This dissertation contains theoretical essays on financial intermediation and regulation. The main theme is the optimal regulation of banks, particularly capital regulation. The key question analyzed is how regulation affects the competitive dynamics of the banking industry, and, vice versa, how competition affects the effectiveness of capital regulation. These issues are analyzed in an industrial organization (IO) setting allowing for inter-bank competition and new entry. The most striking finding is that increasing costly capital requirements can lead to more entry into banking, essentially by reducing the competitive strength of lower quality banks. Other issues analyzed in the dissertation pertain to why banks combine lending and deposit taking, and how inter-bank competition affects the characteristics of an optimal loan contract. The dissertation also includes a review chapter on the foundations of banking, including rationales for regulatory interference in this industry. The latter also culminates in a public policy chapter on the design of regulation.

**Matej Marinč** (born in Ljubljana, 1979) has a bachelor's degree in mathematical physics and an MA in business administration from the University of Ljubljana. In September 2004, he began pursuing his PhD studies at the Amsterdam Center for Law & Economics at the University of Amsterdam. He is currently employed at the Faculty of Economics at the University of Ljubljana. His main area of research is financial intermediation.

ISBN: 978 90 5170 885 1

Essays on Bank Monitoring, Regulation and Competition

Matej Marinč

1

ACLE

# Essays on Bank Monitoring, Regulation and Competition

Matej Marinč

Research Series

Universiteit van Amsterdam
*Amsterdam Center for Law & Economics*

# ESSAYS ON BANK MONITORING, REGULATION AND COMPETITION

ISBN 978 90 5170 885 1

# ESSAYS ON BANK MONITORING, REGULATION AND COMPETITION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 24 april 2008, te 10:00 uur

door

MATEJ MARINČ

geboren te Ljubljana, Slovenië

Promotiecommissie

| | |
|---|---|
| Promotor: | *Prof. dr. A.W.A. Boot* |
| Overige leden: | *Prof. dr. R.M.W.J. Beetsma* |
| | *Prof. dr. D. Mramor* |
| | *Prof. dr. S.R.G. Ongena* |
| | *Prof. dr. E.C. Perotti* |
| | *Prof. dr. I. Ribnikar* |

Faculteit Economie en Bedrijfskunde

# Acknowledgments

It is a strange feeling to be at the end of a long road, where all that it is left is to thank my friends. Without them, this road would not have ended, or would have been much longer and dustier.

The person I have the most to thank for – not only for the academic quality of my work, but also for my personal growth – is my advisor, Arnoud Boot. I cannot count all the ideas that we encountered together. Nor can I forget the countless lines of the text that were improved, or the long equations that were so elegantly simplified during our discussions. Even more vividly, I fondly remember the great times that we had with his family in Spain.

None of this could have been accomplished without my wife Tatjana and our love for each other. She has succeeded in enjoying life with a spouse writing a dissertation, which may be even harder than the challenge of writing one. I also thank my family, who have always backed my (sometimes eccentric) ideas.

The people that I met in the Netherlands made my stay there a wonderful experience. Thanks to Marco and Muriel, Giuseppe, Iwan, Irene, and Audrey. Back in Slovenia, Vasja and I fought side by side.

I would also like to thank Ivan Ribnikar for his ongoing support and my committee members for their suggestions on how to improve the text. I am indebted to Marko, Anjolein, Tanja, Liljana, Aleksandra, Boštjan, and Igor for their insightful discussions.

Matej
Ljubljana, 10 March 2008

# Contents

# Chapter 1

# Motivation and Outline

## 1.1 Introduction and Motivation

This dissertation consists of seven theoretical essays. Its main theme is the analysis of optimal regulation of banks. Broadly speaking, the essays address the following three issues in financial intermediation theory: *i*) why do we need banks, *ii*) how do we regulate them; and *iii*) how does competition impact the main functions that banks perform and the effectiveness of regulation?

Banks can be considered a lubricant for the economy at large. Banks are active on the asset side as well as on the liability side of their balance sheets. On the asset side, banks have loans that need monitoring. These loans are often rather opaque, reflecting the idiosyncracies of bank borrowers that may not have access to financial markets. In particular, information asymmetries could make financial markets inefficient in financing informationally opaque businesses (e.g., small and medium enterprises). Banks aim to resolve this market failure by screening and monitoring borrowers (e.g., businesses that need financing). This allows banks to intermediate between investors (depositors) and borrowers. The additional information that banks possess inevitably makes them opaque (see Morgan (2002)).

On the liability side, banks create liquidity. That is, banks finance illiquid loans with liquid demand deposits. The opaqueness of banks on the asset side, together with liquidity provision on the liability side, makes banks inherently unstable institutions. More specifically, if many depositors unexpectedly withdraw their funds, the bank involved incurs high liquidation costs and even a previously solvent bank might be forced into liquidation. A broader, systemic bank crisis could result because the banking industry is highly interconnected and the failure of one bank can have negative repercussions for other banks and hence for the economy at large.[1]

Because such bank instability comes from a combination of banks' two core functions, the following question analyzed in this dissertation arises: *why do banks combine liquid demand deposits with lending to opaque borrowers that need monitoring?* Several narrow

[1] Empirical evidence points at a cost of a bank crisis ranging from 5-20% of annual GDP (see Hoggarth, Reis, and Saporta (2002) and Bordo, Eichengreen, Klingebiel, and Martinez-Peria (2001)).

banking proposals have emerged that call for the separation of the two core activities of the bank (see Bryan (1988)). In their view, demand deposits should be invested in liquid securities, whereas illiquid loans should be financed with noncheckable long-term liabilities. However, such separation is rarely observed nor is it demanded by the regulator. It seems that a combination of demand-deposits and monitoring gives banks a distinct competitive advantage.[2]

The potential negative externalities of bank instability could provide a justification for bank regulation. More specifically, banks on their own may not fully internalize the costs of failure, i.e. the social cost of failure exceeds the private cost. A widely accepted form of regulation is the implementation of deposit insurance. Deposit insurance prevents bank runs because depositors know that their money will be repaid anyway. However, it creates other distortions. Depositors may no longer carefully examine the stability of their banks. The regulator may now have to take full responsibility for the prudent behavior of each bank.

Several regulatory tools have been created to deal with banks' potential excessive risk taking. Intrusive regulation in the past included restrictions on activities, geographical limitations, and various limitations on banks' prices (e.g., a deposit interest rate ceiling). Nowadays, rather than being "brute forced" into desired behavior, prevailing regulation is more of an indirect nature, in which banks are indirectly "compensated" for being prudent or alternatively charged for risk taking. An example is capital regulation, in which banks are obliged to hold a level of equity capital that corresponds to the riskiness of their activities.[3]

Although the regulator should stand for the stability of the banking system, it also needs to take into account the efficiency of the industry. More specifically, higher competition may improve the efficiency of the banking industry and this could be good for stability, yet, in the extreme with cut-throat competition it may possibly lead to bank failures. Older studies argue that higher competition induces banks to behave more risky, and creates instability (see Keeley (1990)). Although these studies emphasize the potential negative impact of competition, recent studies are more positive (see Vives (2001a) and Boyd and De Nicolo (2005)). The question that emerges here is the following: *how does competition impact bank behavior and, in particular, can competition hamper the fundamental functions that banks perform?*

In addition, fierce competition between banks (domestic or foreign entrants) may also make stability-oriented bank regulation work differently than expected. In particular, I ask the question, *how do competition and capital regulation interact?*

[2]Only a few explanations have been proposed. Calomiris and Kahn (1991) argue that demand deposits exert pressure on banks to behave prudently. Diamond and Rajan (2001) show that demand deposits give power to the bank to prevent borrowers from renegotiating their contracts. Berlin and Mester (1999) argue that core depositors respond less to a potential change in the interest rate. Therefore, banks more heavily funded with core deposits can smooth loan rates more and better insulate borrowers against exogenous credit shocks. See my analysis in Chapter 8.

[3]An example of intrusive regulation was the Glass-Steagall Act in the U.S. that kept investment banking and commercial banking separated. See Boot, Milbourn, and Deželan (2001) on the distinction between direct and indirect regulation.

## 1.2 Analytical Concepts

I analyze these issues using models that build on information economics[4] and specifically game theory.[5] The concepts of information economics are widely used in financial intermediation literature.[6] In the most common situation, a principal hires an agent to perform a certain task. Contracting between the principal and agent is complex because of asymmetric information (i.e. hidden action or type). In the models that I use, borrowers may have an informational advantage over their bank, but also the bank may have proprietary information about its skills that its depositors and/or the regulator may not have. Agents may then act strategically and use information at their own advantage to extract rents. Anticipating this, the principal tries to mitigate the incentive problems and conflicts of interest that the information asymmetries may induce. This then boils down to the design of contracts and regulation.

I will analyze simple models that contain the deposit taking and lending operations of a bank using game theoretic tools. The Nash equilibrium concept is used extensively. I also use the framework of contemporary contract theory, which analyzes contracting between multiple principals and agents.[7]

Another important theme that appears in the analytical models throughout this dissertation is the notion of competition. I use several models to describe different forms of competition. Although pure price competition à la Bertrand is a useful starting point, banks usually operate in a less than perfectly competitive environment. This may for example be due to the presence of switching costs that borrowers incur when switching between banks. In Chapter 3, I use a simple search model in which borrowers search for competitive offers and may find them only with a certain probability.[8] In Chapter 5, I use a spatial model in which borrowers may "travel" to a competing bank to receive a competitive offer; in doing so, they incur transportation costs.[9]

I combine these models of competition with heterogeneity in bank quality. Heterogeneity induces shifts in market share from low quality to high quality players, a positive aspect of competition that has frequently been neglected in the banking literature.[10] Heterogeneity

[4]The literature on information economics starts with Akerlof's (1970) lemon problem and the signalling model of Spence (1973). For a comprehensive recent treatise see Bolton and Dewatripont (2005).

[5]Seminal contributions that have spurred developments in game theory include von Neumann and Morgenstern (1944) and Nash (1950). For a review of this field, see Fudenberg and Tirole (1991) and Vega-Redondo (2003).

[6]Early contributions in financial intermediation theory include Leland and Pyle (1977), Diamond (1984) and Ramakrishnan and Thakor (1984). For overviews, see Bhattacharya and Thakor (1993) and Freixas and Rochet (1999).

[7]See Bizer and DeMarzo (1992), McAfee and Schwartz (1994), Segal (1999) and Prat and Rustichini (2003) for different models of multi-agent multi-principal contracting.

[8]Important contributions to the search literature include Diamond (1971) and Stiglitz (1987).

[9]The spatial literature starts with Hotelling (1929). Salop (1979) presents a circular model of spatial competition with uniform pricing. I use a spatial model of perfect price discrimination as developed also in Thisse and Vives (1988). This better applies to the competitive environment in banking; see also Degryse and Ongena (2005).

[10]Exceptions that account for differences between banks are Kopecky and VanHoose (2006) and Sengupta

may come from inherent differences in skill, but could also come from an information advantage of the incumbent bank. The incumbent bank may have proprietary information about the quality of its borrowers. Both sources of heterogeneity will be used in this dissertation.

## 1.3 Outline of Dissertation

This dissertation consists of seven additional chapters. The core literature review is contained in Chapter 2. This chapter allows me to build a framework for further analysis and to position the main findings of the dissertation in the literature. Other chapters contain new theoretical contributions. I now give a brief overview of the various chapters and their main conclusions.

In Chapter 2 ("Foundations of Banking"), I present a unifying framework of banking in which banks both monitor their borrowers and provide liquidity to depositors. In this framework I review the role of several features of banks. I focus on banks' monitoring incentives, which are particularly relevant for their role in lending. This asset-side perspective is quite prevalent in the banking literature, and is the main focus of this dissertation. I also analyze bank fragility issues that may lead to bank runs and regulatory resolutions designed to contain bank fragility. Moreover, I discuss the potential positive disciplining effect of (the threat of) fragility in mitigating moral hazard. In addition, this review chapter discusses rationales for capital regulation and for banks to voluntarily hold capital.

The main analysis of the dissertation starts in Chapter 3 ("Competition and Entry in Banking: Implications for Capital Regulation"). I assess how capital regulation interacts with the degree of competitiveness of the banking industry. The most striking result is that increasing costly capital requirements can lead to *more* entry into banking, essentially by reducing the competitive strength of lower quality banks. I show that an implication of this is that banks on average may want the regulator to impose a higher capital requirement on the industry than is socially optimal. I also show that competition improves the monitoring incentives of better quality banks and deteriorates the incentives of lower quality banks; and that precisely for those lower quality banks competition typically compromises the effectiveness of capital requirements.

Chapter 4 ("Regulation of a Competitive Banking System") presents the policy implications of my analysis for optimal bank regulation. It seeks to highlight and answer some of the daunting challenges in the regulation of a highly competitive and rapidly changing banking system. Competition typically increases efficiency, yet, is traditionally considered a threat to stability. Recent research, however, argues that competition and stability could (sometimes) also go hand in hand. In this chapter, I discuss direct and indirect barriers to competition stemming from stability-oriented regulation. Despite the potential beneficial effects of competition, regulators should only cautiously liberalize their banking sectors. Liberalization, and the period of transformation that it implies, can induce opportunistic

(2007) Empirical evidence confirms the market shift effect. See Demsetz (1973) for evidence that competition spurs the growth of efficient players at the expense of less efficient ones.

behavior suggesting the need for caution. Moreover, competition might lower the effectiveness of existing regulatory tools. I also consider the efficiency and effectiveness of current regulatory practices in light of recent consolidation in the banking industry.

Chapter 5 ("Competition and Entry in a Spatial Model with Applications to Banking") recasts the analysis of Chapter 3 in a spatial framework (see Hotelling (1929) and Salop (1979)) in which firms are heterogeneous in their cost of production. The key result is that intensifying competition by lowering transportation costs can augment expected profits and entry, essentially by shifting market share from high-cost to low-cost firms. It is shown that increasing the differences in firm costs augments firm profits, entry, and social welfare. Lowering the production costs of high-cost firms augments entry if competition is low (i.e., for high transportation costs) but hampers entry if competition is high (i.e., for low transportation costs). When specifically applied to banking, the results of the analysis show that *i*) deposit insurance may lead to less entry, *ii*) capital regulation may lead to more entry in banking, and *iii*) bailout policies may reduce bank profits and entry, even if they come at no additional cost for banks.

Chapter 6 ("Bank Foreign Entry and the Role of Bank Capital") asks how capital regulation affects foreign entry. I first show that tighter capital regulation may lead to more foreign entry, which is reminiscent of the results in Chapter 3. I show that especially bad domestic banks may have incentives to merge to prevent entry of foreign banks. Interestingly, increasing capital requirements augments the merger incentives of banks. I then allow for discriminatory policies against foreign bank entry. In particular, I allow for higher capital requirements imposed only on foreign banks.

In the last two chapters, I analyze specific functions the banks perform. Chapter 7 ("Demand Deposits and Bank Monitoring") provides a novel rationale for why banks combine lending and deposit taking. I show that demand deposits may commit banks to monitoring in an environment in which monitoring is most valuable for long-term projects. Demand deposits, contrary to straight short-term and long-term debt, help commit a bank not to overleverage itself and this commits the bank to monitoring. I show that banks prefer demand deposits if the cost of early liquidation of borrowers is intermediate and if bank monitoring is costly.

The last chapter, Chapter 8 ("The Maturity of Monitored Finance: Covenants and Insufficient Liquidation"), analyzes the characteristics of an optimal lending contract in a model in which a bank obtains proprietary information by monitoring its borrowers. Key to the analysis is that with short-term contracts the incumbent bank's rent-seeking behavior will trigger competitive offers such that bad borrowers can also obtain financing and liquidation is insufficient. This happens if competition for borrowers is high and competition between banks is based on *simultaneous* offers. Using a long-term contract with covenants adds efficiency and leads to liquidation of more bad borrowers. This chapter shows that lower-risk borrowers use short-term contracts whereas higher-risk borrowers finance themselves with long-term contracts with covenants. As competition for borrowers increases and/or is expected to increase in the future, long-term contracts with covenants become more attractive to borrowers. Moreover, the use of covenants and maturity of contracts are positively correlated with monitoring precision.

# Chapter 2

# Foundations of Banking


**Abstract**

I present a unifying framework of banking in which banks both monitor their borrowers and provide liquidity to depositors. In this framework I review the role of several features of banks. I explain how diversification affects banks' monitoring incentives, which are particularly relevant for their role in lending. This asset-side perspective is quite prevalent in the banking literature, and is the main focus of this dissertation. I also analyze bank fragility issues that may lead to bank runs and regulatory resolutions designed to contain bank fragility. Moreover, I discuss the potential positive disciplining effect of (the threat of) fragility in mitigating moral hazard. In addition, this review discusses rationales for capital regulation and for banks to voluntarily hold capital.



Keywords: Bank Rationale, Bank Regulation, Overview
JEL CLASSIFICATION: G21, G28

## 2.1 Introduction

One of the key objectives of this chapter is to understand the role of banks in monitoring borrowers (the asset side). Through monitoring and screening, banks resolve the inaccessibility of capital markets for information-intensive borrowers. Following this, regulatory issues, in particular bank fragility concerns, are addressed. These originate to a large extent from the mismatch between (opaque) assets and liquid deposits. This chapter concludes with some insights from the literature on bank capital and capital regulation.

I survey some core results of the modern banking literature in a unifying model of banking in which banks monitor their borrowers in the sense that they seek to contain borrowers' behavior. A unifying framework allows me not only to review the key contributions of financial intermediation literature and address some open questions (following several surveys, e.g. Bhattacharya, Boot, and Thakor (1998) and Gorton and Winton (2003)), but also to reveal some new connections between various functions that banks perform. To facilitate this, however, sometimes stronger assumptions are made. I organize this review along the following questions: *i*) How banks can commit to monitoring? *ii*) Why banks provide liquidity to their depositors? *iii*) How a combination of monitoring and liquidity provision may cause fragility? *iv*) How fragility could induce depositors to push banks into monitoring? And *v*) why bank capital and capital regulation are critical in committing banks to monitor?

In addition to monitoring, banks also supply liquidity to depositors. Early contributions describing the creation of liquidity include Bryant (1980) and Diamond and Dybvig (1983). Banks provide intertemporal insurance to their depositors. In particular, demand-deposit contracts give depositors the right to withdraw their deposits upon demand. Such liquid contracts could add value because liquidity needs are random and in a large population of depositors diversification therefore allows banks to invest some of these funds in long-term (illiquid) projects. In doing so, banks enable investors to participate in profitable long-term lending without giving up liquidity. Banks could then potentially offer to pay a higher interest rate on short-term deposits in return for a lower long-term rate on deposits compared to capital market investments. For risk-averse investors, this could increase utility. Following Jacklin and Bhattacharya (1988) and Von Thadden (1999), I discuss the conditions for such smoothing to be possible.

While liquidity provision seems to create value for depositors (and hence the bank), it also has drawbacks. In particular, the combination of liquidity provision (via the sequential service constraint) and illiquid and opaque loans could make banks prone to bank runs. That is, the threat of excessive withdrawals could, via anticipated losses on fire sales of assets, make depositors run on their banks in order to retrieve their funds before default occurs. Such (costly) bank runs could provide a rationale for bank regulation, and deposit insurance in particular. I focus on two regulatory resolutions: suspension of convertibility and deposit insurance.

However, following Calomiris and Kahn (1991), a threat of a bank run may also have positive implications. That is, it may exert pressure on banks to follow safe strategies, and/or monitor their borrowers more rigorously. Such pressure may not be strong enough and/or (credible) deposit insurance – as an ultimate remedy to the bank run problem – may have removed it all together. In this case, bank capital could become even more important. In particular, bank capital could act as a commitment for depositors (or the deposit insurer) that their bank will truly monitor its borrowers. Such commitment might be in the self interest of banks, and thus banks may voluntarily choose positive levels of capital. If banks do not fully internalize the cost of a bank run, the regulator may have to impose capital requirements on banks.

I also reconsider the effect of deposit insurance on banks' monitoring incentives and on the probability of bank runs. I show that deposit insurance may subject banks to moral hazard. In particular, following the analysis of Merton (1977) I show that banks lower the level of monitoring if deposit insurance becomes rather generous.

The chapter is organized as follows. Section 2.2 focuses on the role of banks in monitoring. In Section 2.3, I focus on bank fragility stemming from the liquidity provision role of banks. Section 2.4 describes how the sequential service constraint and the liquidity of deposits may work in favor of bank monitoring; that is, they may commit banks to monitoring. In Section 2.5, I analyze the importance of bank capital and capital regulation. Section 2.6 concludes the chapter.

## 2.2 The Role of Banks in Monitoring

I now highlight and discuss the key conclusions of the literature on the existence of financial intermediaries. A bank's role in monitoring and screening borrowers is central to this literature. Monitoring serves to lower informational asymmetries between investors and borrowers. In particular, following Freixas and Rochet (1999), bank monitoring can be seen as information gathering as in Leland and Pyle (1977), Ramakrishnan and Thakor (1984), and Allen (1990). In this case, banks act on behalf of investors to acquire information about borrowers. Banks either observe the quality of the borrower or verify the cash flows of the borrower (see Townsend (1979) and Krasa and Villamil (1992)).

In addition, monitoring might denote an action of a bank against a deviating borrower as in Diamond (1984). In Holmstrom and Tirole (1997) bank monitoring contains opportunistic behavior. Townsend (1979) focuses on the role of banks and other monitors in auditing borrowers.

Several empirical studies provide evidence on the positive impact that a bank relationship might have on a borrower. James (1987) identifies a positive stock price response to the announcement of a new bank credit arrangement. Building on this, Lummer and McConnell (1989) distinguish between new bank loans and renewals and show that renewals have a positive announcement effect but new bank loans do not. Several other studies, however, show no significant difference between initiations of loans and loan renewals (see Slovin, Johnson, and Glascock (1992), Best and Zhang (1993) and Billett, Flannery, and Garfinkel

(1995)).[1]

I proceed as follows. I first specify a general framework that allows me to capture the core contributions to the literature on bank monitoring. Second, I extend the model to analyze the role of diversification. Third, I relate the analysis to the financial intermediation literature that points to the value of diversification in reducing monitoring costs (Leland and Pyle (1977) and Ramakrishnan and Thakor (1984)). Fourth, I explain the role of financial intermediaries in qualitative asset transformation.

### 2.2.1 Model specification

Here I present my basic model, which I will expand and modify in each section to study the problem at hand. All agents in the model – banks, borrowers, and depositors – are risk neutral.

