318

Bolt, W. and A.F. Tieman (2004), Banking competition, risk, and regulation, *Scandinavian Journal of Economics*, 106(4), 783-804

Boot, A.W. and S. Greenbaum (1993), Bank regulation, reputation and rents: theory and policy implications, in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge, UK: Cambridge University Press, 262-285

Bouckaert, J. and H. Degryse (2004), Softening competition by inducing switching in credit markets, *Journal of Industrial Economics*, 52, 27-52

Boyd, J.H., G. De Nicoló and B.D. Smith (2004), Crises in competitive versus monopolistic banking systems, *Journal of Money, Credit and Banking*, 36(3), 487-506

Boyd, J.H. and G. De Nicoló (2005), The theory of bank risk taking and competition revisited, forthcoming, *Journal of Finance*

Boyd, J.H. and M. Gertler (1993), U.S. commercial banking: Trends, cycles and policy, NBER Macro Annual

Broecker, T. (1990), Creditworthiness tests and interbank competition, *Econometrica*, 58, 429-452

Caminal, R. and C. Matutes (2002), Market power and banking failures, *International Journal of Industrial Organization*, 20(9), 1341-1361

Canoy, M., M. van Dijk, J. Lemmen, R. de Mooij and J. Weigand (2001), 'Competition and stability in banking', CBP Document, no. 015, Netherlands Bureau for Economic Policy Analysis, December

Carletti, E. (1998), Competition, stability and regulation, traduzione italiana in M. Polo (ed), *Industria Bancaria e Concorrenza*, Bologna: Il Mulino, 67-136

Carletti, E. (2004), The structure of bank relationships, endogenous monitoring, and loan rates, *Journal of Financial Intermediation*, 13, 58-86

Carletti, E. and P. Hartmann (2002), Competition and stability: what's special about banking?", in P. Mizen (ed.), *Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart*, vol. 2, Cheltenham, UK: Edward Elgar, 202-229

Carletti, E., P. Hartmann and G. Spagnolo (2003), Bank mergers, competition and liquidity, ECB working paper 292, Frankfurt, Germany

Cerasi, V. and S. Daltung (2000), The optimal size of a bank: Costs and benefits of diversification, *European Economic Review*, 44(9), 1701-1726

Chari, V.V. and R. Jagannathan (1988), Banking panics, information and rational expectations equilibrium, *Journal of Finance*, 43, 749-763

Cordella, T. and L. Yeyati (2002), Financial opening, deposit insurance and risk in a model of banking competition, *European Economic Review*, 46, 471-485

De Bandt, O. and P. Hartmann (2002), Systemic risk in banking: a survey, in C.A.E. Goodhart and G. Illing (eds), *Financial crises, contagion and the lender of last resort – A reader*, London, Oxford University Press, 249-298

Degryse, H. (1996), On the interaction between vertical and horizontal product differentiation: An application to banking, *Journal of Industrial Economics*, 44(2), 169-182

Dell'Ariccia, G. (2001), Asymmetric information and the structure of the banking industry, *European Economic Review*, 45, 1957-1980

Dell'Ariccia, G., E. Friedman and R. Marquez (1999), Adverse selection as barrier to entry in the banking industry, *RAND Journal of Economics*, 30, 515-534

Diamond, D.W. (1984), Financial intermediation and delegated monitoring, *Review of Economic Studies*, 51, 393-414

Diamond, D.W. and P.H. Dybvig (1983), Bank runs, deposit insurance and liquidity, *Journal of Political Economy*, 91, 401-419

Diamond, P. (1971), A model of price adjustment, *Journal of Economic Theory*, 3, 156-168

Edwards, F. and F. Mishkin (1995), The decline of traditional banking: implications for financial stability and regulatory policy, *Federal Reserve Bank of New York Economic Policy Review*, 1, 27-45.

Freixas, X. and B. Parigi (1998), Contagion and efficiency in gross and net interbank payment systems, *Journal of Financial Intermediation*, 7, 3-31

Freixas, X., B. Parigi and J.C. Rochet (2000), Systemic risk, interbank relations and liquidity provision by the central bank, *Journal of Money, Credit and Banking*, 32(3), 611-638

Freixas, X., B. Parigi and J.C. Rochet (2004). The Lender of Last Resort: A 21st century approach, *Journal of the European Economic Association*, 2(6), 1085-1115

Gale, D. and X. Vives (2002), Dollarization, bailouts and the stability of the banking system, *Quarterly Journal of Economics*, 117(2), 467-502

Gehrig, T. (1996), Natural oligopoly and customer networks in intermediated markets, *International Journal of Industrial Organization*, 14(1), 101-118

Gehrig, T. (1998), Screening, cross-border banking and the allocation of credit, *Research in Economics*, 52(4), 387-407

Goldstein, I. and A. Pauzner (2005), Demand deposit contracts and the probability of bank runs, *Journal of Finance*, 60, 1293-1328

González-Maestre, M. and L.M. Granero (2003), Industrial loans and market structure, *European Economic Review*, 47, 841-855

Goodfriend, M. and R. King (1988), Financial deregulation, monetary policy and central banking, *Federal Reserve Bank of Richmond Economic Review*, 74, 3-22

Goodhart, C.A.E. (1987), Why do banks need a central bank?, *Oxford Economic Papers*, 39, 75-89