The basic characteristics of a borrower are as follows. At $t = 0$, a borrower undertakes a project that demands \$1 of investment. Because the borrower has no initial funds, he must borrow from a bank. Monitoring occurs at $t = 1$. At $t = 2$, the project yields $R$ in the case of success and \$0 in the case of failure.

The bank's main role is to monitor the borrower. A bank incurs a cost $c_M$ to monitor the borrower. Although monitoring is costly for the bank, it increases the quality of the borrower's project. In particular, if monitored, the borrower's project becomes a good project that succeeds with probability $p_G$; if not monitored, the borrower's project becomes bad and it only succeeds with probability $p_B$, where $\Delta p \equiv p_G - p_B > 0$.[2]

The bank and depositors can invest in a risk-free asset with a risk-free return $r_F$. I assume that the expected return of a good project net of monitoring costs is higher than the risk-free return, whereas the expected return of a bad project is lower; that is,

$$p_G R - c_M > r_F > p_B R. \tag{2.1}$$

A bank is financed with deposits. For now, there is no deposit insurance. Also capital is not important. The bank promises to depositors a gross interest rate $r_D$, where $r_D \geq r_F$. I assume that the bank is a monopolist that seeks to capture all rents. Summarizing, the timeline is as follows. At $t = 0$, the bank raises funds from depositors and lends to the borrower. At $t = 1$, the bank monitors its borrower and, at $t = 2$, payoffs are realized and depositors are repaid. Initially I abstain from deposit insurance; see Figure 2.1.

[1] More recent evidence suggests that the importance of bank loan agreements is declining (see André, Mathieu, and Zhang (2001) and Fields et al. (2006)). However, loan agreements seem to keep their value for small firms, poorly performing firms, in times of greater economic uncertainty, and in banking systems with high quality lenders (see also Boscaljon and Ho (2005)).

[2] This captures the role that banks play in relationship banking: banks invest in borrower-specific knowledge that might be beneficial to their borrowers; see Boot and Thakor (2000) and Ongena and Smith (2000) for reviews of relationship banking.

| $t = 0$: | $t = 1$: | $t = 2$: |
| --- | --- | --- |
| ♠ The bank gathers funds from depositors and lends to borrower(s). | ♠ The bank can monitor its borrower(s). | ♠ Payoffs are realized. |

Figure 2.1: The basic sequence of events

### 2.2.2 Preliminary discussion: Monitoring incentives

I first analyze the incentives of a bank to monitor its borrower. That is, I compute the profit of the bank conditional on monitoring and compare this with what the bank earns without monitoring.

Conditional on monitoring, the bank's profit is computed as follows. Bank monitoring costs $c_M$ but increases the quality of the borrower's project. In particular, the borrower undertakes a good project that succeeds with probability $p_G$. If the borrower succeeds, the monopolistic bank takes rents $R$ but it must repay the deposit rate $r_D$ at $t = 2$. If the borrower fails, the bank gains zero profit and also fails to repay its depositors. The profit of the bank conditional on monitoring is

$$\Pi_M(r_D) = p_G[R - r_D] - c_M. \tag{2.2}$$

If the bank does not monitor, it saves the cost of monitoring. However, now the borrower undertakes a bad project and hence only succeeds with probability $p_B$. The bank's profit conditional on the absence of monitoring is

$$\Pi_{NM}(r_D) = p_B[R - r_D]. \tag{2.3}$$

Because bad projects have a negative NPV, a bank cannot operate if it abstains from monitoring. Depositors will anticipate this and will not provide funding. That is, depositors would demand a return $r_D$, such that the anticipated return $p_B r_D$ equals the return $r_F$ that they require; that is, $r_D = \frac{r_F}{p_B}$. However, with these costs of deposits the bank's profit is negative; that is,

$$\Pi_{NM}(\frac{r_F}{p_B}) = p_B R - r_F,$$

which is negative due to the condition in (2.1). Hence, the bank can only operate if it convinces depositors that it will monitor its borrower. A bank has an incentive to monitor if $\Pi_M(r_D) > \Pi_{NM}(r_D)$. Using (2.2) and (2.3), one needs

$$p_G[R - r_D] - c_M > p_B[R - r_D]. \tag{2.4}$$

At the same time, depositors demand a return $r_D$ such that their anticipated return $p_G r_D$ equals the return $r_F$; that is, $r_D = \frac{r_F}{p_G}$. Insert this into (2.4) to see that deposit taking and lending is only possible if

$$c_M \leq \bar{c}_M, \text{ where } \bar{c}_M \equiv \Delta p[R - \frac{r_F}{p_G}]. \tag{2.5}$$

The condition in (2.5) shows that the bank is only willing to finance the borrower if the cost of monitoring is not too high. Borrowers with high costs of monitoring (i.e., $c_M > \bar{c}_M$) therefore stay without funds even though they could undertake profitable projects (as guaranteed by the condition in (2.1)). The intuition for this is that a bank cannot commit to monitor such a borrower if the cost of monitoring is too high. In the absence of monitoring, the borrower undertakes an unprofitable project. Depositors anticipate this and are unwilling to finance such a bank. Hence, the bank cannot afford to finance such a borrower.

The reason why a borrower with (potentially) profitable projects can stay without bank financing deserves further explanation. Because monitoring is not contractible, the bank should be incentivized correctly, such that monitoring is also beneficial for the bank. However, if the cost of monitoring is high, this may be difficult. In this case, the bank may prefer to stop monitoring (and spare the cost $c_M$) and gamble on the project. If the borrower succeeds, the bank gains; in the case of a failure, depositors lose.

I follow the banking literature and present several ways how banks commit to monitoring. I start with the role of diversification. Financing several borrowers at the same time and diversifying allows the bank to better commit to monitoring.

### 2.2.3 Diversified intermediaries

One of the main characteristics of financial intermediaries is that they raise funds from many depositors and lend to many borrowers. In the following simplified framework I analyze how diversification affects a bank's incentives to monitor its borrowers. A key result is that a diversified bank can easily commit to monitoring.

I allow for diversification as follows. The bank lends \$1 of funds in total to $N$ equally sized borrowers, such that each borrows $\frac{1}{N}$ with a return of $\frac{R}{N}$. For a more general distribution, see Tirole (2006, p. 158). As before, $p_B$ is small, such that the bad project is negative NPV.

I distinguish between two cases. In the first case, projects are perfectly correlated. A bank that finances $N$ borrowers then receives in total $p_G[N\frac{R}{N} - r_D] - c_M$ if it monitors (which is the same as in (2.2)) and $p_B[N\frac{R}{N} - r_D]$ if it stops monitoring its borrowers (the same as (2.3)).

However, if borrowers' returns are independent, the situation changes. I now establish that a bank is then induced to monitor more. Assume for now that the bank is solvent as long as at least one borrower succeeds. It is easy to see that the bank now succeeds with a higher probability $p_G(N) = 1 - [1 - p_G]^N$ in the case of monitoring and $p_B(N) = 1 - [1 - p_B]^N$ in the absence of monitoring, where $p_B(N) > p_B$ and $p'_G(N) > 0$ and $p'_B(N) > 0$, than without diversification. The expected return of each borrower conditional on monitoring vs. no monitoring is the same as without diversification; that is, $R$ vs. $p_BR$. However, a diversified bank succeeds more often even if it abstains from monitoring. That is, it has to repay its depositors with probability $p(N)$. That is, in expectation it repays to depositors $p(N)r_D$.

Hence, in the case of independent borrowers' returns, the incentive constraint for the bank to monitor its borrowers changes. The following inequality has to be satisfied,

$$p_G R - p_G(N)r_D - c_M > p_B R - p_B(N)r_D. \tag{2.6}$$

Note that the left side of (2.6) describes the profit of the bank in the case of monitoring. Hence the bank receives $p_G \frac{R}{N}$ on each borrower; hence, in total $p_G R$. However, it has to repay depositors, which in expectations cost $p_G(N)r_D$, and incurs monitoring cost $c_M$. The right side of (2.6) is the expected profit if the bank deviates and stops monitoring. In expectations, the bank receives $\frac{p_B R}{N}$ on each borrower; hence in total $p_B R$. The bank now incurs no monitoring cost and repays $r_D$ to depositors with probability $p_B(N)$.

Rearranging (2.6) shows that the bank monitors the borrowers if

$$c_M < \bar{c}_M(N), \text{ where } \bar{c}_M(N) \equiv \bar{c}_M + [\frac{p_B(N)}{p_G(N)} - \frac{p_B}{p_G}]r_F, \tag{2.7}$$

and $\bar{c}_M$ is as defined in (2.5). I can now show the following proposition.

**Proposition 2.1.** *Diversification enables the bank to finance borrowers for which a cost of monitoring is high; that is,* $\bar{c}_M(N) > \bar{c}_M$.

The intuition for Proposition 2.1 is the following. Note that diversification does not change a bank's expected return on its borrowers. However, a well-diversified bank succeeds with higher probability in repaying its depositors. This then contains the risk-shifting incentives of a bank. In particular, now the bank instead of depositors must internalize a larger part of the losses if one of its borrowers fails. Hence, the bank is incentivized more to monitor its borrowers.

Diamond (1984) shows that completely diversified financial intermediary can easily commit to monitoring (see Section 2.2.5 for further discussion). Proposition 2.1 replicates this result. In particular, $\bar{c}_M(N = \infty) = \Delta p R$; hence, (2.1) guarantees that $c_M < \bar{c}_M(N = \infty)$ always holds. That is, the bank can always commit to monitoring if it is completely diversified.

Proposition 2.1 has the following empirical implication. It predicts that a better diversified bank might be willing to finance more opaque borrowers – that is, borrowers that demand higher monitoring (i.e., small, more opaque borrowers). Empirical evidence on this aspect is inconclusive, however. Different lending technologies differently affect banks' abilities to lend to opaque borrowers. Large banks seem to be better at gathering hard information through transaction technologies such as factoring, leasing, small business credit scoring (see Berger, Rosen, and Udell (2007) and Berger and Udell (2006), whereas small banks are better at gathering soft information through relationship lending (see Stein (2002) and Berger et al.

(2005)).

Now I can extend the analysis a bit further to give predictions on how increased competition for deposits would change the incentive of a bank to monitor. If competition for deposits increases, a bank may have to offer depositors a higher deposit rate. A simple way to analyze the effects of higher competition for deposits is to observe the effect of an increase in $r_F$.

**Corollary 2.1.** *An increase in competition (via higher $r_F$) lowers the monitoring incentives of a bank.*

Corollary 2.1 replicates the analysis of Keeley (1990). The intuition is as follows. As competition for deposits increases, the bank expects to earn lower rents, hence lowering the attractiveness of its marginal borrowers. In effect this means that a bank has lower incentives to incur monitoring costs in order to increase its success probability. That is, lowering bank rents makes monitoring less valuable.

The following corollary connects the effects of diversification and competition.

**Corollary 2.2.** *Diversification is (even) more effective if competition for deposits is high (high $r_F$).*

Corollary 2.2 stems from two observations of Corollary 2.1 and Proposition 2.1. First, if competition for deposits is high, the risk-shifting behavior of a bank is more pronounced (see Corollary 2.1). Second, diversification reduces risk shifting because it increases the success probability of a bank. That is, a bank better internalizes the costs of less monitoring (see Proposition 2.1). Hence, diversification that contains risk shifting has the highest effect if risk shifting is high, that is, if competition is high. In contrast, if competition for deposits is low, the risk-shifting behavior of a bank (i.e., lowering monitoring) is relatively limited and diversification has a lower effect.

Up to my knowledge this result has not yet been derived. Winton (1997) establishes a reversed relation between diversification and competition. He argues that bank owners have less incentives to engage in competition if a bigger bank becomes more diversified. In particular, higher competition brings additional market share and additional diversification which benefits mostly debtholders. Corollary 2.2 shows that diversification is especially valuable if competition between banks is high because a diversified bank can commit to monitoring. This contradicts Besanko and Thakor (1993) who argue that higher competition may induce banks to select risky, less diversified portfolio. Note, however, that their result depends on the presence of deposit insurance. Without deposit insurance, a bank in a competitive banking system can not raise deposits without being diversified. Thus, competition in this case forces banks to diversify via a market discipline mechanism.

The implications of Corollary 2.1 and Corollary 2.2 are immediate. Competition may undermine a bank's incentive to monitor. Diversification then helps because it mitigates moral hazard.

I have presented a very simplified model to explain the bank's role in monitoring. Banks that pool many borrowers together and thus diversify are better at committing to monitoring their borrowers. In the next section I link the main insights from my simple model of bank monitoring to literature on banks' existence. Central to this literature is that bank monitoring contains information asymmetries between firms and investors.

### 2.2.4 Information acquisition: Intermediaries as information sellers

I now review the contributions that focus on the role of banks as information producers. In this view, agents endogenously join together and form a bank because this allows for more efficient information production than each individual agent could achieve on his own.

Broadly speaking, one could argue that banks mediate information asymmetries between firms and investors. Normally, firms have more information about the prospects of their projects than investors. Especially if investors are small, they have little incentive to incur monitoring costs in order to lower information asymmetries. Banks, however, pool a large number of investors together and as such could be delegated information producers.

The cost advantages that banks possibly achieve in information acquisition stem not only from pooling together funds from small investors (some type of scale economies), but also could be skill related (e.g. benefits of specialization). In addition, banks could possibly reuse information across time and across different borrowers better than individual agents could (see Chan, Greenbaum, and Thakor (1986)).

Leland and Pyle (1977) were the first to rationalize banks as mediators of information asymmetries. They applied insights from the (then) emerging field of information eonomics to financial intermediation; see Akerlof (1970), Spence (1973) and Rothschild and Stiglitz (1976). In their analysis, entrepreneurs differ in the profitability of their projects. Risk-averse entrepreneurs prefer borrowing funds to self investing; however, borrowing might be limited due to information asymmetries. In particular, because investors cannot determine the profitability of entrepreneurs, they charge entrepreneurs a high lemon cost of financing. That is, charging the actuarially fair average cost is not sustainable because this would induce an adverse selection problem with only risky entrepreneurs raising funds at this average rate. Leland and Pyle (1977) argue that the "good" entrepreneurs choose to (partially) self finance to separate themselves from bad risks. Building on this, Leland and Pyle (1977) suggest that good entrepreneurs may form a coalition to lower the costs of such separation. Diamond (1984) expands this reasoning by showing that potential diversification in the coalition could lower the costs of this separation even further.

Ramakrishnan and Thakor (1984) explain the information acquisition function of financial institutions. They focus on the incentive contract between an investor as a risk-averse principal and risk-averse agents that gather information about the prospects of borrowers; that is, they screen firms. Ramakrishnan and Thakor (1984) show that diversification among information-producing agents may help reduce the agency problem between the principal and information-producing agents. As long as the projects of the firms are uncorrelated and agents can observe their respective efforts (no moral hazard between agents), the formation of a coalition lowers the agency cost per agent.

Ramakrishnan and Thakor (1984) argue that two conditions must be satisfied for financial intermediation to occur; that is, the formation of a coalition of information-producing agents. First, the projects should be diversifiable (i.e., the correlation should not be too high). Second, sufficiently many projects should exist; that is, sufficient diversification should be possible. In this case, the total agency costs of the financial intermediary become small compared to the sum of total agency costs of individual entrepreneurs.[3]

Allen (1990) shows that a financial intermediary might be better at selling information than an individual agent. This is because agents on their own may be forced to leave some rents in order to commit to truthful reporting of information. A coalition of agents in a financial intermediary can better commit to truthful reporting and hence extract additional rents from information selling.

The financial intermediary as described above could be viewed as a pure brokerage institution that produces information for resale. Examples are analysts providing financial advice on investments and credit rating agencies that screen and certify firms, and bond issues.

In summary, agents that produce and resell information may have incentives to form coalitions. I have reviewed several contributions that discuss such endogenous formation of financial intermediaries and their role as information producers. The scope of these papers is quite broad and may be applied to many different types of financial intermediaries. In fact, with the increasing importance of fee business in banking, such brokerage rather than the asset-transformation role is gaining importance. However, depository financial institutions involving asset transformation are (still) of considerable importance. The next chapter focuses on such institutions and, in particular, on the asset transformation role that they perform.

### 2.2.5 Asset transformation

Banks not only produce information but they also act as asset transformers. They provide for transformation of maturity, liquidity, and risk between their assets and liabilities. In particular, banks fund risky borrowers with low-risk deposits. Analyzing this asset transformation role also illuminates why debt contracts may be an optimal contract between banks and depositors and between banks and borrowers.

Diamond (1984) provides the following explanation of the asset transformation role of banks. Without banks, depositors could lend funds directly to borrowers. Doing so, depositors should monitor borrowers to guarantee repayments of the loans. This, however, may

[3] Both Diamond (1984) and Ramakrishnan and Thakor (1984) predict that banks would be of infinite size. In reality, several small banks exist as well. Millon and Thakor (1985) show that the underlying assumption of Ramakrishnan and Thakor's model is that agents costlessly observe the efforts of other agents in the coalition. If the efforts in the coalition are observable only at a cost, several finite financial institutions may appear.

result in several inefficiencies. First, monitoring may be duplicated. Second, depositors may free ride on each other, such that at the end nobody monitors. A bank that gathers funds from depositors and lends them to borrowers resolves these monitoring problems. Although the bank still has to monitor borrowers, Diamond (1984) shows that depositors do not need to monitor a well-diversified bank. He shows that writing a standard debt contract between depositors and a well-diversified bank may guarantee bank monitoring. With this he elegantly solves the problem of who monitors the bank – in his view, nobody needs to monitor a well-diversified bank.

The rationale for why diversification guarantees bank monitoring is the following. Diamond (1984) adds the possibility of non-pecuniary punishment to the standard debt contract between depositors and the bank.[4] In particular, if the bank breaches its contractual obligations and fails to repay depositors, it is punished. Punishment may be an instrument that helps induce the bank to always monitor the borrowers. A dark side of punishment is that an (undiversified) bank may fail even if it monitors the borrower. In this case, punishment is costly for a bank. However, a well-diversified bank that monitors will never fail and costly punishment will never be realized. The sole threat of punishment is enough to incentivize a well-diversified bank to monitor. Hence, punishment is a costless device to induce a well-diversified bank to monitor.

Despite its unquestionable importance, Diamond's analysis has some shortcomings. In particular, the existence of non-pecuniary punishments seems to be unrealistic. While a literal interpretation is far-fetched – banks neither physically punish their defaulted borrowers nor put them in jail – a more realistic interpretation such as the loss of reputation as a mechanism of non-pecuniary punishment also has some difficulties. In particular, investors can hardly fine-tune the loss of reputation to the final outcome of the returns of the bank. However, this is necessary to commit the bank to truthfully report its returns while minimizing the cost of punishment. Moreover, Diamond's analysis suggests that intermediaries should be infinite, which does not seem realistic.

Several other contributions complement Diamond (1984) and further justify the existence of a debt contract. Townsend (1979) and Gale and Hellwig (1985) use a costly state verification approach to justify the existence of the standard debt contract. If it is costly to monitor the performance of borrowers, banks design their lending contracts to minimize monitoring costs. Townsend (1979) and Gale and Hellwig (1985) show that conditioning monitoring on default is an optimal strategy. This leads to a debt contract in which equal payments are made in non-default. A few assumptions are that banks are limited to deterministic monitoring and borrowers are risk neutral. In other words, they show that a standard debt contract minimizes the monitoring cost. In addition, the bank should monitor more if the firm performs badly.

Another rationale for the use of debt contracts is given by Gorton and Pennacchi (1990).

[4]In Diamond (1984), the punishment is a non-pecuniary penalty such as the loss of reputation or physical punishment.

In Gorton and Pennacchi's contribution, the role of a financial intermediary is to design securities that split the cash flows of the underlying assets into securities of different informational content. In their setting, debt contracts are used to protect uninformed traders on the capital markets. They show that firms optimally divide their cash flows into two securities, equity and debt. Uninformed traders protect themselves by buying a security with low informational content; that is, debt security. In contrast, informed traders buy equity contract that contains a lot of information about firm profitability. This is reminiscent of later work on optimal security design by Boot and Thakor (1997) and DeMarzo and Duffie (1999).

To summarize, financial intermediation occurs because markets may be unable to resolve information asymmetries. The various contributions discussed above consider various types of informational frictions and present rationales for different forms of financial intermediaries such as banks, investment banks, credit-rating agencies, and mutual funds.

## 2.3 Bank Fragility and Regulation

Another feature that distinguishes banks from other firms is that they provide liquidity to their depositors. Liquidity provision is intrinsically linked to the sequential service constraint (SSC) in demand deposits. A feature that could make liquidity provision beneficial is that banks pool together large group of deposits and insure depositors against shocks in their consumption needs. Banks can invest part of the gathered funds in liquid assets (to cover the stochastic early consumption need of depositors) and another part in illiquid loans. They then use the liquid assets to repay depositors with early consumption needs while depositors with late consumption needs are repayed with illiquid loans. Having many depositors makes this liquidity need rather predictable in the aggregate. Yet, liquidity provision also makes banks intrinsically fragile institutions. That is, if many depositors unexpectedly withdraw their funds for reasons other than liquidity needs, a bank might have to liquidate its illiquid loans at possibly high costs.

In this section I first evaluate the benefits of liquidity provision as put forward by Diamond and Dybvig (1983). Key in Diamond and Dybvig is the possibility of a sunspot bank run. I also evaluate two regulatory mechanisms that prevent bank runs: deposit insurance and limitations on convertibility.

### 2.3.1 Liquidity provision

Following Diamond and Dybvig (1983), I use the following simple model. Depositors are risk-averse and initially identical. Each depositor invests his endowment of \$1 at $t = 0$. At $t = 1$, he finds out whether he must consume immediately or must wait on consumption until $t = 2$. With the short-term return $r_{D1}$ and the long-term return $r_D$, consumption at $t = 1$ brings him utility $U(r_{D1})$. If a depositor must wait on consumption until $t = 2$, he obtains utility $U(r_D)$. For simplicity, I assume that the depositor's utility function has a constant

| $t = 0$: | $t = 1$: | $t = 2$: |
|---|---|---|
| ♠ The bank gathers funds from depositors and invests in long-term projects. | ♠ Depositors may have a liquidity need.<br>♠ The bank liquidates some long-term projects to repay early withdrawals. | ♠ Payoffs are realized. |

Figure 2.2: The sequence of events accounting for depositors' liquidity needs

intertemporal relative risk aversion:

$$U'(r) = r^{-a}, \text{ and } U(r) = \frac{r^{1-a}}{1-a},$$

where $a > 1$.[5] The probability of an early consumption need is $\rho$ and the probability of late consumption is $[1-\rho]$. Probabilities are independently distributed among depositors. Hence, the expected utility of a depositor is

$$u(r_{D1}, r_D) = \rho U(r_{D1}) + [1-\rho]U(r_D). \tag{2.8}$$

There exist only one type of (long-term) project to invest in. Investments are made at $t = 0$. The (long-term) project payoff is $R$ at $t = 2$ per \$1 of invested funds. Early liquidation, however, is possible. If the project is prematurely liquidated, it returns only the nominal invested funds; see Figure 2.2.

I first analyze a situation of autarky in which depositors directly invest in the project (finance the borrower). With probability $\rho$ depositors must consume early; that is, at $t = 1$. This forces them to liquidate their investments, realizing each a return of \$1. With probability $[1-\rho]$ they consume late. In this case, they receive a return $R$. A depositor's expected utility with such direct investment is (use (2.8))

$$U_A = \frac{\rho}{1-a} + \frac{1-\rho}{1-a}R^{1-a}. \tag{2.9}$$

When there is only one investment opportunity, a trade between depositors cannot improve the outcome in autarky. This is because everybody invests in the same type of project and trade cannot occur. This is different if depositors can invest in projects of different types. Then trade occurs between depositors that have a short-term project but a long-term liquidity need and depositors with a long-term project but a short-term liquidity need. The option to trade then increases depositors' utilities (see Bhattacharya and Gale (1987) and Von Thadden (1997)).

Now I compare autarky with the outcome that a bank can provide. For the moment, I assume that the bank only performs liquidity provision ignoring asset-side monitoring. The

[5] The assumption of $a > 1$ is common in the literature (see Diamond and Dybvig (1983) and Von Thadden (1999)). It guarantees that the liquidity provision role of a bank is valuable.

bank's added value thus should come completely from its role in liquidity provision. A bank promises its depositors the deposit rate $r_{D1}$ for early withdrawals and $r_D$ for late withdrawals. Although initially the bank invests all its funds in (long-term) projects, it must liquidate the proportion $\rho r_{D1}$ of its investments to repay early depositors. The proportion $1 - \rho r_{D1}$ of investments is not liquidated and yields a final return $R$.

The bank solves the following optimization problem:

$$\Pi^* = \max_{r_{D1}, r_D} [1 - \rho r_{D1}]R - [1 - \rho]r_D \tag{2.10}$$

$$s.t.$$

$$\rho U(r_{D1}) + [1 - \rho]U(r_D) \geq U_A. \tag{2.11}$$

The bank maximizes its expected profit in (2.10) under the condition that the incentive constraint in (2.11) is satisfied. In particular, the bank must offer depositors at least the utility $U_A$ that they can obtain in autarky (see (2.9)).

In equilibrium, a bank offers $\{r^*_{D1}, r^*_D\}$, which solves (2.10), where $1 < r^*_{D1} < r^*_D < R$. The bank makes positive profits, that is, in equilibrium $\Pi^* > 0$. This shows that banks create value for depositors through liquidity provision. More specifically, banks provide intertemporal insurance for depositors. That is, intertemporarily risk-averse depositors value intertemporarily smoothed payoffs. Put another way, in expected value sense the depositors' utility increases if the bank subsidizes depositors with early consumption needs at the expense of depositors with late consumption needs. This replicates Diamond and Dybvig (1983). In Diamond and Dybvig (1983), however, banks maximize the depositors' utilities whereas in my case banks maximize their own profits. That is, banks may achieve equal utility for depositors compared to autarky and at the same time realize some profits.

The literature includes several follow-up studies extending the findings of Diamond and Dybvig (1983). Jacklin (1987) shows that if deposit contracts are allowed to be traded at $t = 1$ this would unravel the bank's added value; that is, in such a case carefully designed equity contracts providing dividends in the interim period could provide just as much liquidity provision as a demand-deposit contract.

In response to this critique, Bhattacharya and Gale (1987) and Von Thadden (1997) argue that, in addition to liquidity provision, banks also transform the maturities of assets. In particular, allowing for a second, short-term investment opportunity, they show that banks better allocate funds between short-term and long-term assets providing for liquidity *and* maturity transformation superior to what markets can achieve.

The rationale why banks provide better intertemporal insurance and maturity transformation than capital markets deserves further attention. With capital markets, depositors cannot commit to the level of price at which assets will be traded. Consequently, the optimal allocation between short-term and long-term assets is not an equilibrium outcome. In particular, if depositors select the optimal allocation, each depositor has an opportunity for a valuable deviation. That is, a depositor can invest more in a long-term and less in a short-term asset than in the optimal allocation. By doing this, the depositor can gain in trade, after the price is formed on the market. In contrast, a bank can commit to the socially optimal allocation because it can commit ex-ante to the fixed payoffs of depositors.

Several other contributions further highlight the role of liquidity provision. Von Thadden (1997) shows that the liquidity provision and the monitoring function of banks interact and should not be viewed separately. Pooling of intermediation functions may reduce moral hazard problems and enhance bank stability by mediating the liquidity problems of demand-deposit contracts.

Allen and Gale (1997) use a standard overlapping generations model to show that financial intermediaries provide for better intertemporal insurance than capital markets. More specifically, people cannot insure against a bad outcome ex ante; that is, before they are born. This makes capital markets incomplete such that underinvestment in safe assets occurs. Financial intermediaries can smooth the returns by accumulating reserves of safe assets and this may lead to a Pareto efficient outcome.

So far, I have used the word *liquidity* in the context of Diamond and Dybvig (1983). In particular, an asset is considered liquid as long as agents can choose when to use it (intertemporarily) for consumption when they would like to do so. In contrast, Holmstrom and Tirole (1998) provide for a slightly different definition of liquidity. In their view, instruments are liquid if they allow intertemporal transformation of welfare. Holmstrom and Tirole (1998) argue that banks act as coordinators of the use of scarce liquidity. That is, banks redistribute liquidity from firms that have excess liquidity to firms that need liquidity.

Several distinct notions of liquidity might be the reason why empirical research on bank liquidity creation is scarce. Kashyap, Rajan, and Stein (2002) find evidence that banks provide liquidity both on the liability side (via demand deposits) as well as on the asset side via loan commitments. To the extent that deposit withdrawals and commitment takedowns are imperfectly correlated, banks create synergies in combining both functions.

Deep and Schaefer (2004) and Berger and Bouwman (2007) attempt to directly measure liquidity creation by banks. Deep and Schaefer (2004) only consider on–balance sheet items. They compute the liquidity transformation gap as the difference between liquid liabilities and assets held by a bank, scaled by its total assets. Using this measure they show that liquidity transformation of the U.S. banking industry is, at best, modest. However, Berger and Bouwman (2007) include off–balance sheet items in their analysis and obtain a reversed result. They estimate that the U.S. banking industry creates \$2.5 of liquidity per \$1 of capital. They also find that bank liquidity creation has increased substantially in the last decade. Nevertheless, these studies on liquidity creation are still very incomplete. For example, Berger and Bouwman (2007) do not capture the liquidity created by non-banking vehicles (e.g., ABCP conduits).

To summarize, several papers highlight banks as liquidity providers. Future research should develop testable hypotheses pinpointing the exact role that banks play in liquidity provision. In the next section, I turn to the fragility that a combination of liquidity provision and investing in opaque assets may bring.

### 2.3.2 Sunspot bank runs vs. information-based bank runs

So far, I have presented a rather positive perspective on banks and in a sense on bank stability as well. In the view of Diamond (1984) and Ramakrishnan and Thakor (1984), banks can completely diversify and hence they never fail (see Section 2.2). However, this assumption is frequently violated in reality. That is, banks cannot completely diversify themselves and hence default risk may exist. In the context of Diamond and Dybvig (1983), bank activity is more at risk. The combination of opaque assets and liquid (withdrawable on demand) deposits could expose banks to costly runs. I now first discuss the potential fragility of banks stemming from banks' liquidity provision function. Second, I turn to an information-based rationale for bank runs; namely, averse information about a bank's loans that may trigger a bank run.

I first focus on a purely liquidity-driven bank run as first described by Diamond and Dybvig (1983). In terms of this model, if "many" depositors withdraw early a bank may have to liquidate more projects than anticipated. Because early liquidation is costly (assets need to be liquidated forgoing returns) and early withdrawals have a subsidized rate, late depositors suffer. With sufficiently high unexpected early withdrawals, late depositors might be better off withdrawing their funds prematurely as well. Even worse, trying to beat the crowd by being first in line is becoming an increasingly attractive option when many other early withdrawals are anticipated. This triggers what is known as a sunspot bank run.

The optimization problem in (2.10) has two equilibria. In the first, only a proportion $\rho$ of the depositors withdraw early, and the others withdraw late. In the second, Pareto inferior equilibrium, all depositors withdraw early.

A prevalent criticism of sunspot bank runs is that they lack a trigger mechanism. Sunspot bank runs in Diamond and Dybvig (1983) are not correlated with the economic variables. Empirical findings contradict this. Empirical studies persistently show that bank runs are a consequence of deterioration of bank quality reflected in different economic indicators; for example, increasing small business failure rates (see Gorton (1988) and Demirgüç-Kunt and Detragiache (2002)). I respond to this criticism first by reviewing several contributions in the literature that describe the notion of the informational bank run (see Chari and Jagannathan (1988) and Jacklin and Bhattacharya (1988)). In Section 2.4, I extend the model by analyzing the potential disciplining role that the threat of a bank run could have.

Banks are characterized both by liquidity provision *and* by having opaque assets. In particular, giving credit to monitoring-intensive borrowers makes banks opaque on the asset side (see Morgan (2002)). A combination of opaque assets and liquid deposits makes banks sensitive to what is called information-based bank runs. Chari and Jagannathan (1988) and Jacklin and Bhattacharya (1988) analyze such information-induced bank runs. In their setting, a group of depositors observe the risk of a bank and, when they observe (too) high risk, they withdraw early. Other depositors, however, observe only the amount of the withdrawals (i.e., the line in front of the bank), but not the underlying information of bank asset quality. Hence, they do not know whether early withdrawals are a consequence of early liquidity needs or are information based. Consequently, information-driven bank runs and also sunspot bank runs can both occur as equilibrium events.

In a related model, Goldstein and Pauzner (2005) connect the probability of a bank run to underlying economic fundamentals. In their model, the beliefs of investors are determined by economic fundamentals. That is, economic fundamentals do not directly induce withdrawals, but instead serve as a device that coordinates the beliefs of investors on a particular outcome. They show that inefficient (i.e., sunspot) bank runs occur more often at a bank that offers higher short-term interest rates; that is, at banks that offer more intertemporal risk sharing. A bank will then weight the benefits of risk sharing with the drawback of inducing an inefficient bank run in determining the level of liquidity provision for its depositors.

To summarize, financial intermediation theory argues that bank runs may either be pure sunspot phenomena or be information based. In both cases, bank runs are costly events that may provide a rationale for bank regulation.

### 2.3.3 Cost of bank runs and regulatory resolutions

The banking sector is considered special because of the externalities that banks could impose on the economy at large. A bank run imposes externalities and possibly negatively affects social welfare for several reasons. First, a bank run may potentially spread and provoke a systemic bank crisis. Second, bank failures can produce a sharp monetary contraction and induce a recession (see Bernanke (1983) for an argument that banking crises deepened the severity of the Great Depression). Third, the failure of banks may reduce the supply of bank loans, which is especially detrimental to small- and medium-sized business financing (see Hubbard, Kuttner, and Palia (2002)). A total collapse of a banking system might even cause a breakdown of the payment system and impair trade.

Empirical research confirms that the costs of bank crises are high. However, no agreement exists on how costly crises are. First, in cross-country studies, Hoggarth, Reis, and Saporta (2002) assess the costs of banking instability at 15 to 20% of annual GDP. Second, Bordo, Eichengreen, Klingebiel, and Martinez-Peria (2001) extend their analysis over the past 120 years of financial history to show that the frequency of crises has been rising. They assess the output losses of a crises at 6 to 8% of an annual GDP for a single banking crisis, but well over 10% if a banking crisis is accompanied by a currency crisis. In any case, these estimates confirm that banking crises are costly.

The broad economic importance of bank stability and considerable costs of bank instability to the economy at large justify the existence of extensive banking regulation. There are two main resolutions to costly bank runs: suspension of convertibility and deposit insurance.

Suspension of convertibility seems to be the optimal response in a case in which the liquidity shocks are i.i.d. among depositors. In this case, the bank only repays an expected proportion of depositors in the first period. Announcing this, and assuming an infinitely large bank, provides perfect resolution. In particular, suspension of convertibility precludes losses because it prevents excessive early withdrawals, which would otherwise lead to firesales. Depositors anticipate that there will be no losses and are comfortable with leaving their funds in the bank. Consequently, only depositors with early liquidity needs raise their funds.

Suspension of convertibility becomes more problematic if liquidity shocks are partially correlated (and/or not fully diversified). In this case, suspension of convertibility induces suboptimal liquidity provision because not all depositors with an early liquidity need are able to withdraw their funds (see Diamond and Dybvig (1983)). In the case of information-based bank runs, matters are more complicated. Suspension of convertibility may again lead to a suboptimal liquidity provision (see Chari and Jagannathan (1988)).

If the implicit costs of suspension of convertibility are high, deposit insurance may become a preferred regulatory policy. Credible deposit insurance eliminates bank runs. Excessive early withdrawals are then backed by government bonds and taxed. However, deposit insurance is *not* socially costless. The cost of deposit insurance stems from the deadweight costs that taxation brings. In a richer setting, either regulatory intervention could also have an impact on the given problem between a bank and investors.

Gorton and Pennacchi (1990) justify the existence of deposit insurance in another way. In their view, deposit insurance is valuable because it protects uninformed agents (i.e., depositors), from being abused from the informed traders that hold bank equity.

To summarize, deposit insurance is an optimal regulatory policy if the deadweight costs of taxation are lower than the costs of suboptimal liquidity provision that the suspension of convertibility could bring. In this section, I have presented a view in which bank runs are purely negative phenomena and regulation is needed to prevent them. In the next section, however, I present an argument why (the threat of) bank runs might serve a useful economic function. In particular, the threat of a bank run could discipline banks to behave prudently.

## 2.4 Is There Something Good about the SSC?

So far, the liquidity of deposits via the sequential service constraint (SSC) was shown to have drawbacks; that is, it may cause bank runs. Why then does the SSC exist and, more specifically, what could be the positive role of the SSC? This section shows that the SSC may act as a threat that commits banks to monitoring.

Liquidity provision and the sequential service constraint may now play a valuable role.

### 2.4.1 Analysis

I now analyze a situation in which depositors can monitor their banks and pressure them to behave prudently. In particular, depositors may make a run on the bank after obtaining a negative signal, and this may be damaging for the bank. Such a prospect could discipline the bank.

| $t = 0$: | $t = 1$: | $t = 2$: |
|---|---|---|
| ♠ The bank gathers funds from depositors and lends to a borrower. | ♠ The bank can monitor its borrower.<br>♠ Depositors obtain a signal about bank monitoring.<br>♠ Depositors can threaten to run on the bank and force it to invest in a transparent monitoring technology. | ♠ Payoffs are realized. |

Figure 2.3: Sequence of events with a bank run as a threat

I assume that there is no deposit insurance. This is obviously important; otherwise, no depositors would choose to run. Depositors are risk neutral.[6] With probability $\alpha$, all depositors obtain a (costless) signal at $t = 1$. With probability $1 - \alpha$, depositors receive no signal. The signal perfectly reveals whether the bank has monitored its borrower. If the signal shows no monitoring, depositors run on the bank. Calomiris and Kahn (1991) model a bank run as liquidation of the bank. In contrast, I allow for a less punishing role of a bank run. I assume that depositors that run on the bank may be persuaded to leave their funds at the bank.

In particular, the bank can restore the confidence of depositors by making its monitoring technology transparent. In this case, bank monitoring is contractible; however, the transparent monitoring technology is less effective. That is, in the case of the transparent monitoring technology, the success probability $p_T$ of the borrower's project is lower than the success probability of a good project, but higher than the success probability of a bad project; that is, $p_G > p_T > p_B$. Figure 2.3 depicts the sequence of events.

Resolution of a bank run in this way seems to better match reality. More specifically, in the case of a bank run usually the regulator (or acquiring bank) steps in. The regulator may force the troubled bank to become transparent (i.e., to invest in the transparent monitoring technology) such that monitoring of the bank may be easily controlled. This may mean that the bank can only use verifiable hard data in its monitoring technology but has to disregard unverifiable soft information about its borrowers. Hence, the transparent monitoring technology is verifiable but is less valuable for borrowers.

The bank profit, conditional on monitoring, see (2.2), is left unchanged. Conditional on no monitoring, the bank profit is left unchanged as well, see (2.3), if depositors receive no signal. This happens with probability $1 - \alpha$. However, with probability $\alpha$ depositors receive a negative signal. In this case, the bank profit is lower. That is, depositors threaten to run on the bank which forces it to invest in the transparent monitoring technology. In this case, a borrower's project differs from a good project only in its success probability; that is, $p_T$ is lower than $p_G$. Hence, (2.3) becomes

$$\Pi_{NM} = [1 - \alpha]p_B[R - r_D] + \alpha\{p_T[R - r_D] - c_M\}. \tag{2.12}$$

[6] To maintain focus, I do not include random liquidity needs of depositors as analyzed in Section 2.3.

The incentive constraint in (2.4) now becomes (use (2.2) and (2.12))

$$p_G[R - r_D] - c_M \geq [1 - \alpha]p_B[R - r_D] + \alpha\{p_T[R - r_D] - c_M\}. \tag{2.13}$$

Note that, as long as the condition in (2.13) is satisfied, the bank always monitors its borrower. Anticipating this, depositors demand a return $r_D = \frac{r_F}{p_G}$. Combining this with (2.13) establishes that the bank monitors if

$$c_M < \tilde{c}_M \text{ where } \tilde{c}_M \equiv \bar{c}_M + \frac{\alpha}{1-\alpha}[p_G - p_T][R - \frac{r_F}{p_G}], \tag{2.14}$$

where $\bar{c}_M$ is as given in (2.5). I can now show the following result.

**Proposition 2.2.** *Building on a threat of a bank run strictly* increases *the threshold $\tilde{c}_M$ of monitoring costs below which financing is available; that is, $\tilde{c}_M > \bar{c}_M$.*

This result shows that the threat of a bank run also allows the bank to finance a borrower that has relatively high costs of monitoring. Such a borrower is left without funding in the absence of such a bank run threat.

In the model so far, banks always behave prudently (the incentive constraint in (2.13) is satisfied). Depositors never run on the bank even though they could do so. That is, a bank run only acts as a threat but it never materializes. I now enrich the analysis by allowing depositors to observe the monitoring of their bank only at a certain cost $c_D$. I also make the following assumption:

**Assumption 2.1.** *$p_T < 2p_B$.*

Assumption 2.1 guarantees that investing in a transparent monitoring technology is not an attractive option. In particular, it is never so attractive that depositors would automatically (i.e., without observing monitoring of their bank) force the bank to invest in the transparent monitoring technology.

It is easy to see that, for high monitoring costs $c_M > \bar{c}_M$, there exists no Nash equilibrium in pure strategies with monitoring (for low monitoring costs $c_M \leq \bar{c}_M$, Section 2.2.2 shows that the bank always monitors). That is, if depositors always "observed" the bank, the bank would behave prudently and monitor its borrower. Knowing that, depositors would refrain from "observing" the bank to save on $c_D$. However, then anticipating this, the bank would stop monitoring. As a consequence, no pure strategy equilibrium with monitoring would exist.

I can show the following result.

**Proposition 2.3.** *For low monitoring costs (i.e., $c_M \leq \bar{c}_M$), the bank always monitors and there is no bank run. For high monitoring costs (i.e., $c_M > \bar{c}_M$), the following equilibrium in mixed strategies exists. A bank monitors its borrower with probability $\beta$ and it refrains from monitoring its borrower with probability $1 - \beta$. Depositors observe the monitoring of their bank with probability $\alpha$ and do not observe it with probability $1 - \alpha$.*

When depositors invest in observing the monitoring choice of the bank, this not only induces the bank to monitor (Proposition 2.3) but also introduces the possibility of bank runs. I now determine the probability of a bank run. A bank run occurs if a bank does not monitor its borrower (with probability $[1 - \beta]$) and depositors observe this (with probability $\alpha$). Hence, the probability of a bank run is

$$p_{BR} = [1 - \beta]\alpha. \tag{2.15}$$

The proof of Proposition 2.3 also shows that depositors only run if they receive a negative signal. If they receive no signal, they prefer to stay in a bank. That is, Assumption 2.1 guarantees that depositors cannot profit from a bank run without observing bank behavior.

I can now show the following result.

**Corollary 2.3.** *Increasing the cost of bank monitoring, $c_M$, and/or the cost of depositors' observing the bank, $c_D$, increases the probability of a bank run.*

Increasing the cost of bank monitoring, $c_M$, increases the probability of a bank run because the higher costs of bank monitoring negatively affect a bank's incentives to monitor, which induces depositors to invest in observing monitoring of their bank; that is, they "observe" more often ($\alpha$ increases). This increased pressure from depositors restores banks' incentives to monitor. In fact, the bank monitors with the same probability as before (i.e., $\beta$ is independent of $c_M$). Yet, depositors more frequently spot risky behavior of their bank. Hence, a bank run occurs more often.

Increasing the costs to depositors of $c_D$ for observing bank behavior augments the probability of a bank run through two channels. First, depositors may be less willing to "observe" the bank. The bank anticipates this and monitors less. Lower monitoring increases the probability of a bank run. Second, depositors anticipate that the bank monitors less and they demand higher returns. These higher returns induce depositors to "observe" the bank more, which again leads to more bank runs.

Proposition 2.2 above shows that demand deposits can act as a disciplining device, as in Calomiris and Kahn (1991). They argue that demand deposits serve as a threat for managers to behave in the interest of depositors. In particular, in their analysis demand deposits deter managers from absconding with the funds of the bank. More specifically, demand deposits induce depositors to check whether the bank monitors its borrowers. This is valuable because depositors can vote with their feet (i.e., run), and in doing so bring down the bank. Calomiris and Kahn (1991) also show that having a larger number of depositors augments the accuracy of depositors' actions if their signals are imperfectly correlated.