Gorton, G. and A. Winton (2003), Financial intermediation, in G.M. Constantinides, M. Harris and R. Stulz (eds.), *Handbook of Economics and Finance*, North Holland, Amsterdam

Hauswald, R. and R. Marquez (2005), Competition and strategic information acquisition in credit markets, forthcoming in *Review of Financial Studies*

Hellman, T.F., K. Murdock and J. Stiglitz (2000), Liberalization, moral hazard in banking and prudential regulation: are capital requirements enough?, *American Economic Review*, 90(1), 147-165

Holmstrom, B. and J. Tirole (1997), Financial intermediation, loanable funds and the real sector, *Quarterly Journal of Economics*, 112, 663-691

Jacklin, C.J. and S. Bhattacharya (1988), Distinguishing panics and information-based bank runs: welfare and policy, *Journal of Political Economy*, 96, 568-592

Jensen, M. and W. Meckling (1976), Theory of the firm: Managerial behaviour, agency costs and ownership structure, *Journal of Financial Economics*, 3, 305-360.

Keeley, M. (1990), Deposit insurance, risk and market power in banking, *American Economic Review*, 80, 1183-1200

Koskela, E. and R. Stenbacka (2000), Is there a tradeoff between bank competition and financial fragility?, *Journal of Banking and Finance*, 24(12), 1853-1874

Marquez, R. (2002), Competition, adverse selection and information dispersion in the banking industry, *Review of Financial Studies*, 15(3), 901-926

Matutes, C. and A.J. Padilla (1994), Shared ATM networks and banking competition, *European Economic Review*, 38, 1057-1069

Matutes, C. and X. Vives (1996), Competition for deposits, fragility and insurance, *Journal of Financial Intermediation*, 5(2), 184-216

Matutes, C. and X. Vives (2000), Imperfect competition, risk taking and regulation in banking, *European Economic Review*, 44(1), 1-34

McAndrews, J.J. and R. Rob (1996), Shared ownership and pricing of a network switch, *International Journal of Industrial Organization*, 14(6), 727-745

Merton, R. (1977), An analytical derivation of the cost of deposit insurance and loan guarantees, *Journal of Banking and Finance*, 1, 3-11

Nagarajan, S. and C.W. Sealey (1995), Forbearance, deposit insurance pricing and incentive compatible bank regulation, *Journal of Banking and Finance*, 19(6), 1109-1130

Parigi, B. (1998), Competition in banking: a survey of the literature, traduzione italiana in M. Polo (ed), *Industria Bancaria e Concorrenza*, Bologna: Il Mulino, 19-66

Perotti, E. and J. Suarez (2002), Last bank standing: what do I gain if you fail?, *European Economic Review*, 46(9), 1599-1622

Postleweite, A. and X. Vives (1987), Bank runs as an equilibrium phenomenon, *Journal of Political Economy*, 95, 485-491

Rajan, R. (1992), Insiders and outsiders: The choice between informed and arm's-length debt, *Journal of Finance*, 47, 1367-1400

Repullo, R. (2004), Capital requirements, market power and risk-taking in banking, *Journal of Financial Intermediation*, 13, 156-182

Rey, P. and J. Tirole (2000), Loyalty and investment in cooperatives, mimeo, University of Toulouse, Toulouse, France

Riordan, M. (1993), Competition and bank performance: A theoretical perspective, in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge, UK: Cambridge University Press

Rochet, J.C. and J. Tirole (1996), Interbank lending and systemic risk, *Journal of Money, Credit and Banking*, 28, 733-762

Rochet, J.C. and J. Tirole (2002), Cooperation among competitors: The economics of credit card associations, *RAND Journal of Economics*, 33(4), 1-22

Rochet, J.C. and X. Vives (2004), Coordination failures and the lender of last resort: Was Bagehot right after all?, *Journal of European Economic Association*, 2-6, 1116-1147

Sáez, L. and X. Shi (2004), Liquidity pools, risk sharing and financial contagion, *Journal of Financial Services Research*, 25(1), 5-23

Salop, S. (1979), Monopolistic competition with outside goods, *Bell Journal of Economics*, 10, 141-156

Schoenmaker, D., (1996), Contagion risk in banking, LSE Financial Market Group Discussion Paper, n. 204, London

Sharpe, S. (1990), Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships, *Journal of Finance*, 45, 1069-1087

Shy, O. and R. Stenbacka (2004), Market structure and risk taking in the banking industry, *Journal of Economics*, 82, 249-280

Stahl, D.O. (1988), Bertrand competition for inputs and Walrasian outcomes, *American Economic Review*, 78, 189-201

Smith, B.D. (1984), Private information, deposit interest rates and the "stability" of the banking system, *Journal of Monetary Economics*, 14, 293-317.

Vives, X. (1991), Banking competition and European integration, in A. Giovannini and C. Mayer (eds.), *European Financial Intermediation*, Cambridge, Cambridge University Press, 9-31

Vives, X. (2001), Competition in the changing world of banking, *Oxford Review of Economic Policy*, 17, 535-54

Vives, X. (2002), External discipline and financial stability, *European Economic Review*, 46, 2002, 821-828

Yanelle, M.O. (1989), The strategic analysis of intermediation, *European Economic Review*, 33, 294-301

Yanelle, M.O. (1997), Banking competition and market efficiency, *Review of Economic Studies*, 64, 215-239