Diamond and Rajan (2000) justify why banks combine monitoring of their borrowers with a liquidity provision to depositors. They claim that it is difficult to lend to the borrower, which is indispensable for the project that he is undertaking. If such a borrower cannot commit to staying with the project, he may hold up the bank. Anticipating this, the bank can only lend a limited amount. The relationship bank may partially solve this problem by gathering specific knowledge related to the project that the borrower undertakes. In this case, if the borrower leaves, the relationship bank may overtake and continue the project instead of the borrower. However, a new problem arises. If struck by a liquidity need, the relationship bank may have difficulties in obtaining funds to the full value of its loan. This is because the relationship bank may now hold up investors due to its specific knowledge related to the project. Diamond and Rajan (2000) show that withdrawable demand deposits prevent this renegotiation problem. In particular, depositors successfully threaten the bank with withdrawal in the case of any renegotiation.[7]

In Chapter 7, I provide another rationale for why banks combine liquid demand deposits with lending to borrowers that need monitoring. Following Jensen (1986), I show that demand deposits may commit banks to monitoring. In particular, withdrawals on demand deposits could force banks to liquidate some of their borrowers early to repay debt and, in doing so, limit debt financing in later periods. This alleviates moral hazard. In addition, demand deposits, contrary to straight short-term and long-term debt, help commit the bank not to overleverage itself and this commits the bank to monitoring. I show that banks prefer demand deposits if the cost of early liquidation of borrowers is intermediate and if bank monitoring is costly.

Proposition 2.3 and Corollary 2.3 explicitly model the occurrence of a bank run in a mixed strategy Nash equilibrium. This allows for the identification of parameters in the economy (i.e., the cost of monitoring the borrower and the cost of "observing" the bank's action) that have an impact on the probability of a bank run. This is related to the approach taken by Goldstein and Pauzner (2005). Whereas Goldstein and Pauzner (2005) as well as Chari and Jagannathan (1988) connect the probability of a bank run to an exogenous event (the state of the economy or a signal about the bank's action), this analysis endogenously determines the probability that depositors "observe" the bank's action, which subsequently affects the probability of a bank run.

Although the threat of a bank run may be viable, the theoretical banking literature is not yet conclusive on the exact mechanism that triggers a bank run. In the following section I present a short inroad to the analysis of this question.

### 2.4.2 Sequential service constraint: Further thoughts

I now present a more in-depth analysis of the role of the SSC. In particular, the SSC and the observed pattern of withdrawals provide for a type of information aggregation among depositors. This is important because this type of informational aggregation may on the one hand be inefficient and induce bank runs, yet on the other hand may be efficient in resolving free-riding in the monitoring of the banking depositors.

Instructive is the bank run on the UK bank Northern Rock that occurred in September

[7]Despite the benefit of demand deposits, Diamond and Rajan (2001) show that a level of capital is valuable because it limits bankruptcy costs. Therefore, a bank selects its optimal capital structure by weighting the benefits of liquidity creation and the enforcement of borrowers' repayment with the costs of bankruptcy.

2007. The Treasury Committee in its report (House of Commons (2008)) analyzes the causes and consequences of the run on Northern Rock. Interestingly, the lending side of Northern Rock before the run seemed quite solid. The regulator (Financial Service Authority) also allowed Northern Rock to adopt the advanced approach of risk measurement under the Basel II capital accord. As Mervyn King, the Governor of Bank of England stated: "What I would say about Northern Rock (and this is the tragedy of Northern Rock) is that most of the staff [...] did an excellent job in appraising the loans that they were making, and that they monitored very carefully and they did not lend money to people who should not be borrowing from them. The lending side was handled extremely well."

The main cause of a bank run on Northern Rock was the improper structure of its liabilities. Northern Rock was heavily financed on the capital markets, with a substantial maturity mismatch. It was the credit squeeze on capital markets that drove Northern Rock illiquid. When it became public that Northern Rock had liquidity problems and that it had to arrange a support facility at the Bank of England, a bank run was triggered. One interpretation is that information about the unwillingness of informed investors to provide financing triggered the bank run of less informed, small depositors. This is related to the work of Ó Gráda and White (2002) who analyze bank runs in 1854 and 1857 and argue that the bank run triggered by informed market participants in 1857 had much graver consequences than the bank run triggered by less wealthy, less experience – uninformed – depositors in 1854. My analysis addresses the following questions that these examples provoke: *i*) what can increase the precision of a bank run; and *ii*) how can a bank run be halted?

I model a sequential withdrawal of depositors as an informational cascade following Bikhchandani, Hirshleifer, and Welch (1992). In particular, each depositor, in addition to monitoring the bank on his own, also observes the actions of depositors in the line in front of him. These actions reveal the information of other depositors and may trigger an informational cascade. An informational cascade denotes a situation in which a depositor may decide on his action (i.e., to withdraw or not) entirely on the basis of the actions of his predecessors regardless of his own signal. This analysis points to the main drawback of the sequential service constraint: depositors may optimally choose to disregard their own information and in doing so trigger bank runs even though the bank is sound. This could make information aggregation inefficient.

By modeling the withdrawal of deposits as a cascade, I obtain a few new insights. I show that the sequential service constraint allows for cheaper monitoring of a bank. That is, it suffices for only a few depositors to monitor the bank, and the SSC makes it optimal for a few to monitor and hence resolves the free-rider problem among depositors. Others may then simply follow their actions. Thus monitoring by a few is enough to provide incentives to the bank to behave safely. Moreover, monitoring by a few also presents the drawback of the sequential service constraint: inefficient aggregation of information via cascades. In particular, depositors may rely completely on the actions of their predecessors, disregarding their own information about the quality of the bank.

Interestingly, I show that borrowing from depositors with an exogenous liquidity need may limit the loss of information due to the cascading effect. That is, a liquidity need lowers the information content of the withdrawals of depositors early in line. More specifically, if the depositor at the front of the line withdraws, this might be either because he has obtained a negative signal about the bank's quality, or because this depositor simply has a liquidity need. Thus, liquidity needs reduce the "stigma fraction" of withdrawals. This partially limits the herding behavior of depositors and improves the precision of bank runs.

Several contributions aim to explain bank runs as a consequence of an informational cascade. Yorulmazer (2003) applies the same model to explain the occurrence of bank runs on safe banks. He shows that completely eliminating bank runs may create costs. In particular, the demand-deposit contract that eliminates bank runs must discourage late depositors from withdrawing early even if they obtain a negative signal about the bank's quality. This means that short-term returns must be sufficiently small. Although this may prevent bank runs, it limits the intertemporal insurance function of the bank. More specifically, an intertemporally averse depositor values the most demand-deposit contracts of comparable early and late returns. In the case of a run-proof demand-deposit contract, however, the early return has to be substantially smaller than the late return.

The main difference between Yorulmazer (2003) and this analysis is in the observability of the type of depositors. In Yorulmazer (2003), depositors may observe whether depositors in line are late or early depositors. In my model, types of depositors are not observed. Hence, it is not known whether depositors withdraw due to their liquidity needs or due to the signal about the bank's quality. This contributes some noise to the informational cascade.

The closest paper to my analysis is Gu (2007). In his model bank runs are similarly triggered by the herding behavior of depositors. In particular, a depositor's decision to withdraw depends on both his signal about the bank's quality and the withdrawal decisions of other depositors. He shows that deposit contracts that allow for bank runs can be optimal because the bad bank might be liquidated before it invests in bad projects. My analysis generalizes Gu (2007) by showing that withdrawals motivated by liquidity needs could "confuse" the inference of depositors from previous withdrawals and be beneficial. In particular, when liquidity needs are common, depositors put a higher weight on their own signal about the bank's quality and less eagerly draw inferences from withdrawals by other depositors. This partially limits herding and could improve the precision of bank runs.

The observation that withdrawals based on liquidity needs may actually improve the precision of a bank run is in contrast with the existing literature in which a bank run occurs as a consequence of (excessive) simultaneous withdrawals by uninformed depositors. Diamond and Dybvig (1983) argue that liquidity needs may trigger a sunspot bank run. Chari and Jagannathan (1988) consider both informed and sunspot bank runs. They show that withdrawals due to liquidity needs might be mistakenly considered a consequence of negative information about the bank and may trigger a run of uninformed depositors. In their model, however, uninformed depositors obtain no signal whereas informed depositors perfectly observe bank monitoring. Hence, there is no role for beneficial pooling of information that my analysis predicts. That is, I consider a setting where depositors' signals are not perfectly informative; hence, depositors benefit from the informational content of the history of withdrawals.

Chen (1999) stresses that the first-come, first-serve rule that the SSC contains creates a negative payoff externality among depositors. In particular, if a depositor withdraws after everybody else has already withdrawn, he may receive very small returns. In this model, informed depositors can withdraw early based on their signals about the quality of the bank. Uninformed depositors cannot evaluate the quality of the bank directly. To limit the negative payoff externality, they nevertheless withdraw soon based on other less precise measures, such as bankruptcy of other banks. Consequently, bank runs may be triggered by noisier information of uninformed depositors and this is socially costly. Chen (1999) proposes that deposit insurance should cover only losses to uninformed depositors but not losses to informed depositors. In this analysis, noise in the form of a potential liquidity need does not distort the signal of each depositor. It distorts information about the actions of the other depositors. In this case, noise is beneficial and it increases the precision of bank runs and may therefore be valuable.[8]

I proceed as follows. First, I analyze the case in which depositors have no liquidity needs. Hence, they withdraw only on the basis of their signal. Second, I focus on the case with a liquidity need. In particular, if a depositor is struck by a liquidity need, he must withdraw regardless of his signal.

The following stylized model captures the key characteristics of the sequential service constraint. The bank may be either good or bad. The bank is good if it monitors its borrowers. In this case, it is optimal for depositors to keep their deposits in the bank. The bank is bad if it does not monitor its borrowers. In this case it is optimal for depositors to prematurely withdraw their deposits. However, depositors cannot directly observe the quality of the bank. Each depositor obtains a costless signal about the bank's quality. More specifically, the signal is symmetric and informative in the sense that a depositor obtains a positive signal with probability $\Phi > \frac{1}{2}$ and a negative signal with probability $1 - \Phi$ if the bank is good. If the bank is bad, a depositor obtains a negative signal with probability $\Phi$ and a positive signal with probability $1 - \Phi$. Hence, a depositor can only imprecisely anticipate a bank's quality from his signal. The signals that depositors receive are independent.[9] Additionally, a depositor observes withdrawals (or no withdrawals) of his predecessors. This gives him additional but imprecise information about the signals about

[8]This relates to the observation of the theory of informational cascades that it is more beneficial for society if agents rely on their own information than if they follow the actions of others; see Bikhchandani, Hirshleifer, and Welch (1992) and Bikhchandani, Hirshleifer, and Welch (1998).

[9]Kelly and Ó Gráda (2000) argue that the rumors about bank stability spread through social networks. Using correlated instead of independent signals of subsequent depositors would allow for this effect. However, this extension is left for future work. Iyer and Puri (2007) stress that not only social networks are important but also the strength of the bank-depositor relationship. This suggests that depositors with more precise information about their bank run less willingly.

the bank's quality that other depositors have received.

This setting allows for the occurrence of several informational cascades. For example, an informational cascade starts if the first two depositors do not withdraw. In this case, the subsequent depositors do not withdraw regardless of their signals. A different informational cascade starts if the first two depositors obtain a negative signal and withdraw. In this case, the third depositor observing these withdrawals withdraws irrespective of his own signal. All the subsequent depositors withdraw as well and an informational cascade starts. In this case, an informational cascade resembles a bank run. That is, one of the key characteristics of a bank run is that depositors withdraw simply because of the long line in front of the bank and regardless of their own knowledge about the bank's quality.

Representing a bank run as an informational cascade points to the following problem. Bank runs may not necessarily happen to bad banks and they may also happen to good banks. I now compute the probability that a bank run occurs if the bank is good. First, I compute the probability that the first two depositors in line withdraw. With probability $[1-\Phi]^2$, both depositors mistakenly receive a negative signal. With probability $\Phi[1-\Phi]$, the first depositor receives a positive signal and the second depositor receives a negative signal. In this case, the second depositor anticipates that the first depositor has received a positive signal. Due to his negative signal, he cannot make any assessment of the quality of the bank. Hence, he flips the coin and withdraws with probability $\frac{1}{2}$. That is, both depositors mistakenly withdraw if

$$\Phi_M \equiv [1-\Phi]^2 + \frac{\Phi}{2}[1-\Phi]. \tag{2.16}$$

Interestingly, if the first two depositors withdraw, the third depositor and all subsequent depositors withdraw regardless of their signals. That is, even though the third depositor obtains a positive signal, he anticipates that the first depositor has obtained a negative signal. Because the second depositor has also withdrawn, the third depositors anticipates that the bank is more likely bad than good and withdraws. Subsequent depositors withdraw as well.

Now I compute the probability of a bank run on a good bank. The bank run occurs if the first two depositors withdraw. If the first two depositors correctly leave their funds in the bank, which happens with probability

$$\Phi_C \equiv \Phi^2 + \frac{\Phi}{2}[1-\Phi], \tag{2.17}$$

the bank run does not occur. Moreover, in this case, all the subsequent depositors leave their funds in the bank. However, if only one of the two depositors, withdraws which happens with probability

$$\Phi_{MC} \equiv \Phi[1-\Phi], \tag{2.18}$$

the cascade does not start (see Table 2.1). The third depositor gains no additional information from the actions of the first two. Hence, the whole game starts from scratch; in particular, the bank run starts with probability $\Phi_{WW}$ if the third and fourth depositors

| | Good bank | Bad bank |
|---|---|---|
| Both stay → All stay | $\Phi_C$ | $\Phi_M$ |
| One stays, one withdraws | $\Phi_{MC}$ | $\Phi_{MC}$ |
| Both withdraw → Bank run | $\Phi_M$ | $\Phi_C$ |

Table 2.1: The probabilities of various actions that the first two depositors may take (to withdraw or stay in the bank) in the absence a liquidity need

withdraw. In sum, the probability that a bank run on a good bank starts is

$$Pr(BR,G) = \Phi_M + \Phi_{MC}\Phi_M + \Phi_{MC}^2\Phi_M + \ldots = \frac{\Phi_M}{1-\Phi_{MC}}. \tag{2.19}$$

Inserting (2.16) and (2.18) into (2.19) yields

$$Pr(BR,G) = \frac{[2-\Phi][1-\Phi]}{2[1-\Phi+\Phi^2]}. \tag{2.20}$$

I compute the probability of a bank run on a bad bank as follows. A bank run on a bad bank occurs if the first two depositors correctly withdraw, which occurs with probability $\Phi_C$. A bank run may also occur if only one of the first two depositors withdraws (with probability $\Phi_{MC}$). In this case, the third and fourth depositors again rely only on their signals. A bank run on a bad bank occurs with probability

$$Pr(BR,B) = \Phi_C + \Phi_{MC}\Phi_C + \Phi_{MC}^2\Phi_C + \ldots = \frac{\Phi_C}{1-\Phi_{MC}}. \tag{2.21}$$

Inserting (2.17) and (2.18) into (2.21) yields

$$Pr(BR,B) = \frac{\Phi^2 + \frac{\Phi}{2}[1-\Phi]}{1-\Phi[1-\Phi]}. \tag{2.22}$$

Now the main part of the analysis of the SSC starts. I allow for a liquidity need. That is, with small probability $\epsilon \to 0$ a depositor may have a liquidity need. Conditional on having a liquidity need, a depositor must withdraw (early with probability $\frac{1}{2}$ or late with probability $\frac{1}{2}$) regardless of his signal. Such a liquidity need introduces noise into the information content of withdrawals. In particular, a depositor cannot anticipate with certainty that the withdrawals are triggered by negative signals. They may be triggered by the liquidity needs of depositors.

Now I compute the probability of a bank run on a good bank in the presence of a liquidity need. In the presence of a liquidity need, the first two depositors withdraw less often than without a liquidity need (see (2.16)). That is, the first two depositors only withdraw if they both mistakenly receive negative signals, which happens with probability $[1-\Phi]^2$. Upon observing the first depositor to withdraw, the second depositor may stay in the bank if he receives a positive signal. This is because the second depositor relies more on his own signal knowing that the withdrawal of the first depositor may well be liquidity-driven. Hence, the probability that the first two depositors mistakenly withdraw is lower than (2.16); that is,

$$\bar{\Phi}_M \equiv [1-\Phi]^2. \tag{2.23}$$

If the first two depositors withdraw, the third depositor and all subsequent depositors withdraw regardless of their signals and a bank run starts.

The bank run occurs if the first two depositors withdraw. If the first two depositors leave their funds in the bank, a bank run does not occur. This happens only if they both correctly receive a positive signal; that is, with the probability

$$\bar{\Phi}_C \equiv \Phi^2. \tag{2.24}$$

However, if only one of the two depositors withdraws, which happens with probability

$$\bar{\Phi}_{MC} \equiv 2\Phi[1-\Phi], \tag{2.25}$$

the cascade does not start (see Table 2.2). The third depositor gains no additional information from the actions of the first two. Hence, the whole game starts from scratch. Thus, the probability that a bank run on a good bank occurs in the presence of a liquidity need is

$$\bar{Pr}(BR,G) = \bar{\Phi}_M + \bar{\Phi}_{MC}\bar{\Phi}_M + \bar{\Phi}_{MC}^2\bar{\Phi}_M + \ldots = \frac{\bar{\Phi}_M}{1-\bar{\Phi}_{MC}}. \tag{2.26}$$

Inserting (2.23) and (2.25) into (2.26) yields the probability of a bank run on a good bank

$$\bar{Pr}(BR,G) = \frac{[1-\Phi]^2}{1-2\Phi[1-\Phi]}. \tag{2.27}$$

In the same way, I compute the probability of a bank run on a bad bank

$$\bar{Pr}(BR,B) = \bar{\Phi}_C + \bar{\Phi}_{MC}\bar{\Phi}_C + \bar{\Phi}_{MC}^2\bar{\Phi}_C + \ldots = \frac{\bar{\Phi}_C}{1-\bar{\Phi}_{MC}}. \tag{2.28}$$

Inserting (2.24) and (2.25) into (2.28) yields

$$\bar{Pr}(BR,B) = \frac{\Phi^2}{1-2\Phi[1-\Phi]}. \tag{2.29}$$

I can now show the following result.

**Proposition 2.4.** *The presence of a liquidity need lowers the probability of a bank run on a good bank and increases the probability of a bank run on a bad bank.*

The intuition for Proposition 2.4 is the following. Adding noise – a probability of a liquidity need – lowers the information content of the actions of previous depositors. In particular, depositors cannot observe whether previous withdrawals are liquidity- or information-driven.

| | Good bank | Bad bank |
|---|---|---|
| Both stay → All stay | $\bar{\Phi}_C$ | $\bar{\Phi}_M$ |
| One stays, one withdraws | $\bar{\Phi}_{MC}$ | $\bar{\Phi}_{MC}$ |
| Both withdraw → Bank run | $\bar{\Phi}_M$ | $\bar{\Phi}_C$ |

Table 2.2: The probabilities of various actions that the first two depositors may take (to withdraw or stay in the bank) in the presence of a liquidity need

This induces each depositor to rely more on his own signal and leads to less herding and consequently to better information aggregation. Figure 2.4 graphically shows that introducing noise into the actions of depositors reduces the probability of a bank run on a good bank and increases the probability of a bank run on a bad bank.

Observe that the probabilities of a bank run in the case of a liquidity need given in (2.26) and (2.29) do not depend on the size of the liquidity need $\epsilon$. This is because I have assumed that a liquidity need occurs with a small probability ($\epsilon \to 0$). Alternatively, I could allow for a higher probability of a liquidity need (higher $\epsilon$). In this case, the information content of previous withdrawals may even be lower and each depositor would rely more on his own signal. Consequently, a bank run would start only after several withdrawals occur in a row.

Interestingly, a release of a small amount of new information may halt a bank run. Note that if information released contains the information content of two subsequent withdrawals, the subsequent depositor again follows his own signal. This shows that purely sunspot bank runs might be stopped by early intervention. A release of new information or backing by more informative investors suffices. Ó Gráda and White (2002) confirm this by showing that the sunspot bank run in 1854 was successfully stopped by intervention of more informed investors.

Although a liquidity need improves the precision of a bank run, it also introduces a drawback. That is, with a liquidity need the bank run starts later. In particular, more depositors should observe the bank before a bank run starts. This is costlier and may create a free-rider problem if observing the bank is costly. However, this would make the analysis more complex and it is left for further work.

The main implication of the analysis in this section is the following. Recall that, with the possibility of a liquidity shock, withdrawals may occur because of negative signals about the quality of the bank *or* due to liquidity needs of depositors. Such noise decreases the probability of a bank run on a good bank and increases the probability of a bank run on a bad bank. This shows that the sequential service constraint induces better aggregation of information if depositors may be hit by an exogenous liquidity shock.

Despite this potential effectiveness of "depository discipline," the prevalence of deposit insurance poses some questions regarding the relevance of this mechanism. In the following section, capital and capital regulation are introduced as potential alternative (and more realistic) mechanisms.

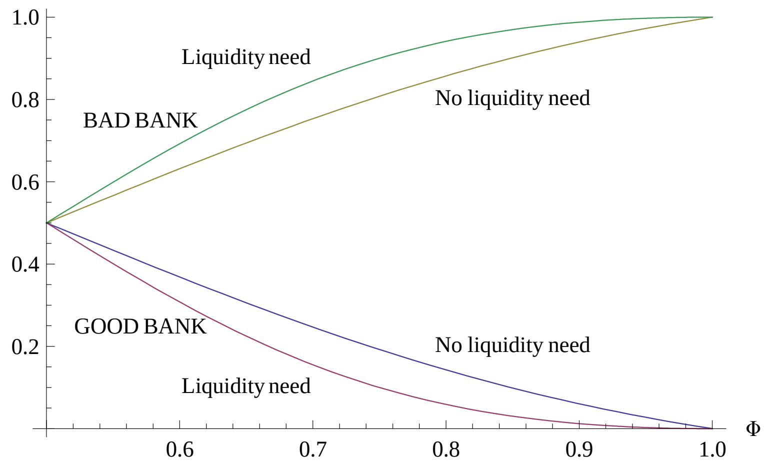


Figure 2.4: The probability of a bank run on a good bank (and on a bad bank) as a function of precision of a signal (in the case of a liquidity need and in the absence of a liquidity need)

## 2.5 Bank Capital

In this section I focus on an alternative mechanism that commits banks to monitoring. Banks can commit to monitoring by raising sufficient amounts of capital. This suggests that it could be privately optimal for a bank to be well capitalized. In this context I also discuss the recent work of Allen, Carletti, and Marquez (2007), which analyzes precisely this issue. An important caveat is that banks may not internalize the cost of a bank run sufficiently. In this case, capital regulation may be necessary.

### 2.5.1 Bank capital

Now I analyze the monitoring incentives of a bank that is financed with deposits and bank capital. I follow the work of Allen, Carletti, and Marquez (2007) to show that banks may voluntarily hold capital. In particular, capital induces banks to monitor their loans and hence credibly communicates to borrowers that banks will monitor. This is valuable for borrowers and hence can improve the profitability of the banks. I connect the level of bank capital (i.e., the level of capital) with the strength of the borrower (i.e., the size of the collateral).

I now expand the model that was introduced in Section 2.2.1 along the lines of Holmstrom and Tirole (1997) in two directions. First, I allow for bank capital. The bank collects the proportion $k$ of the total funds needed from the providers of bank capital at a cost $r_E$ per unit. The proportion $[1-k]$ it collects from depositors at a cost $r_D$. Capital is more costly than deposits; that is, $r_E > r_D$.[10] Second, I denote the strength of the borrower by the size of his collateral $A$, where $A < r_F$. That is, if the borrower fails, the bank and in particular

[10]See Holmstrom and Tirole (1997) and Diamond and Rajan (2000) for explicit models of why the cost of capital might be higher than the cost of deposits. Note that this assumption bypasses the question *how* capital is raised, including potential adverse selection problems, as in Myers and Majluf (1984), or moral hazard problems, as in Thakor (1990).

| $t = 0$: | $t = 1$: | $t = 2$: |
|---|---|---|
| ♠ The bank gathers funds from depositors and capital providers and lends to a borrower. | ♠ The bank can monitor its borrower.<br>♠ Depositors obtain a signal about bank monitoring.<br>♠ Depositors can run on the bank and force it to invest in a transparent monitoring technology. | ♠ Payoffs are realized.<br>♠ If necessary depositors seize a collateral. |

Figure 2.5: Sequence of events accounting for bank capital and collateral

its depositors are still able to size at least $A$. See Figure 2.5.

If depositors anticipate bank monitoring, they demand

$$p_H r_D + [1 - p_H]A = r_F,$$

which yields

$$r_D = \frac{r_F - [1 - p_G]A}{p_G}. \tag{2.30}$$

Because the bank finances the investment with $1 - k$ deposits, it pays out to depositors $r_D[1 - k]$ in the case of success. It also incurs the cost of capital financing $kr_E$.[11] The expected profit of the bank conditional on monitoring its borrower is

$$\Pi_M = p_G\{R - r_D[1 - k]\} - c_M - kr_E. \tag{2.31}$$

The expected profit of the bank conditional on not monitoring its borrower is

$$\Pi_{NM} = p_B\{R - r_D[1 - k]\} - kr_E, \tag{2.32}$$

where I have assumed that depositors anticipate that the bank will monitor. The bank monitors if (2.31) is higher than (2.32). That is,

$$p_G\{R - r_D[1 - k]\} - c_M - kr_E \geq p_B\{R - r_D[1 - k]\} - kr_E. \tag{2.33}$$

I can now show the following result.

**Proposition 2.5.** *A borrower with high collateral (i.e., $A > \bar{A}$) can borrow from banks. A borrower with low collateral (i.e., $A \leq \bar{A}$) cannot obtain a bank loan, where*

$$\bar{A} \equiv \frac{r_F}{1 - p_G} - \frac{p_G}{[1 - p_G][1 - k]}[R - \frac{c_M}{\Delta p}]. \tag{2.34}$$

The intuition for Proposition 2.5 is the following. On the one hand, if the borrower with low collateral fails, the bank and its depositors receive only low returns from his collateral. Anticipating this, depositors demand a relatively high deposit rate. However, a high deposit

[11] Note that $kr_E$ reflects the cost of outside equity.

rate lowers the bank profit and consequently the bank cannot commit to monitoring. Hence, the borrower with low collateral cannot obtain a bank loan. On the other hand, if the borrower with high collateral fails, the bank and depositors still receive decent returns. Consequently, depositors demand low returns and the bank anticipating high profits can commit to monitoring.

Proposition 2.5 partially replicates Holmstrom and Tirole (1997). Holmstrom and Tirole (1997) further extend the analysis to include the possibility of issuing direct debt. They show that undercapitalized borrowers (i.e., those with low levels of collateral) cannot obtain finance at all, whereas reasonably capitalized borrowers (i.e., those with intermediate levels of collateral) borrow from banks and well-capitalized borrowers can issue direct debt. In my model above, I have precluded the option of direct debt by assuming that a borrower with direct debt would always decide on a bad project due to his high private benefits. If the level of private benefits is sufficiently small, a borrower with high collateral could issue direct debt, which would replicate the result of Holmstrom and Tirole (1997).

I can now show the following proposition.

**Proposition 2.6.** *A better capitalized bank can lend to a borrower with lower collateral (i.e., $\frac{\partial \bar{A}}{\partial k} < 0$). A borrower with low collateral (i.e., $A < \bar{A}(k=0)$) can only borrow from a bank with the minimum level of capital*

$$k^* = 1 - \frac{p_G[R - \frac{c_M}{\Delta p}]}{r_F - [1 - p_G]A}. \tag{2.35}$$

Proposition 2.6 shows that the level of capital of the bank is important if the bank wishes to finance a borrower with low levels of collateral. This is because a well-capitalized bank has its own capital at stake and this induces it to monitor even a borrower with low levels of collateral.

The next corollary shows that interbank competition affects the minimum level of bank capital.

**Corollary 2.4.** *The bank needs more capital if competition for deposits increases (i.e., $\frac{\partial k^*}{\partial r_F} > 0$).*

The intuition for Corollary 2.4 stems from the negative effect of competition on bank rents. Because competition undermines the bank's expected profits, the bank is less incentivized to monitor its borrower. The only way that it can reestablish monitoring incentives and attract depositors is by holding more capital.

Proposition 2.6 and Corollary 2.4 follow the work of Allen, Carletti, and Marquez (2007) by showing that banks may choose to voluntarily hold positive levels of capital and that these positive levels of capital increase with competition. However, they go even further in their analysis. They show that banks may even hold more capital than is socially optimal if competition between banks is high. This is because banks use capital to commit to monitoring in order to better compete for borrowers. In particular, monitoring is valuable for borrowers because it increases borrowers' success probability. If competition for borrowers is high, banks would like to commit to high levels of monitoring (as Corollary 2.4 also shows). However, higher competition lowers bank rents and makes it increasingly difficult for banks to commit to monitoring. This is why banks use high levels of capital. Bank capital commits them to high monitoring despite strong competition.

The empirical literature confirms that banks often hold considerably higher levels of capital than prescribed by the regulator (see Flannery and Rangan (2004) and Calem and Rob (1999)).

Boot and Greenbaum (1993) and Gehrig and Jost (1995) show that banks may use not only financial capital, but also reputational capital, to enforce credibility in their own monitoring. Boot and Greenbaum (1993) show that reputation allows banks to lower their market-determined funding costs. Such reputational benefits encourage banks to monitor and are especially important if competition between banks is high. They also show that deposit insurance fixes future funding costs for banks and, in doing so, destroys the funding-related benefits of reputation.

Above I have explained the incentive role of bank capital in which bank capital serves as an incentive mechanism for bank monitoring. Calem and Rob (1999) point to another role of capital. In their view, capital acts as a buffer against bank losses. More specifically, banks hold excessive capital to assure that they satisfy the capital requirements demanded by the regulator even if they are hit by a solvency shock. In the intertemporal model, they show that the relation between bank capital and its risk is a $U$-shaped curve. Banks with the lowest levels of capital are the riskiest. Then bank risk declines with bank capital. At higher levels of capital, however, bank risk starts increasing again. Calem and Rob (1999) also point to the influence of market pressure (i.e., pricing of the uninsured depositors) on prudent bank behavior.

Peura and Keppo (2006) assume that banks can only recapitalize at a delay. If the losses exceed the capital buffer, banks may be driven to liquidation. Their model, calibrated to the data of U.S. banks, reasonably well predicts why banks hold excessive capital.

### 2.5.2 Bank capital and bank runs

Now I combine bank capital with the possibility of a bank run. I observe how bank capital affects the possibility of a bank run. I modify the model in the previous section by introducing the possibility that depositors at a cost $c_D$ observe whether a bank monitors its borrower (see Section 2.4). For simplicity, I set the borrower's collateral to zero.

The bank now has two options. It may raise sufficient capital and always monitor its borrower. In this case, depositors are safe and they do not need to observe a bank's behavior. Alternatively, the bank may hold lower levels of capital. In this case, Proposition 2.3 shows that, when depositors observe the bank at cost $c_D$, the bank only monitors its borrower with certain probability. However, this then may lead to a bank run.

I can now show the following.

**Proposition 2.7.** *If capital is not too costly (i.e., $r_E < \bar{r}_E$), the bank raises $k^*$ capital and monitors. If capital is costly, however, (i.e., $r_E \geq \bar{r}_E$), the bank finances entirely with deposits and follows a random monitoring strategy, where*

$$\bar{r}_E \equiv r_F + \frac{c_D[p_G - p_B]}{p_T - p_B}. \tag{2.36}$$

Proposition 2.7 shows that bank capital – as discussed by Allen, Carletti, and Marquez (2007) – and the threat of a bank run as discussed by Calomiris and Kahn (1991) act as complementary mechanisms that commit the bank to monitoring. The bank decides on the cheaper one; that is, the bank weights the cost of capital with the cost of a potential bank run. If capital is relatively costly, the bank uses as little capital as possible. However, if a bank run becomes more probable, the expected cost of a bank run increases and the bank may prefer to raise capital.

Proposition 2.7 also connects the characteristics of depositors with the level of bank capital (note that (2.36) is increasing in $c_D$). The empirical prediction is the following. Banks use less capital if they finance from depositors with a low cost of observing the bank. That is, banks that use a lot of interbank lending (where the cost of observing is low) would stick to the regulatory-prescribed capital requirements. However, banks may use capital significantly above the regulatory-prescribed levels of capital if they finance from depositors with a high cost of observing the bank.

Several contributions in the literature jointly describe and compare the use of bank capital and demand deposits. Diamond and Rajan (2000) claim that, if borrowers' returns are certain, demand deposits are an optimal contract because they prevent borrowers and a relationship bank from holding up its depositors. However, if borrowers' returns are uncertain, a bank may become insolvent, which would trigger a bank run and costly fire sales. Capital is then seen as a buffer against potential losses. Banks set their capital by trading the lower cost of bankruptcy with lower liquidity creation, and the lower ability to force a borrower's repayment.

Dowd (2000) points to the tradeoff between liquidity provision of demand deposits and additional stability that bank capital provides. He expands Diamond and Dybvig's model. In addition to depositors, he allows for agents that provide bank capital. He shows that banks would choose capital on their own to cushion losses and prevent bankruptcy.

Chapter 3 shows that higher levels of capital are value enhancing for the banking industry as a whole because they weed out inefficient banks and make room for efficient banks. Interestingly, banks may want the regulator to impose higher than socially optimal levels of capital. This is because high capital requirements by weeding out inefficient banks may provide more space for efficient banks and therefore may be preferred by the average bank.

### 2.5.3 Capital regulation

A bank run may be costly for the economy as a whole (see Section 2.3.3). For example, a bank run may spread to other banks. Considering Proposition 2.7, this points to the following problem. A bank maximizing its own profit may not fully consider this externality of a bank run. Hence, banks may on average raise too little capital compared to the welfare optimal value. Consequently, the regulator may have an incentive to impose capital requirements on banks in order to limit the probability of a bank run. Observe from Proposition 2.6 that banks will always monitor (and hence be safe) if the regulator demands that banks raise more capital than $k^*$. In this case, each bank has sufficient capital at stake. This induces it to high levels of monitoring with a positive impact on bank stability.

The literature offers a highly divergent view on the influence of capital regulation on banks' risk-taking and on the overall risk of the banking system (for a review, see Santos (2001) and VanHoose (2006)). In the approach in which banks are considered as pure portfolio managers, Kim and Santomero (1988) show that only risk-adjusted capital regulation is effective. If capital requirements are independent of risk, a bank replaces safe assets with risky ones. In addition, the analysis by Kim and Santomero (1988) holds only if a bank's losses are not constrained by a limited liability clause. Rochet (1992) appropriately considers the limited liability of a bank. He shows that, for low levels of capital, the bank relies on its limited liability and chooses a portfolio with maximum risk and minimum diversification. As a response, Rochet (1992) suggests an additional regulation of a minimum level of capital regardless of the risk and the size of a bank.

Diamond and Rajan (2001) show that an increase in capital requirements may force the bank to liquidate some borrowers, which may increase a bank's risk. It may also result in spillover effects among borrowers. In particular, the liquidation threat that higher capital requirements introduce may result in a credit crunch for a cash-poor borrowers and smaller loan repayments for cash-rich borrowers.

A study that points to a detrimental intertemporal effect of tightening capital requirements is Blum (1999). He shows that a bank that expects an increase in capital requirements in the future might have an incentive to immediately become riskier. This is because a unit of capital is worth more if capital requirements are tight. One way of increasing capital in the future is to boost profits by immediately adopting riskier strategies. Besanko and Kanatas (1996) show that increasing capital requirements augments the proportion of outside ownership and lowers the proportion of inside (managerial) ownership. Consequently, managers employ less effort in assuring bank stability and banks may become less stable.

Chapter 3 allows for the heterogeneity of banks. Banks differ in the qualities of their monitoring technologies. I analyze more in depth the linkages between competition and capital regulation. In particular, I show that competition may increase the effectiveness of capital regulation for high quality banks. In particular, high-quality banks would respond more to increases in capital requirements. I highlight the potential danger of capital regulation. Capital regulation may be less effective when needed the most: that is, for low-quality banks. Low-quality banks would raise monitoring only a little as a response to an increase in capital requirements.

Another potential danger of capital requirements may be adverse macroeconomic effects. Blum and Hellwig (1995) prove that capital adequacy regulation may act procyclically. That is, in the case of an economic downturn, firms often renege on servicing their debt obligations. Banks suffer losses. In order to fulfill capital requirements, they have to cut back on lending, which may further depress the economy. Risk-adjusted capital regulation (such as that proposed by the Basel II accord) may act even more procyclically, because bank risk increases in economic downturns (see Borio (2003) and Gordy and Howells (2006)).

### 2.5.4 Deposit insurance reconsidered

In spite of some exceptions, capital regulation generally has a positive effect on bank monitoring and reduces the probability of bank runs. Deposit insurance, however, may work in the other direction.

In Section 2.3.3, I argued that bank runs may be prevented by the introduction of deposit insurance. Now I highlight the drawback that deposit insurance has on banks' incentives to behave prudently. In particular, I extend the model of an information-based bank run to incorporate the effect of deposit insurance. I show that deposit insurance undermines banks' monitoring incentives. Interestingly, I show that deposit insurance contains the probability of a bank run if a borrower is risky on average, but increases the probability of a bank run if a borrower is safe on average.

I consider a situation in which the regulator repays a proportion of $\delta$ promised returns in the case of a bank failure. Hence, depositors receive their promised return with probability $\bar{p}_G = p_G + [1 - p_G]\delta$ in the case of a good project and $\bar{p}_B = p_B + [1 - p_B]\delta$ in the case of a bad project. A run on a bank (that forces it into investment in a transparent monitoring technology) also brings $\{p_T + [1 - p_T]\delta\}r_D \equiv \bar{p}_T r_D$ instead of $p_T r_D$ to depositors.

I can now show the following result.

**Lemma 2.1.** *More generous deposit insurance increases a bank's risk (i.e.,* $\frac{\partial \beta}{\partial \delta} < 0$*).*

The intuition for Lemma 2.1 stems from the impact of deposit insurance on depositors. Insured depositors have lower incentives to observe the bank. The bank anticipates this and takes advantage by lowering the level of monitoring.

This lemma builds on the huge body of literature that describes the pro-risk effect of deposit insurance. Merton (1977) assesses the value of deposit insurance for a bank using option valuation methodology. In particular, he shows an isomorphic structure between deposit insurance and the put option on the bank asset value with a strike price equal to the promised maturity value of bank debt. He predicts that deposit insurance is more valuable for a bank the riskier the bank is, assuming that the deposit insurance premium is insensitive to a bank's risk. Consequently, under flat-rate deposit insurance, banks may be inclined to follow risky strategies.

In Chapter 3 I point to the cross-subsidization effect of deposit insurance. In particular, in the presence of deposit insurance, which makes the deposit rate insensible to a bank's risk, good banks subsidize bad banks. I show that the introduction of higher capital requirements limits the distortion of deposit insurance.

A direct solution to the distortions of deposit insurance seems to be to adjust the deposit insurance premium to a bank's risk. However, Chan, Greenbaum, and Thakor (1992) show that fairly priced deposit insurance cannot be implemented in a competitive banking system due to information problems. Fairly priced deposit insurance can only be implemented if banks are subsidized through deposit insurance.

Gropp and Vesala (2004), however, present a rather positive view of deposit insurance. They claim that a bank might behave more safely in the presence of carefully designed explicit deposit insurance. They argue that deposit insurance that is limited only to depositors commits the regulator not to bail out other creditors, such as large creditors or holders of subordinated debt.

Now I explore the impact of deposit insurance on the probability of a bank run.

**Proposition 2.8.** *Increasing the coverage of deposit insurance augments the probability of a bank run if a borrower is safe on average (i.e., for $p_G > \bar{p}_G$), but lowers the probability of a bank run if a borrower is risky on average (i.e., for $p_G \leq \bar{p}_G$).*

Proposition 2.8 stems from two effects of deposit insurance. First, deposit insurance increases bank risk taking; that is, banks monitor less. If a borrower is safe on average, this effect prevails and increasing the coverage of deposit insurance negatively affects the probability of a bank run. However, the opposite effect of deposit insurance is at work as well. Deposit insurance guarantees the promised returns to depositors. Consequently, depositors have no incentives to monitor their bank. However, then they cannot spot risky bank behavior and there are no bank runs. This opposite effect prevails if a borrower is risky on average. In this case, depositors gain a lot when deposit insurance is introduced and it is less valuable to make a run on a bank.

Proposition 2.8 shows that deposit insurance may be detrimental for the stability of a well developed banking system (in which borrowers are safe on average) but may increase the stability of a less-developed banking system (in which borrowers are risky on average).

Proposition 2.8 complements the empirical study by Demirgüç-Kunt and Detragiache (2002). They show that, in countries with explicit deposit insurance, banks are more prone to bank runs and the possibility of a bank run is increasing in the coverage of deposit insurance premiums. This shows that deposit insurance might work exactly in the opposite way it is intended to.

This section shows that banks may commit to high monitoring on their own using high levels of capital. Sometimes, however, this is not enough. That is, banks may not fully internalize the costs of several externalities that bank runs may bring. Consequently, banks may use too little capital. Capital regulation may then help to restore prudent behavior. If deposit insurance exists, banks' risk-taking may be even more problematic. Capital regulation in the presence of deposit insurance seems crucial.

## 2.6 Conclusion

This chapter has reviewed some fundamental features of banking in a unifying framework. This allows for a condensed presentation of key insights of the modern banking literature. In addition, it offers some new insights and connections between several insights from the literature.

I start with an ambitious goal to answer the question of why banks are better in monitoring borrowers. In Section 2.2, I first show that banks that pool together many borrowers (with uncorrelated returns) have higher incentives to monitor. I relate this result to several contributions in the financial intermediation literature. In Section 2.3, I discuss liquidity provision, which in combination with opaque bank loans makes banks inherently unstable. I describe the possibility of sunspot and information-based bank runs. I also focus on stability-oriented bank regulation that aims at preventing bank crises. The threat of a bank run may, however, have a positive role; that is, it could exert pressure on banks to behave prudently.

Finally, I turn to the role of capital and capital regulation. Banks may hold capital on their own to commit themselves to high(er) monitoring. If this effect is insufficient, the regulator may impose capital requirements on banks. Although the literature is not conclusive, capital and capital requirements in general induce more bank monitoring. Interestingly, one of the insights coming from the survey in this chapter is that deposit insurance may increase the probability of a bank run in a well-developed banking system but decrease it in a less-developed banking systems.

## 2.7 Appendix

*Proof of Proposition 2.1*

Because $p_G > p_B$, one can easily show that $\frac{p_B(N)}{p_G(N)}$ is increasing in $N$ for $N \geq 1$ and is therefore higher than $\frac{p_B}{p_G}$ for $N > 1$. ■

*Proof of Corollary 2.1*

Rewrite (2.7) to obtain $\bar{c}_M(N) = [R - r_F]\frac{p_G(N)-p_B(N)}{p_G(N)}$. Note that $\frac{\partial \bar{c}_M(N)}{\partial r_F}$ is negative. ■

*Proof of Corollary 2.2*

Use the definition in (2.7) to compute $\frac{\partial^2 \bar{c}_M(N)}{\partial r_F \partial N} = \frac{\partial}{\partial N}[\frac{p_B(N)}{p_G(N)}]$, which is positive. ■

*Proof of Proposition 2.2*

Use (2.14) to see that $\frac{\partial \tilde{c}_M}{\partial \alpha} < 0$. ■

*Proof of Proposition 2.3*

For $c_M \leq \bar{c}_M$, the bank always monitors (see (2.5)) and there is no bank run. For $c_M > \bar{c}_M$, the following analysis applies. If a depositor observes the level of a bank's monitoring, he expects to obtain

$$V_O = \beta p_G r_D + [1-\beta] p_T r_D - c_D. \tag{2.37}$$

That is, with probability $\beta$, the bank monitors and a depositor expects $p_G r_D$. With probability $1-\beta$, the bank does not monitor and a depositor withdraws early and receives $p_T r_D$. A depositor also incurs the cost of observing the level of a bank's monitoring $c_D$.

If a depositor does not observe the level of a bank's monitoring, he expects

$$V_{NO} = \beta p_G r_D + [1-\beta] p_B r_D. \tag{2.38}$$

That is, with probability $\beta$, the bank monitors and a depositor expects $p_G r_D$. With probability $1-\beta$, the bank does not monitor. However, a depositor does not observe this. Hence, he does not withdraw funds and he receives $p_B r_D$.

In equilibrium, probability $\beta$ is such that depositors are indifferent between observing bank's behavior or not. That is, (2.37) equals (2.38); that is,

$$\beta p_G r_D + [1-\beta] p_T r_D - c_D = \beta p_G r_D + [1-\beta] p_B r_D.$$

Rearranging yields

$$[1-\beta][p_T - p_B] r_D = c_D. \tag{2.39}$$

In equilibrium, probability $\alpha$ is such that the bank is indifferent between monitoring and not monitoring; that is, (2.13) is satisfied with equality. In addition, depositors can directly invest in borrowers, which yields the following condition on the promised deposit rate $r_D$ at $t = 2$,

$$r_F = \{\beta p_G + [1-\beta] p_B\} r_D. \tag{2.40}$$

Solving (2.39) and (2.40) shows that the bank monitors its borrowers with probability

$$\beta = 1 - \frac{c_D p_G}{c_D[p_G - p_B] + r_F[p_T - p_B]}, \tag{2.41}$$

and that the depositor demands a deposit rate of

$$r_D = \frac{r_F}{p_G} + \frac{c_D[p_G - p_B]}{p_G[p_T - p_B]}. \tag{2.42}$$

The probability that the depositor monitors its bank is (use (2.13))

$$\alpha = \frac{[p_G - p_B][R - r_D] - c_M}{[p_T - p_B][R - r_D] - c_M}. \tag{2.43}$$

Depositors would run on the bank *without* observing its monitoring if

$$\beta p_G r_D + [1-\beta]p_B r_D > p_T r_D. \tag{2.44}$$

The condition in (2.44) can be rewritten as $\beta > \frac{p_T - p_B}{p_G - p_B}$. Note from (2.41) that $\beta > \frac{p_B}{p_G - p_B}$. This and Assumption 2.1 guarantee that the condition in (2.44) is always satisfied. This concludes the proof. ■

*Proof of Corollary 2.3*
Use (2.41) to see that $\frac{\partial \beta}{\partial c_D} < 0$. Insert (2.42) into (2.43) to see that $\frac{\partial \alpha}{\partial c_D} > 0$. Combine this with (2.15) to see that $\frac{\partial p_{BR}}{\partial c_D} > 0$.

Observe from (2.13) that $R > r_D$, otherwise the bank's profit is negative. Use (2.43) to see that $\frac{\partial \alpha}{\partial c_M} > 0$. Combine this with (2.15) to see that $\frac{\partial p_{BR}}{\partial c_M} > 0$. ■

*Proof of Proposition 2.4*
Subtract (2.20) from (2.27) and (2.22) from (2.29) to obtain

$$\bar{Pr}(BR, G) - Pr(BR, G) = -[\bar{Pr}(BR, B) - Pr(BR, B)] = \frac{\Phi[1-\Phi][1-2\Phi]}{2[1-\Phi+\Phi^2]\{[1-\Phi]^2 + \Phi^2\}}. \tag{2.45}$$

Observe that (2.45) is negative because the signal is informative; that is, $\Phi > \frac{1}{2}$. ■

*Proof of Corollary 2.4*
Differentiating (2.35) yields $\frac{\partial k^*}{\partial r_F} = \frac{p_G}{r_F^2}[R - \frac{c_M}{\Delta p}]$, which is positive because of the condition in (2.1). ■

*Proof of Proposition 2.5*
Insert (2.30) into (2.33) and solve for $A$ from $\Pi_M = \Pi_{NM}$ to obtain (2.34). ■

*Proof of Proposition 2.6*
If the condition in (2.5) is satisfied, the bank does not need to raise costly capital to commit to monitoring. Hence, $k^* = 0$. However, if $c_M \geq \bar{c}_M$, the bank must hold sufficiently high

capital to commit to monitoring. That is, the condition in (2.33) is satisfied and must be satisfied with equality because capital is costly. Solve (2.33) for $k$ to obtain (2.35). ■

*Proof of Proposition 2.7*

Observe from (2.31) that the profit of the bank increases with $k$ as long as $r_E < \bar{r}_E$ where

$$\bar{r}_E \equiv p_G r_D. \tag{2.46}$$

Insert (2.42) into (2.46) to obtain (2.36). ■

*Proof of Lemma 2.1*

Replacing $p_B$, $p_G$, and $p_T$ with $\bar{p}_B$, $\bar{p}_G$, and $\bar{p}_T$ in (2.41) and (2.42) one obtains

$$\begin{aligned} \beta &= 1 - \frac{c_D \bar{p}_G}{c_D[\bar{p}_G - \bar{p}_B] + r_F[\bar{p}_T - \bar{p}_B]}, \\ r_D &= \frac{r_F}{\bar{p}_G} + \frac{c_D[\bar{p}_G - \bar{p}_B]}{\bar{p}_G[\bar{p}_T - \bar{p}_B]}. \end{aligned} \tag{2.47}$$

Rearrange (2.47) to obtain

$$\beta = 1 - \frac{c_D\{p_G + [1 - p_G]\delta\}}{c_D[p_G - p_B] + r_F[p_T - p_B]} \frac{1}{1-\delta}.$$

Observe that $\frac{\partial \beta}{\partial \delta} < 0$. ■

*Proof of Proposition 2.8*

For $p_G = 1$, observe that neither $r_D$ nor $\alpha$ depends on $\delta$. Hence, due to Lemma 2.1, $\frac{\partial p_{BR}}{\partial \delta} > 0$ for $p_G = 1$. Observe that $r_D$ at $p_G = 1$ must be smaller than $R$. This yields the condition

$$r_F + \frac{c_D[1 - p_B]}{p_T - p_B} < R.$$

For $p_G < 1$, we have $\frac{\partial \alpha}{\partial \delta} < 0$. Now note that

$$\frac{\partial r_D}{\partial p_G} = \frac{1}{p_G^2}\{-r_F + \frac{c_D p_B}{p_T - p_B}\} < \frac{1}{p_G^2}\{-r_F + \frac{p_B}{1 - p_B}[R - r_F]\} < 0.$$

Observe that for $p_G \to p_B$ the deposit rate $r_D$ increases above $R$. At the point where $p_G$ is such that $r_D$ approaches $R$, the negative effect of $\delta$ on $\alpha$ is extremely high and prevails over its effect on $\beta$. Hence, there exists $\bar{p}_G$ such that for $p_G > \bar{p}_G$ I have $\frac{\partial p_{BR}}{\partial \delta} > 0$ and for $p_G \leq \bar{p}_G$ I have $\frac{\partial p_{BR}}{\partial \delta} \leq 0$. ■

# Chapter 3

# Competition and Entry in Banking: Implications for Capital Regulation


**Abstract**

This chapter assesses how capital regulation interacts with the degree of competitiveness of the banking industry. It particularly asks two questions: *i*) how does capital regulation affect endogenous entry; and *ii*) how do changes in the competitive environment affect bank monitoring choices and the effectiveness of capital regulation? The approach deviates from the extant literature in that it allows for heterogeneous bank quality and recognizes the fixed costs associated with banks' monitoring technologies. The most striking result is that increasing costly capital requirements can lead to *more* entry into banking, essentially by reducing the competitive strength of lower quality banks. I show that an implication of this is that banks on average may want the regulator to impose a higher capital requirement on the industry than is socially optimal. I also show that competition improves the monitoring incentives of better quality banks and deteriorates the incentives of lower quality banks; and that precisely for those lower quality banks competition typically compromises the effectiveness of capital requirements.[1]



Keywords: Competition, Entry, Capital Regulation, Banking
JEL CLASSIFICATION: G21, G28


[1] This chapter is part of joint work with Arnoud Boot.

## 3.1 Introduction

A key public policy issue concerning the banking sector is how competition and regulation affect the functioning of financial institutions, and specifically, what the interaction is between competition and the effectiveness of regulation. In this chapter, I particularly ask two questions: *i*) how does capital regulation affect endogenous entry; and *ii*) how do changes in the competitive environment affect bank monitoring choices and the effectiveness of capital regulation?

The importance of these issues is unquestionable. The increasingly competitive and dynamic environment of banking puts severe strains on the viability and effectiveness of regulation. Competition also affects the behavior of the players in the industry directly. More competition could induce banks to take more risks, which could undermine the stability of the industry (see Vives (2001a) for a review). Simultaneously, there is a concern about the impact of capital regulation on the competitive dynamics, including level-playing-field issues.

These issues are analyzed in an industrial organization framework in which I distinguish multiple banks, and I let banks differ in quality. These quality differences are linked to banks' abilities to monitor potential borrowers, and affect the riskiness of banks and the profitability of their lending operations. I let banks compete for borrowers and analyze how their choices of monitoring technology, and hence risk, are affected by capital regulation and the intensity of competition. I show that increasing interbank competition – that is, opening up locally segmented markets for cross-market competition (holding the total number of banks fixed) – improves the monitoring incentives of better quality banks and deteriorates the incentives of lower quality banks, and that precisely for those lower quality banks competition typically also compromises the effectiveness of capital requirements. These results point to the difficulty of introducing more competition into protected markets when the local banking system is of low(er) quality.

When I permit endogenous entry, and thus allow for an increase in the aggregate number of banks, I obtain arguably the most striking result of this analysis. I show that existing work has overlooked a key benefit of increasing capital requirements in that it reduces the competitive strength of lower quality banks vis-à-vis high(er) quality banks, and this effect encourages entry. This insight complements observations by practitioners and policymakers that have sometimes argued that the real contribution of the existing Basel I capital requirements is that they have raised capital levels across the industry and, in doing so, improved the stability of the financial system. I show that discouraging weaker players is an important aspect of the link between capital requirements and the quality of the industry. This cleansing effect of capital regulation also gives a moment of pause for the ambitions of the new Basel II capital requirements. Trying to differentiate capital requirements between banks and tailor them to the exact risks taken by these institutions might truly be of secondary importance if raising capital requirements across the board has had such favorable effect on the industry.

While increasing (costly) capital requirements always has a cleansing effect on the industry by discouraging weaker banks, the net effect on entry is positive only when there are sufficiently many lower quality banks in the economy *and* local banking markets are not fully segmented, i.e. interbank competition should be feasible such that market share can be captured from lower quality banks. In such an environment there is a distinct benefit to discouraging lower quality banks that exceeds the direct cost that costly capital requirements impose on the industry.

The reason why capital requirements work against the competitive strength of low-quality banks deserves some further discussion. In this analysis, this is a consequence of deposit insurance. As long as the deposit insurance premium cannot be made fully type (and/or risk) dependent, deposit insurance effectively subsidizes low-quality banks relative to high(er) quality banks. This makes low-quality banks more competitive than they would otherwise be, and makes it more difficult for good banks to gain market share at their expense.[2] The consequence of this is that lending rates are pushed down by the over-competitive low-quality banks, and this discourages entry. Increasing capital requirements mitigates this by reducing the deposit insurance subsidy for lower quality banks, thereby reducing their competitive strength and encouraging entry.[3]

This approach deviates from the extant literature in that it recognizes the fixed costs associated with banks' monitoring technologies. These fixed costs give importance to a bank's scale and hence market share.[4] The fixed costs put low-quality banks at a double competitive disadvantage: they are subjected to a higher unit cost of monitoring (which is an artifact of their intrinsically less efficient monitoring technology) *and* face an amplification of this because of their anticipated smaller scale of operations compared to more competitive good banks. These effects lead to less monitoring and hence more risk for low-quality banks. For good banks, competition allows them to gain market share and this encourages monitoring. I show that these issues turn out to be particularly relevant when countries with different quality banking systems open up their domestic markets to cross-border competition. Strong countries gain, but substantial instability could be expected in weaker countries.

Key to this analysis are the quality differences between banks. These differences create an asymmetric impact of competition on the behavior of banks.[5] This work contrasts with

[2]Although lack of contractability generally makes it infeasible to have deposit insurance premiums fully risk-based (i.e., type and risk dependent) and effectively introduces cross-subsidies, systemic concerns in the banking industry create all kinds of other cross-subsidies and interdependencies. For example, many agree that the functioning of the banking sector depends crucially on the confidence that the public has in the financial system at large. Any such interdependencies could induce similar competitive distortions as analyzed in this chapter.

[3]I assume that capital requirements are binding. Some have observed, however, that banks choose levels of capital above the minimum (see Flannery and Rangan (2004)). This analysis shows that capital requirements are needed for low-quality banks that seek to maximize the deposit insurance subsidy, and thus have little interest in being well capitalized. In contrast, good banks can be expected to be adequately capitalized; for example, to protect their franchise values.

[4]Recent empirical evidence points at scale economies in banking; see, for example, Focarelli and Panetta (2003). For older, more mixed, evidence, see the survey paper by Berger, Demsetz, and Strahan (1999).

[5]Allowing for quality differences introduces effects similar to the ones analyzed in the industrial orga-

the extant literature on banking competition that has primarily been analyzed in a symmetric context with equally capable banks.[6] However, quality differences between banking institutions and banking systems are of primary concern to regulators and policy makers. For example, countries with weak banking systems are reluctant to open up their market to competition because this could undermine their banking systems further. Indeed, this analysis confirms that competition has a negative impact on weak banking systems. Possibly even more troublesome, I show that competition makes capital regulation typically less effective precisely in those weak banking systems, whereas it strengthens the incentive effects of capital regulation in high(er) quality banking systems. Nevertheless, this analysis is rather positive on the role of capital requirements. Capital requirements do mitigate risk-taking incentives, and, when I allow for endogenous entry (and no longer hold the aggregate number of banks fixed), they "cleanse" the banking system by reducing the competitive strength of weak banks, and in doing so could encourage entry.

I also ask how the social welfare maximizing level of capital requirements relates to banks' privately optimal choices of capital, and particularly to the level of capital requirements that the banks would prefer the regulator to impose on the industry. I show that banks may want to have the regulator impose a level of capital requirements that exceeds the social welfare maximizing level.

This chapter is organized as follows. In Section 3.2 I develop the model, including the specification of the competitive environment. Section 3.3 presents some basic results. Section 3.4 analyzes how interbank competition affects the effectiveness of the capital requirements. In Section 3.5, I endogenize entry and analyze how entry is affected by changes in capital requirements. Section 3.6 contains the social welfare analysis and the empirical predictions. Section 3.7 concludes the chapter. All proofs are relegated to Appendix A. Social welfare is extensively analyzed in Appendix B.

## 3.2 The Model

### 3.2.1 Preliminaries

There are four players in the model: borrowers (companies asking for loans), depositors (and providers of capital), commercial banks, and the regulator (who sets the capital requirement and provides deposit insurance).

Banks specialize in lending and fund themselves with deposits and capital. I assume that banks acquire core expertise in monitoring borrowers, and that this expertise is valuable to

nization literature that focuses on non-price competition through product differentiation (see, for example, Shaked and Sutton (1982)). Interfirm reallocations to more productive firms are also analyzed in the trade literature (see Melitz (2003) and Syverson (2004)).

[6]See Repullo (2004), Matutes and Vives (2000), and Boyd and De Nicolo (2005). Exceptions are the recent papers by Freixas, Hurkens, Morrison, and Vulkan (2004) and Kopecky and VanHoose (2006) that also allow for heterogeneity in ability between banks. However, neither focuses on the interaction between capital regulation, deposit insurance, and competition, which is the focus of this analysis.

the companies that they finance. In particular, I have the monitoring technology of a bank affect the success probability of the project that the bank is financing. This captures the role that banks play in relationship banking: banks invest in borrower-specific knowledge that might be beneficial to their borrowers.[7]

The funding of the banks comes from (liquid) deposits and capital. The liquidity of deposits is rooted in deposit insurance that I assume to be present. Deposit insurance is available at a fixed cost. This potentially introduces moral hazard on the part of banks and helps explain the role of capital requirements: capital requirements may contain asset substitution moral hazard. Thus this chapter is related to studies of the role of capital in reducing risk-taking; see, for example, Hellmann, Murdock, and Stiglitz (2000).[8] I assume that bank management is aligned with shareholders.

The regulator sets the capital requirement and provides deposit insurance.

### 3.2.2 Model details

*Preferences and time line:* There is universal risk neutrality, with $r_f$ representing the riskless interest factor (one plus the interest rate). There are four dates, $t = 0$, 1, 2, and 3. At $t = 0$, the regulator sets the capital requirement $k$ and banks decide whether or not to enter the banking industry. At $t = 1$, each borrower is matched with a bank, and banks decide on their investments in monitoring technology. I call the initial bank that the borrower is matched with the "incumbent bank." This bank makes the borrower an initial offer. At $t = 2$, the borrower might find a second competing bank. If this happens, the initial incumbent bank and competing bank compete as Bertrand competitors. The borrower chooses the best offer. Subsequently, the winning bank collects the necessary capital and deposits, makes the loan, and the borrower invests. Payoffs are realized at $t = 3$. In Figure 3.1 I have summarized the sequence of events.

*Borrowers:* A borrower needs a single-period loan of \$1 to finance a project at $t = 2$, with a payoff at $t = 3$. All borrowers are identical. A borrower's project has a return of either $Y$ or 0 (zero). The probability of success (i.e. the pay-off $Y$) depends on a bank's investment in monitoring technology $\nu_j$ with $j \in \{I, C\}$, where $j = I$ refers to the incumbent bank and $j = C$ is the competing bank. I let the probability of success be equal to the investment $\nu_j$, and hence normalize $\nu_j$ to $\nu_j \in [0, 1]$. All other things being equal, when a borrower can choose between two competing offers, he will choose the bank with the highest $\nu_j$.[9] The aggregate demand for loans from all borrowers is normalized to 1.

*Depositors and providers of capital:* With complete deposit insurance, depositors are

[7]See Boot and Thakor (2000) and Ongena and Smith (2000) for reviews of relationship banking.

[8]Allen, Carletti, and Marquez (2007) analyze a related rationale for capital. In their analysis, institutions choose capital in response to lending-market induced market discipline. In Morrison and White (2005), raising capital requirements could be an appropriate response to counter a confidence crisis.

[9]Actually, I assume (see later) that a borrower can only switch at a cost. Consequently, the incumbent bank has an incumbency advantage, and the competing (second) bank must overcome this when making its offer.

$t = 0$:

♠ The regulator sets the capital requirement $k$.

♠ Banks enter the banking industry (if applicable).

$t = 1$:

♠ Each borrower is matched with a bank.

♠ Each bank discovers its type.

♠ Banks invest in monitoring technology.

♠ Each borrower receives an initial offer from his bank.

$t = 2$:

♠ Each borrower searches for a competing bank.

♠ If a second bank materializes, the incumbent bank and the "second bank" compete as Bertrand competitors. If no second bank is available, only the borrower's first offer is available.

♠ Each bank collects the necessary capital and deposits, and funds its borrowers.

♠ Borrowers undertake their projects.

$t = 3$:

♠ Payoffs are realized.

Figure 3.1: Timeline

willing to supply their funds at the risk-free interest rate $r_f$. The deposit insurance premium is fixed, and to simplify matters I assume that this premium is included in the gross costs of deposits. Hence, the cost of deposits is $r_D > r_f$. Banks face a binding capital requirement $k$. They collect this proportion $k$ of the total funds needed from the providers of bank capital and $[1 - k]$ from depositors.

Capital is costly. I let the cost of capital equal $\rho$, where $\rho > r_D$.[10]

*Commercial banks:* Banks choose to enter the banking industry at $t = 0$. All banks are initially (perceived as) identical. At $t = 1$, with $N$ banks present, each bank is matched with $1/N$ of the borrowers.[11] Banks then learn whether their type $\tau$ is good ($G$) or bad ($B$), thus $\tau \in \{B, G\}$, and following this they choose their investments in monitoring technology. The cross-sectional probability of being good ($\gamma$) or bad ($1 - \gamma$) is known to all. Banks have an intrinsic monitoring ability $\underline{\nu}_\tau$, with $0 < \underline{\nu}_B < \underline{\nu}_G$. A bank can increase its monitoring ability to a higher level $\nu$ at a cost $\frac{c}{2}[\nu - \underline{\nu}_\tau]^2$.[12]

*Competitive environment:* Competition between banks occurs in two phases. In the first phase (at $t = 1$), all $N$ banks are allocated $1/N$ of the total borrowers. Each bank specifies an interest rate offer $R$ for its allocated borrowers. At $t = 2$, borrowers succeed in locating a competing offer with probability $q$. With probability $[1 - q]$, they do not find a competing offer. When this happens, borrowers have no choice but to accept the initial offer, provided this gives them a non-negative expected return. When a second bank is found, both the initial (incumbent) and the second bank compete as Bertrand competitors. I assume that

[10]See Holmstrom and Tirole (1997) and Diamond and Rajan (2000) for explicit models of why the cost of capital might be higher than the return that depositors demand. Note that this assumption bypasses the question *how* capital is raised, including potential adverse selection problems.

[11]Because all banks are perceived as identical at that moment, this even distribution of borrowers over all banks is quite natural.

[12]Using a generalized cost function satisfying the Inada conditions produces similar results but at a cost of substantial complexity.

at this stage the borrowers and the competing banks can observe the monitoring technology adopted by each bank and their types. Each borrower then chooses the bank that gives the highest expected return net of funding costs.[13]

One important additional consideration is that, if a borrower switches to a competing bank, he incurs a fixed switching cost $S$. This allows the incumbent bank to earn rents even if the competing bank is equally capable. In other words, the incumbent bank effectively has an incumbency advantage vis-a-vis the competing banks.

## 3.3 Initial Analysis: Some Basic Results

I solve the model using backward induction. I first determine the strategies and the valuation of the incumbent bank at $t = 2$ conditional on the levels of investment in monitoring technology $\nu_j$. Subsequently, I compute the optimal investments in monitoring technology $\nu_j$ at $t = 1$, anticipating the events at $t = 2$.

At $t = 1$ each borrower is matched with a bank (i.e., the incumbent bank). The initial offer that this bank makes is a monopolistic offer. The bank can always improve on this offer when its borrower succeeds in obtaining a competing offer. The incumbent bank optimally sets the interest rate equal to the maximum payoff of the borrower such that it obtains all surplus when competition would materialize; thus

$$R^{\max}(\nu_I|\text{no competition}) = Y. \tag{3.1}$$

At $t = 2$, the borrower finds with probability $q$ a competing bank; with probability $[1 - q]$ the borrower only has access to the offer of the incumbent bank. When the borrower has no access to a competing offer, he accepts the monopolistic offer and loses all rents. When the borrower has a competing offer, both banks compete for the borrower as Bertrand competitors.

The investment that a bank is prepared to make in its monitoring technology depends crucially on the profitability of the lending operation, and hence the competition it anticipates. Recall that each of the $N$ banks is allocated $1/N$ borrower. For this initial allocation, a bank has a role as incumbent bank. Competition implies that it may lose this borrower (and/or be forced to lower its lending rate), but the bank could also gain new borrowers by challenging other (incumbent) banks. I first derive some preliminaries.

A bank maximizes its market value of equity; that is, its expected profits net of costs of debt, discounted by the cost of capital. The value that the incumbent bank derives from its

[13]This formulation is qualitatively identical to a Hotelling framework. It gives us rents in the banking system that are decreasing in the level of interbank competition $q$, similar to a Hotelling-type specification with transaction costs that are decreasing in $q$. Actually choosing a Hotelling framework would have complicated the analysis substantially, given the heterogeneity across banks that I have.

$1/N$ initial borrower, conditional on having no competing offer, equals

$$V(\nu_I|\text{no competition}) = \frac{1}{N}\{-k + \nu_I \frac{R^{\max}(\nu_I|\text{no competition}) - [1-k]r_D}{\rho}\}.$$

The bank then obtains all surplus. Inserting (3.1), I can write

$$V(\nu_I|\text{no competition}) = \frac{1}{N}\{-k + \frac{\nu_I X}{\rho}\}, \tag{3.2}$$

where $X \equiv Y - [1-k]r_D$.

The value that the incumbent bank derives from its borrower when he obtains a competing offer is computed as follows. The lowest interest rate $R^{\min}(\nu_C)$ that a competing bank with investment in monitoring technology $\nu_C$ is just willing to offer follows from its zero NPV condition[14]

$$-k + \frac{\nu}{\rho}\{R^{\min}(\nu_C) - [1-k]r_D\} = 0. \tag{3.3}$$

The incumbent bank is able to outbid the competing bank if it can make an offer such that the borrower obtains a surplus at least equal to what he could obtain with the best competing bank's offer $R^{\min}(\nu_C)$. I proceed as follows. The maximum interest rate that the incumbent can charge the borrower without losing him to a competitor with $\nu = \nu_C$ is $R^{\max}(\nu_I|\nu_C)$, where $R^{\max}(\nu_I|\nu_C)$ is such that the borrower is indifferent between this offer and the best offer of the competing bank. That is,

$$\nu_I[Y - R^{\max}(\nu_I|\nu_C)] = \nu_C[Y - R^{\min}(\nu_C)] - S, \tag{3.4}$$

where I have taken into account that the borrower incurs a switching cost $S$ when he switches to the competing bank.

Conditional on a competing bank being present with monitoring technology $\nu_C$, the value that the incumbent bank derives from its initial borrower if it decides to respond with offer $R^{max}(\nu_I|\nu_C)$ is

$$V(\nu_I|\text{competition, } I \text{ responds}) = \frac{1}{N}\{-k + \nu_I \frac{R^{\max}(\nu_I|\nu_C) - [1-k]r_D}{\rho}\}.$$

Using (3.3) and (3.4), I can rewrite this as

$$V(\nu_I|\text{competition, } I \text{ responds}) = \frac{S + [\nu_I - \nu_C]X}{\rho N}.$$

If $S + [\nu_I - \nu_C]X < 0$, the incumbent bank is not willing to respond (i.e. does not offer $R^{\max}(\nu_I|\nu_C)$) and hence the competing bank prevails. The value that the incumbent bank

[14]Note that the cost of investing in monitoring technology is incurred at $t = 1$. This is sunk once the competition phase is reached at $t = 2$, and thus is not considered when the bank sets the interest rate.

derives from its initial borrower then equals zero. In total,

$$V(\nu_I|\text{competition}) = \max(0, \frac{S + [\nu_I - \nu_C]X}{\rho N}). \tag{3.5}$$

The incumbent bank can also compete for the borrowers of other banks. Strictly speaking, these other banks are the incumbent banks for those borrowers. To prevent confusion, I will continue to call "this bank" the incumbent bank, and to use $\nu_I$ for its technology and $\nu_C$ for the technology of the other banks. If the incumbent bank competes for the borrower of another bank with monitoring technology $\nu_C$, the value that it derives from the possibility of obtaining this new borrower is

$$V(\nu_I|\text{new borrower}) = \max(0, \frac{-S + [\nu_I - \nu_C]X}{\rho N}). \tag{3.6}$$

The expression (3.6) is very similar to (3.5), but note that the incumbency advantage now works against "this bank."

An useful result relates to the expected number of other borrowers that a bank can make an offer to:

**Lemma 3.1.** *The expected number of other borrowers that a bank can make an offer to is $q/N$.*

To see this, observe that there are $[N-1]$ other banks in the economy. The incumbent bank has a probability $q/[N-1]$ that it can compete for the borrower of any one of these banks.[15] Recall that each of these banks has $1/N$ borrower. Thus the expected number of other borrowers that the incumbent bank can make an offer to is $[N-1] \times \frac{q}{N-1} \times \frac{1}{N} = \frac{q}{N}$.

This lemma highlights that there is a degree of symmetry in this model. That is, the way that I have structured the competition between banks implies that any incumbent bank faces a probability $q$ that others will bid for its $1/N$ borrower. Thus, the fraction $q/N$ of its borrower is – in expected value sense – at risk. However, Lemma 3.1 shows that the flip side is that any incumbent bank can bid in expected value for the fraction $q/N$ of borrowers of other banks.[16] The actual outcome will depend on the quality differentials between banks and their potentially different levels of (investment in) monitoring technology.

I now derive the equilibrium investments in monitoring technology. At $t = 1$, the $N$ banks first learn their types, and then individually choose their levels of investment in monitoring technology. I consider a simultaneous move game and derive a separating Nash equilibrium in pure strategies. In choosing their individual levels of investment in monitoring technology, each bank makes a conjecture about the choices of the other banks. In deriving this separating

[15] For this, see that a borrower receives a competing offer with probability $q$ and there are $[N-1]$ banks that could obtain the opportunity to make this competing offer.

[16] In Section 6.2.1 I also analyze how competition evolves if there is only one-sided competition. What I mean by this is that a bank may face competition from other banks for its own borrowers, but has no access to the borrowers of the competing banks; that is, the borrowers of these other banks are shielded from competition.

Nash equilibrium, I need to put some constraints on the incumbency advantage $S$. More specifically, I assume

**Assumption 3.1:** $\frac{X^2}{c\rho N} < S < [\underline{\nu}_G - \underline{\nu}_B]X$.

This assumption can be explained as follows. The lower bound on the incumbency advantage ensures that when an incumbent bank competes with a bank of equal quality its incumbency advantage prevails. That is, this competing bank of equal quality will not find it optimal to overcome the incumbency disadvantage by choosing a much higher investment in monitoring technology. Without an incumbency advantage this could be optimal because capturing the incumbent bank's borrower offers scale advantages justifying the higher investment in monitoring technology. The incumbency advantage makes this strategy too costly and ensures that banks of the same type will choose identical strategies; that is, they will choose the same level of investment in monitoring technology. Thus banks of the same type will not take market share at each others expense.[17]

The upper bound on the incumbency advantage ensures that quality matters in competition; that is, a good bank can overcome the incumbency advantage of a bad bank and take its borrower.

I now proceed as follows. Each bank chooses its investment in monitoring technology $\nu$ holding the strategy of other banks fixed. I continue to analyze the problem from the perspective of the incumbent bank. Its investment in monitoring technology is $\nu_I$. The other banks choose $\nu_C^j$, where $j$ refers to one of the other $[N-1]$ banks. I can now write the expected value of the incumbent bank of type $\tau$, $\tau \in \{B, G\}$, net of funding costs, as

$$\begin{aligned} V_\tau(\nu_I) &= \frac{1-q}{N}[-k + \frac{\nu_I X}{\rho}] + \frac{q}{\rho N}\sum_{j=1}^{N-1}\frac{1}{N-1}\max(0, S + [\nu_I - E(\nu_C^j)]X) + \\ &+ \frac{q}{\rho N}\sum_{j=1}^{N-1}\frac{1}{N-1}\max(0, -S + [\nu_I - E(\nu_C^j)]X) - c\frac{[\nu_I - \underline{\nu}_\tau]^2}{2}. \end{aligned} \quad (3.7)$$

In (3.7), the first expression is the bank's profitability when there is no competition; see (3.2). This happens with probability $[1-q]$. The second expression is the expected profit on its initial borrower when there is competition; see (3.5). The summation is over all $[N-1]$ competing banks. The third expression is the incumbent bank's profit from successfully attracting borrowers away from other banks, as given in (3.6). The last expression is the cost of investing in monitoring technology.

Each bank maximizes its analogous expression (3.7). I now have the following result.[18]

[17] Similarly, the lower bound on $S$, $\frac{X^2}{c\rho N} < S$, effectively puts a lower bound on $c$. This is important because the lower bound on $S$ ensures that it is prohibitively costly for a bank to overcome its intrinsically lower quality ($\underline{\nu}_B < \underline{\nu}_G$) by choosing a (much) higher level of investment in monitoring technology.

[18] I impose restrictions to guarantee that the monitoring choices are in the interior and the borrowers' projects are sufficiently attractive that all banks are willing to provide funding. These restrictions are shown to be compatible with Assumption 3.1 (see the proof of Proposition 3.1).

**Proposition 3.1.** *There exists a separating Nash equilibrium consisting of the strategies* $\nu_B^*$ *for bad banks and* $\nu_G^*$ *for good banks, where* $\nu_G^*$ *and* $\nu_B^*$ *equal*

$$\nu_B^* = [1 - q\gamma]\frac{X}{c\rho N} + \underline{\nu}_B, \quad (3.8)$$

$$\nu_G^* = \{1 + q[1 - \gamma]\}\frac{X}{c\rho N} + \underline{\nu}_G. \quad (3.9)$$

From this proposition it readily follows that in equilibrium good banks choose a strictly higher level of monitoring than bad banks.[19] That is, when comparing (3.9) and (3.8), I see that good banks have a higher intrinsic monitoring ability than bad banks ($\underline{\nu}_G > \underline{\nu}_B$), and invest more in additional monitoring because of their anticipated gains in market share due to competition. To see this, observe that $\nu_G^*$ is positively affected by the competition parameter $q$, whereas $q$ affects $\nu_B^*$ negatively.

I can now derive a corollary that relates to the effect of capital requirements on monitoring incentives.

**Corollary 3.1.** *Higher capital requirements improve the monitoring incentives of both good and bad type banks.*

Capital requirements favorably affect monitoring incentives in this model because higher capital forces banks to internalize more risk, which in turn reduces risk-taking incentives, implying more monitoring.[20] This is a typical result, and follows from the objective function of banks in this analysis; that is, banks maximize the value of capital.

## 3.4 Competition and the Effectiveness of Capital Regulation

I continue to hold the number of banks $N$ fixed; in Section 3.5, I will allow for entry. The focus for now is on interbank competition. The key question analyzed is how relaxing barriers between existing banks (inducing more interbank competition) affects the strategies of banks and the effectiveness of capital regulation.

The type of competition that I analyze in this section could be interpreted as opening up national markets to foreign competitors. Across the globe, I increasingly see that banks are challenged in their home markets by foreign players, but also themselves challenge other banks in their home markets. The reasons for this include globalization, developments in

[19] When good and bad banks are very similar to each other and the incumbency advantage is very high (note that this would violate Assumption 3.1), there exists another (pooling) Nash equilibrium in which all banks focus only on their incumbent borrowers. Neither the good nor the bad banks try to win borrowers from other banks, simply because the high incumbency advantage prevents any type of bank from profiting from non-incumbent borrowers. In the absence of an incumbency advantage (again a violation of Assumption 3.1), no equilibrium exists in pure strategies.

[20] Recall from the specification of the model in Section 3.2 that higher monitoring increases the expected returns on the projects. Strictly speaking, it only reduces risks for all $\nu_\tau > 1/2$.

information technology, and deregulation. In particular, developments in information technology could potentially allow banks to enlarge their geographic area of operations without having a local presence in those markets; this possibly reduces the competitive advantage of local players (see, for example, Petersen and Rajan (2002)).

In this model, these developments positively impact $q$, the probability that borrowers have access to a competing second offer. I continue to assume symmetry in the structure of competition. That is, in the model that I have developed so far, an incumbent bank faces competition for its own $1/N$ borrower with probability $q$, but it also receives access to an equal number of borrowers (in expectation) from other banks; see Lemma 3.1. Thus, in expected value sense, the number of borrowers at risk equals the number it could gain. A bank's actual success depends both on its inherent quality and on its investment in monitoring technology relative to that of its competitors.

I now analyze how relaxing barriers to competition between existing banks (i.e., increasing $q$) affects monitoring incentives and the effectiveness of capital regulation. Following this, I analyze how capital requirements affect the values of good and bad banks.

I first analyze the effect of competition on monitoring incentives. From Proposition 3.1 I can directly show that

**Corollary 3.2.** *Increasing interbank competition (higher $q$; holding the number of banks $N$ fixed) decreases the optimal level of monitoring of bad banks ($\nu_B^*$) but increases the optimal level of monitoring of good banks ($\nu_G^*$).*

The intuition for this corollary is as follows. Higher competition reduces the probability that bad banks can hang on to their own borrowers. This diminishes their anticipated market share and hence lowers their incentives to invest in monitoring technology. Good banks, however, benefit from a higher $q$ in that they can steal more borrowers from bad banks. Hence, they expect to gain market share, effectively increasing the returns on their investments in monitoring technology.

This differential impact of competition on monitoring incentives highlights an interesting property of this model. For bad banks, competition implies losing market share and hence higher per-unit costs due to the presence of fixed costs in monitoring technology. For good banks, this is precisely the reverse: competition allows for an increase in market share, and effectively helps to lower the per-unit costs.

A related question is what happens to the effectiveness of capital requirements when competition heats up. From Corollary 3.1 I know that higher capital requirements increase the investments in monitoring technology by both types. What I show next is that competition *strengthens* this positive effect for good banks, but weakens it for bad banks.

**Proposition 3.2.** *Higher interbank competition (higher $q$) negatively affects the effectiveness of the capital requirements for bad banks, but it increases the effectiveness of capital regulation for good banks.*

The intuition for this is directly related to that of Corollary 3.2. Competition reduces the marginal benefit of investing in monitoring technology for bad banks but increases this for good banks. Not surprisingly, then, the favorable impact that capital regulation has on monitoring incentives is strengthened for good banks but not for bad banks.

The results so far show that competition has a positive impact on the monitoring incentives of good banks, but undermines those of bad banks both directly (anticipating the smaller market share) and indirectly via reducing the effectiveness of capital regulation. This has implications for regulatory policy. Most importantly, competition undermines the effectiveness of capital regulation precisely for those banks for which it is needed most; that is, the bad banks. For higher quality banks, competition positively impacts the effect of capital regulation on monitoring incentives. Because elevating interbank competition also has a direct positive effect on monitoring incentives for high-quality banks, competition and stability go hand in hand. For bad banks, the opposite holds.[21]

The differential impact of capital regulation on good and bad banks is further highlighted when I look at the effect of capital regulation on the values of good and bad banks. I can derive the following proposition.

**Proposition 3.3.** *Higher capital requirements* always *reduce the value of a bad bank $V_B(\nu_B^*)$, but increase the value of a good bank $V_G(\nu_G^*)$ as long as competition is sufficiently strong (high $q$) and the quality of the banking industry is sufficiently low (low $\gamma$).*

The key to understanding this result is that capital regulation has two effects on the industry. The first effect is that capital imposes a cost on each bank because capital is more expensive than deposits. This, in isolation, reduces the value of each bank, and is the familiar result. However, a second more subtle effect is at work as well: capital regulation reduces the deposit insurance subsidy that goes to low-quality banks. That is, flat-rate deposit insurance is most valuable to bad banks, and this makes them artificially stronger competitors.[22] Capital regulation mitigates this and helps good banks capture higher rents when competing with bad banks. This has a positive impact on the value of good banks and reduces the value of bad banks.

Proposition 3.3 shows that the positive effect of capital regulation on the value of a good bank depends crucially on $q$ and $\gamma$. Good banks can only gain from higher capital requirements when $q$ is high, meaning that the banking system is rather open and competitive, such

[21] A qualification can be made. Diversification effects across banks are not present in this model. Consequently, only the success probability matters for stability. For good banks, this success probability is positively affected by competition via an increase in monitoring incentives. However, competition will generally reduce rents and this could negatively affect stability when I take into account diversification effects. Bank stability would then not only depend on the failure probability of one borrower, but also on diversification effects across borrowers and hence the level of rents the bank earns on borrowers that succeed. Similarly, taking into account diversification effects across banks would lend importance to the level of rents. What I get is that if diversification effects are considered competition has a smaller positive effect on stability for good banks. For bad banks, things would become even worse.

[22] In a very different analysis, Winton (1997) argues that deposit insurance may facilitate entry by effectively underwriting de novo banks that investors are not familiar with.

that much is gained by weakening the competitive strength of bad banks. This effect is most important when many bad banks are present (i.e., $\gamma$ is low).

To understand this further, let's reexamine the competition between good and bad type banks. I focus on the case in which an incumbent good bank faces competition from a bad bank. The rents that the good bank earns equal $\frac{1}{\rho N}[S + [\nu_G^* - \nu_B^*]\{Y - [1-k]r_D\}]$; see (3.5). Observe that these rents are increasing in the capital requirement $k$. This is the consequence of the negative effect that capital requirements have on the rents that a bad bank derives from the flat-rate deposit insurance; this reduces its competitive strength and benefits the good banks. To see this, note that a good bank faces a net cost of deposits equal to $\nu_G^*[1-k]r_D$ whereas for a bad bank this is $\nu_B^*[1-k]r_D$. Because $\nu_G^* > \nu_B^*$ deposits are effectively subsidized for bad banks. This mispricing of flat-rate deposit insurance thus unfairly helps bad banks, and makes them fiercer competitors for good banks. Higher capital requirements partially eliminate this distortionary effect.

Proposition 3.3 provides an intriguing perspective on the impact of capital requirements. Capital requirements, despite their costs, could benefit good banks under well-defined circumstances. In Section 3.5 I explore this further, and focus in particular on the impact of capital requirements on entry; that is, I endogenize $N$.

The competition that I have analyzed so far focuses on interbank competition. In the model this means that I focus on $q$, while keeping the number of players $N$ fixed. What this implies is that *within-market* competition intensifies, for example due to developments in information technology or deregulation. In the context of two countries that introduce cross-border competition, the results show that the country with low-quality banks will become even riskier and the country with high-quality banks gains and becomes safer. The direct consequence is that opening up borders is bad for the stability of a low-quality banking system and good for the stability of a high-quality banking system.

Similarly, the effectiveness of capital regulation is typically negatively affected in a low-quality system, yet favorably affected in a high-quality system. The impact of capital regulation on the valuation of banks is different as well. Low-quality banks lose value whereas high-quality banks gain value as long as the quality of the banking system is sufficiently low and $q$, the parameter of within-market competition, is sufficiently high.

## 3.5 Endogenous Entry

I now allow for entry into banking by endogenizing the number of banks $N$. The probability that a borrower finds a competing bank, $q$, now also depends on the number of banks $N$ operating in the banking system. In particular, I assume that the probability of finding a competing bank is increasing in $N$; that is, $\frac{\partial q}{\partial N} > 0$.[23]

[23] Observe that $q$ can still be largely determined by local institutional arrangements. I also let $\frac{\partial [q/N]}{\partial N} < 0$. This is a quite natural property that implies that the probability that a borrower receives his competing (second) offer from any one particular bank is decreasing in $N$.

I first analyze how monitoring choices and bank values are affected by $N$, the number of banks in the economy. Subsequently, I endogenize $N$.

**Lemma 3.2.** *An increase in the number of banks $N$ decreases both the investments in monitoring, $\nu_G^*$ and $\nu_B^*$, and the values of banks, $V_G(\nu_G^*)$ and $V_B(\nu_B^*)$.*

This lemma is intuitive. A higher number of banks reduces the anticipated market share of each bank, and this discourages investments in monitoring technology.

I now endogenize $N$, and hence allow for entry. The entry decision is made at $t = 0$. At that moment, each prospective bank does not yet know its own (future) type, but assesses its expected quality based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$. Each bank computes whether its expected profits from entering exceed the cost of entry F, anticipating the competitive environment (including the number of banks already present).[24]

To prevent complexity due to discreteness in the number of banks, I let $N$ be a continuous variable, such that $N^*$ is determined by the equilibrium condition:

$$[1-\gamma]\bar{V}_B^* + \gamma\bar{V}_G^* = F. \tag{3.10}$$

The values $\bar{V}_B^*$ and $\bar{V}_G^*$ are the equilibrium valuations of the bad and good banks at the point where $N = N^*$.

I am particularly interested in how capital regulation affects entry. The next proposition shows that higher capital requirements could encourage entry. The competition parameter $q$ is the one that obtains in equilibrium before I change the level of capital requirements.

**Proposition 3.4.** *The effect of capital regulation on entry is as follows:*

1. *When competition is low ($q < \bar{q}$), higher capital requirements decrease entry.*

2. *When competition is high ($q \geq \bar{q}$), higher capital requirements:*

   *(a) increase entry for $\gamma \in [\gamma_1(q), \gamma_2(q)]$;*

   *(b) decrease entry for $\gamma \in [0, \gamma_1(q)) \cup (\gamma_2(q), 1]$.*

This proposition points at a striking feature of capital regulation: higher capital requirements could – despite their costs – induce more entry into the industry. This happens when the banking industry is of intermediate quality, $\gamma \in [\gamma_1(q), \gamma_2(q)]$, and competition is sufficiently high ($q \geq \bar{q}$). To see this, note from Proposition 3.3 that higher capital requirements can only induce more entry when these requirements positively affect the value of good banks (otherwise both the bad and the good banks' valuations would be decreasing in the level of capital requirements, which would certainly lead to less entry). Proposition 3.3 then tells us that the level of competition should be sufficiently high (high $q$), and $\gamma$ should be sufficiently

[24] I consider only the following simple entry procedure: banks decide on entering sequentially in random order. Note that the order does not matter because all prospective entering banks are identical, and they assess their quality based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$.

low. What Proposition 3.4 shows is, that for higher capital requirements to induce more entry, a lower bound is needed on $\gamma$ as well. This can be easily understood. If $\gamma$ is too low, a prospective entering bank believes that it will turn out to be of low quality as well. In this case, it expects its value to be negatively affected by higher capital requirements (see Proposition 3.3), which discourages it from entering.

I can now analyze what happens to the effectiveness of capital regulation as an instrument to encourage monitoring when entry is endogenous. Observe that in the absence of endogenous entry (see Corollary 3.1) capital regulation always has a positive impact on monitoring incentives. I am now ready to prove the following corollary, which shows that this positive impact could be dampened by endogenous entry.

**Corollary 3.3.** *The effect of capital requirements on investment in monitoring technology for both good and bad banks is weakened when capital regulation encourages entry and strengthened when capital regulation induces less entry.*

Corollary 3.3 in combination with Proposition 3.4 offers some challenges for regulators. Capital regulation has a direct positive effect on investments in monitoring technology (Corollary 3.1), but this effect is mitigated by the higher entry that capital regulation could induce (Case 2a in Proposition 3.4). What this indicates is that under circumstances such as those in Case 2a restrictions on entry could improve on the effectiveness of capital regulation.

If capital regulation discourages entry (Cases 1 and 2b), the effectiveness of capital regulation is actually enhanced and hence entry restrictions would be redundant.

## 3.6 Social Welfare and Empirical Predictions

This section does three things. First, it analyzes what the optimal capital requirements are from a social welfare point of view and how this relates to the private incentives of banks. Further analysis of social welfare is relegated to Appendix B and Appendix C. Subsequently, this section focuses on the empirical predictions that are generated by this analysis. Third, it relates this analysis to the literature.

### 3.6.1 Social welfare

In the analysis so far I have assumed that the capital requirements set by the regulator are binding. An important issue is what precise incentives banks have. What level of capital would they prefer? Proposition 3.3 shows that for $q$ sufficiently high and $\gamma$ relatively low, a good bank wants the regulator to impose a higher capital requirement on the industry. This will elevate its value.[25] This is not surprising, observing the "cleansing role" of capital that I have identified. However, a bad bank would always prefer capital requirements to be as

[25]The regulator imposes a uniform capital standard on the industry that is consistent with the lack of observability of bank type ex ante in this setting. If capital requirements could be made contingent on bank type, good banks would prefer the regulator to impose capital requirements only on bad banks.

low as possible. As Proposition 3.3 shows, its value is decreasing in the level of the capital requirements. For now I hold the number of banks $N$ fixed.

The question I would like to answer now is how the level of capital requirements preferred by good banks compares to the welfare optimal level. Observe that the welfare optimal level of capital requirements takes into account the externality that banks impose on the deposit insurer. That is, a failure of a bank typically imposes a loss for the deposit insurer. Because banks do not bear these losses, they would not have an incentive to privately choose capital to reduce this externality. As is well known, in a homogeneous banking industry this will always put the welfare maximizing level of capital above the level that banks choose privately (see Hellmann, Murdock, and Stiglitz (2000) and Repullo (2004)). However, the heterogeneity in banking quality in this analysis gives good banks an incentive to favor high(er) capital because this reduces the competitive strength of bad banks. Actually, in the spirit of Proposition 3.4, I can prove that a bank that does not know its quality might favor positive capital requirements. That is,

**Proposition 3.5.** *For sufficiently high interbank competition $q$, $q > \hat{q}$, and intermediate values of $\gamma$, banks prefer the regulator to impose a positive level of capital on the industry.*

The intuition is very similar to Proposition 3.4. Imposing positive (costly) capital requirements only helps banks if this reduces the competitive strength of bad banks (which is only valuable if interbank competition $q$ is sufficiently high). Also, each bank should anticipate that "enough" bad banks will be present (i.e., $\gamma$ not too high) and the banks' expectations about their own quality should not be too low (i.e., lower bound on $\gamma$ needed).

Proposition 3.5 offers some interesting insights in that it shows that in a heterogeneous industry capital requirements are desired by the industry itself. This contrasts with a homogeneous industry, in which banks would privately set capital as low as possible. Observe that this contrast is somewhat subtle. In a heterogeneous industry, banks would *individually* choose low capital, but realize that they are better off if the regulator enforces high(er) capital levels across the industry. So banks prefer capital regulation to be present.[26]

What the preceding discussion implies is that I have uncovered a new rationale for capital regulation. Banks, realizing that the competitive playing field will be spoiled by some that turn out to be of low quality later, prefer that capital regulation be put in place to guarantee a more balanced competitive environment.

I have not yet provided an answer to the question how the level of capital that the banks want to have imposed relates to the socially optimal level of capital. Recall that the regulator wants to have capital regulation in place to limit the losses to the deposit insurance

[26]In this analysis, capital regulation and capital levels are committed to *prior* to banks discovering their type (and borrowers discovering the bank type). Hence, the chosen levels of capital do not reveal information about type. If I allowed banks to choose their levels of capital *after* they get to know their own type, this would make bad banks interested in choosing a positive level of capital only if that could prevent borrowers from finding out the banks' low quality (i.e., they then may want to mimic good banks). Observe that this will create a very sophisticated competition game between banks for borrowers because borrowers would not know the quality of banks making offers and hence would have difficulty choosing between offers.

fund, whereas banks may desire capital regulation to guarantee a "balanced" competitive environment. In deciding on the welfare optimal level of capital, the regulator maximizes the total surplus of banks, borrowers, and deposit insurance fund.[27] I now establish that the private optimum may exceed the welfare maximizing level of capital requirements.

**Proposition 3.6.** *For high values of interbank competition $q$, $q > \bar{q}$ (with $\bar{q} \geq \hat{q}$), and intermediate values of $\gamma$, banks prefer the regulator to impose capital requirements that are strictly above the welfare optimal level.*

The intuition is that the level of capital that banks would want to have imposed is increasing in the level of interbank competition $q$; that is, at higher $q$ it becomes more important to contain the competitive distortion that low-quality banks inflict. In the limit (for $q$ high enough), $k$ approaches 1. The regulator's optimum for the level of capital is, however, always strictly less than one. To see this, note that, for $k$ approaching 1, banks no longer fund themselves with deposits and hence the externality imposed on the deposit insurer vanishes. Thus, the regulator's choice of capital would – given its cost – never approach 1.[28]

### 3.6.2 Empirical predictions

This analysis produces several predictions that should be brought to the data. Various pieces of existing empirical evidence are available and will be discussed where applicable. An important step in testing the various predictions is distinguishing between the two competition measures, $q$ and $N$. The measure $q$, reflects the intensity of competition between existing banks. The other measure of competition is the number of banks $N$. Observe that in the model $q$ is the probability with which borrowers can receive a competing offer. This is affected by the number of banks $N$ in the market, but is also (or even primarily) determined by institutional factors such as the degree of stringency of anti-trust enforcement. The number of banks $N$ also measures bank size and degree of concentration.

The predictions are as follows:

i. Increasing the openness/competition measure $q$ shifts market share from bad to good banks. This follows from the discussion surrounding Proposition 3.1. Good banks benefit from a higher $q$ and gain market share, whereas bad banks lose market share. This prediction is supported by Stiroh and Strahan (2003), who observe that competition reallocates assets from badly performing banks to good ones.

ii. Competition (increasing $q$) undermines stability in a low-quality banking market but strengthens it in high-quality banking markets. This prediction follows from the results

[27] Observe that this implies that the regulator does not care about competition per se (only the total surplus of banks and borrowers matters). In the welfare optimization, I have not taken into account the potential externality that bank instability could impose on the economy at large.

[28] I have not analyzed what happens when I allow for entry. Under the assumptions of Proposition 3.5, this will weaken the banks' preferences for positive capital requirements.

in Corollary 3.2. There is some supporting evidence for this in the recent literature. In particular, Boyd and De Nicolo (2005), Beck, Demirgüç-Kunt, and Levine (2005) show that competition and stability could go hand in hand. This analysis points to the importance of the quality of the banking system for this to hold.

iii. The effectiveness of capital regulation in discouraging risk-taking is negatively affected by competition ($q$) for low-quality banks but not so for high-quality banks. This follows from Proposition 3.2 that shows that for bad (good) banks capital is less (more) effective in encouraging investments in monitoring technology when competition heats up.

iv. Raising capital requirements positively affects the values of good banks when competition ($q$) is sufficiently high, and the average quality of the banking system is not too high (see upper bound on $\gamma$ in Proposition 3.3). The value of bad banks is always negatively affected. A way of testing this prediction is by looking at the valuation effects of the introduction of higher capital requirements.

v. Strengthening capital requirements encourages entry in banking markets that are of *intermediate* quality and sufficiently competitive (high $q$); otherwise it discourages entry. This prediction follows from Proposition 3.4.

vi. Increasing the number of players $N$ in the industry (such that average market share is diluted) reduces investments in monitoring technology and reduces the effectiveness of capital regulation for all banks. This prediction follows from Lemma 3.2 and Corollary 3.3 and comes from the scale economies in the monitoring technology. What this prediction implies is that augmenting competition via the number of players $N$ differs radically from augmenting competition via the openness parameter $q$. As predictions ii and iii show, increasing $q$ has a favorable effect on high-quality banks.

### 3.6.3 The impact of heterogeneity

Now I will connect my analysis to the literature that involves heterogeneity. I will first link it to contributions by Melitz (2003) and Syverson (2004) to the general IO literature. Subsequently, I will focus on the literature on heterogeneity in banking. In the trade literature Melitz (2003) analyzes the intra-industry effects of international trade using a Dixit and Stiglitz (1977) model of competition extended for firm heterogeneity. He shows that opening up borders for trade increases the total productivity of firms and hence total welfare, essentially by a reallocation of resources from low-quality to high-quality firms. I extend Melitz (2003) to the case of banking, and show that increasing capital requirements spurs the reallocation of market share toward better banks. If this effect is sufficiently strong, the banking industry as a whole may benefit from higher (although costly) capital requirements.

Syverson (2004) shows that pressure from the demand side affects firm heterogeneity in the industry. High competition for customers forces less productive firms to exit the market and, in doing so, lower firm heterogeneity. In my analysis, higher capital requirements "cleanse" the banking system by reducing the competitive strength of weak banks. This effectively neutralizes them, albeit there is strictly speaking no exit.

My analysis contrasts with the extant literature on banking competition that has primarily been analyzed in a symmetric context with equally capable banks. Keeley (1990) argues that competition lowers bank rents and may induce risk taking. My analysis shows that heterogeneity reinforces Keeley's result for low-quality banks but not necessarily for high-quality banks. That is, low-quality banks lose market share in competition with high-quality banks. Hence, the incentives of low-quality banks for prudent behavior (i.e., investments in monitoring technology) decline together with their franchise values. In contrast, good banks gain market share in competition with bad banks. This makes their investments in monitoring technology more valuable. Hence, competition can increase the stability of higher quality banking system.

These results are also in contrast with the results of Hellmann, Murdock, and Stiglitz (2000). They show that higher capital requirements may encourage bank risk taking. The reason is that when capital is costly, higher capital requirements negatively affect the franchise value of banks, and at such lower franchise value, a bank has less incentives to invest in monitoring technology. Note that these results critically depend on homogeneity in quality in the banking system. Repullo (2004) mediates their concern by showing that higher capital requirements contain risk shifting behavior of a bank directly and this effect is always stronger than the indirect effect through lowered franchise values. My analysis shows that higher capital requirements may increase a good bank's value although capital is costly. In particular, higher capital requirements protect higher quality banks by discouraging low-quality banks. This might be beneficial for the industry as a whole.

Only few contributions in the banking literature allow for heterogeneity in ability between banks. Kopecky and VanHoose (2006) argue that higher capital requirements may only affect a subset of banks with the lowest levels of capital. They show that higher capital requirements unambiguously increase the market loan rate and reduce aggregate lending, but have an ambiguous effect on loan quality. In my analysis higher capital requirements lead to less risk taking in banking. However, this effect is smaller for low-quality banks (see Proposition 3.2) and if there is more entry in the banking industry (see Corollary 3.3). Finally, Freixas, Hurkens, Morrison, and Vulkan (2004) show that low-quality borrowers would borrow at a bank with a weaker screening technology whereas high-quality borrowers would select a bank with a strong screening technology. This borrower "preference" effect is not part of my analysis. They do not focus on the interaction between capital regulation, deposit insurance and competition, which is the focus of my analysis.

## 3.7 Conclusions

I believe that this chapter contributes some key insights to understanding the interaction between competition and regulation. Heterogeneity between banks and the fixed costs of monitoring technology are important building blocks for understanding banking. I have shown that these lead to drastic shifts in market shares when competition heats up.

This analysis of competition between banks of different quality shows that capital regulation has a substantial impact on the competitive dynamics. The most striking conclusion from this analysis is that increasing costly capital requirements could encourage entry in markets that are sufficiently open for interbank competition. This result comes from the distortions that flat-rate deposit insurance introduces in banking. Implicitly, such deposit insurance benefits lower quality banks most, and makes them fiercer competitors than they otherwise would have been. Capital requirements are an effective regulatory tool that mitigates this distortion, and in doing so increases the value of entry. This points at a complementarity between capital regulation and deposit insurance that goes further than the typical insight that capital regulation mitigates the risk-taking incentives induced by deposit insurance. Capital requirements have a "cleansing" effect mitigating the artificial competitive advantage of low-quality banks that deposit insurance induces.

This insight also addresses a potential criticism of this analysis. I have assumed that capital requirements are binding; however, in the real world I often see banks operate at levels of capital significantly above the regulatory minimum (see Flannery and Rangan (2004)). Note however that in this analysis capital plays a crucial role in disciplining lower quality banks, and arguably precisely for these riskier banks capital regulation should be expected to be most binding. This analysis shows that capital regulation protects higher quality banks (and the financial system at large) from low-quality "fly-by-night" operators.

An arguably less surprising insight from this analysis is that competition weakens low-quality banking systems even further, including the effectiveness of capital regulation in such systems, while strengthening high-quality banking systems. This result confirms the anxiety that regulators may have about opening up their weak domestic banking markets to foreign competition; the stability consequences could be quite negative. However, it would be wrong to use this as an argument against opening up domestic markets. Rather, it points at the way in which domestic markets should be opened to competition. This chapter shows that having low-quality domestic banks compete with higher quality foreign banks will cause substantial instability. Anticipating a loss in market share, the weak domestic banks will cut back on investments in monitoring and in doing so elevate their riskiness. This may not happen if foreign entry leads to takeovers of domestic institutions. Such takeovers would not cause a reduction in monitoring because market share is no longer at risk.

In future work, the optimality of capital regulation and deposit insurance deserves further study. The optimality of these instruments in the face of even more competitive environment of banking is a key public policy issue. This chapter has taken these arrangements as given, and focused on their impact on the competitive dynamics. The good news that I have uncovered is that capital requirements help mitigate the competitive distortions that deposit insurance induces.

## 3.8 Appendix A

*Proof of Lemma 3.1*

Observe that there are $[N-1]$ other banks in the economy. The incumbent bank has a probability $q/[N-1]$ that it can compete for the borrowers of any one of these banks. Recall that each of these banks has $1/N$ borrower. Thus the expected number of other borrowers that the incumbent bank can make an offer to is $[N-1] \times \frac{q}{N-1} \times \frac{1}{N} = \frac{q}{N}$. ■

*Proof of Proposition 3.1*

Conjecture that good banks prevail over incumbent bad banks and, when banks of the same type compete, the incumbency advantage prevails. Using (3.7) one has

$$V_B = \frac{1-q}{N}[-k + \frac{\nu_B X}{\rho}] + \frac{q}{\rho N}[1-\gamma]\{S + [\nu_B - \nu_B^*]X\} - c\frac{[\nu_B - \underline{\nu}_B]^2}{2}, \quad (3.11)$$

$$V_G = \frac{1-q}{N}[-k + \frac{\nu_G X}{\rho}] + \frac{q\gamma}{\rho N}\{S + [\nu_G - \nu_G^*]X\} + \frac{2q}{\rho N}[1-\gamma][\nu_G - \nu_B^*]X - c\frac{[\nu_G - \underline{\nu}_G]^2}{2} \quad (3.12)$$

The first terms in (3.11) and (3.12) represent the profits of the incumbent bank from its borrower without a competing offer. This happens with probability $[1-q]$. With probability $q$, the borrower finds a competing bank. A bad incumbent bank only retains its borrower when he gets the second offer from another bad bank. This happens w.p. $q[1-\gamma]$, see the second term in (3.11). A good bank can retain its incumbent borrower when he gets an offer from another good bank. This occurs w.p. $q\gamma$, see the second term in (3.12). In addition, a good bank retains its incumbent borrower when he receives an offer from a bad bank. This happens with probability $q[1-\gamma]$. Moreover, it can take new borrowers from other bad banks, also with the same probability $q[1-\gamma]$; see the third term in (3.12).

Implicitly in (3.7), (3.11), and (3.12), I have used the assumption that banks are always willing to bid for borrowers. That is, borrowers' projects are sufficiently profitable such that banks are willing to lend. Whether a bank succeeds in holding on to, or acquiring, a borrower depends on its own strength (quality and investment in monitoring technology), the strength of its competitor, and the incumbency advantage. A sufficient condition for this is

$$-k + \frac{\underline{\nu}_B X}{\rho} - \frac{S}{\rho} > 0. \quad (3.13)$$

The condition in (3.13) implies that a bad bank at its minimum intrinsic monitoring level $\underline{\nu}_B$ could profitably lend to a borrower of another bank (but loose out in the competition!). I further use this condition in the proof of Lemma 3.2.

Each type maximizes its value, holding the strategy of the other type fixed. Use (3.11) and (3.12) to obtain

$$\frac{\partial V_B}{\partial \nu_B}(\nu_B^*) = \frac{1-q}{\rho N}X + \frac{q}{\rho N}[1-\gamma]X - c[\nu_B^* - \underline{\nu}_B] = 0, \tag{3.14}$$

$$\frac{\partial V_G}{\partial \nu_G}(\nu_G^*) = \frac{1-q}{\rho N}X + \frac{q}{\rho N}\gamma X + 2\frac{q}{\rho N}[1-\gamma]X - c[\nu_G^* - \underline{\nu}_G] = 0, \tag{3.15}$$

which imply (3.8) and (3.9). Note from (3.14) and (3.15) that $\frac{\partial V_B}{\partial \nu_B}(\nu_B = 0) > 0$ and $\frac{\partial V_G}{\partial \nu_G}(\nu_G = 0) > 0$. This shows that each bank's investment in monitoring technology is positive. Note also that the second-order conditions are negative. Thus, the optimal levels of monitoring are (3.8) and (3.9). Insert $\nu_B = \nu_B^*$ and $\nu_G = \nu_G^*$ from (3.8) and (3.9) into (3.11) and (3.12) to obtain

$$V_B^* = \frac{1-q}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + q[1-\gamma]\frac{S}{\rho N} + \frac{X^2}{2c[\rho N]^2}\{1 - 2q + q^2[2-\gamma]\gamma\}, \tag{3.16}$$

$$V_G^* = \frac{1-q}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + q\gamma\frac{S}{\rho N} + \frac{1+q[1-2\gamma]}{\rho N}[\underline{\nu}_G - \underline{\nu}_B]X + \frac{X^2}{2c[\rho N]^2}\{[1-q]^2 - q^2\gamma^2\}. \tag{3.17}$$

Now I check that (3.8) and (3.9) indeed satisfy my conjectures. Assumption 3.1 guarantees that $[\underline{\nu}_G - \underline{\nu}_B]X > S$. Use this and (3.8) and (3.9) to obtain $[\nu_G^* - \nu_B^*]X > S$, hence

$$\nu_G^* X - S > \nu_B^* X. \tag{3.18}$$

The expression in (3.18) implies that a good bank prevails over an incumbent bad bank; obviously, then, an incumbent good bank prevails over a competing bad bank.

I show next that for a good bank it is not profitable to increase its level of monitoring sufficiently to steal borrowers from other good banks; that is, to deviate from $\nu_G^*$ to $\hat{\nu}_G \gg \nu_G^*$. Use (3.7) to compute the value of a good bank that chooses the level of monitoring $\hat{\nu}_G$,

$$\hat{V}_G = \frac{1-q}{N}[-k + \frac{\hat{\nu}_G X}{\rho}] + 2\frac{q}{\rho N}\gamma[\hat{\nu}_G - \nu_G^*]X + 2\frac{q}{\rho N}[1-\gamma][\hat{\nu}_G - \nu_B^*]X - c\frac{[\hat{\nu}_G - \underline{\nu}_G]^2}{2}. \tag{3.19}$$

Maximizing (3.19) w.r.t. $\hat{\nu}_G$ yields

$$\hat{\nu}_G^* = \underline{\nu}_G + [1+q]X/c\rho N. \tag{3.20}$$

Insert $\hat{\nu}_G = \hat{\nu}_G^*$ from (3.20) into (3.19) and use (3.8) and (3.9) to obtain

$$\hat{V}_G^* = \frac{1-q}{N}\{-k + \frac{\underline{\nu}_B X}{\rho}\} + \frac{1+q[1-2\gamma]}{\rho N}[\underline{\nu}_G - \underline{\nu}_B]X + \frac{X^2}{2c[\rho N]^2}[1-q]^2. \tag{3.21}$$

To show that the deviation to $\hat{\nu}_B^*$ is not profitable, observe from (3.21) and (3.17) that

$$V_G^* - \hat{V}_G^* = \frac{q\gamma}{\rho N}S - \frac{X^2}{2c[\rho N]^2}q^2\gamma^2. \tag{3.22}$$

Because $\frac{X^2}{c\rho N} < S$ (see Assumption 3.1), it immediately follows that the expression in (3.22) is positive; hence a good bank will not steal borrowers from other good banks.

I now show that a bad bank does not have an incentive to increase its investment in monitoring from $\nu_B^*$ to $\hat{\nu}_B \gg \nu_B^*$ to attract borrowers from other bad banks. From (3.7) one has

$$\hat{V}_B = \frac{1-q}{N}[-k + \frac{\hat{\nu}_B X}{\rho}] + 2\frac{q}{\rho N}[1-\gamma][\hat{\nu}_B - \nu_B^*]X - \frac{c}{2}\hat{\nu}_B^2. \tag{3.23}$$

Maximizing (3.23) w.r.t. $\hat{\nu}_B$ yields

$$\hat{\nu}_B^* = \{1 + q[1-2\gamma]\}X/c\rho N + \underline{\nu}_B. \tag{3.24}$$

Insert $\hat{\nu}_B = \hat{\nu}_B^*$ from (3.24) into (3.23) and use (3.8) to obtain

$$\hat{V}_B^* = \frac{1-q}{N}\{-k + \frac{\underline{\nu}_B X}{\rho}\} + \frac{X^2}{2c[\rho N]^2}[1-q]^2. \tag{3.25}$$

Observe that the deviation to $\hat{\nu}_B^*$ is not profitable (use (3.16) and (3.25)):

$$V_B^* - \hat{V}_B^* = q[1-\gamma]\frac{S}{\rho N} - \frac{X^2}{2c[\rho N]^2}q^2[1-\gamma]^2. \tag{3.26}$$

Because $\frac{X^2}{c\rho N} < S$ (see Assumption 3.1), it follows that (3.26) is positive, and a bad bank will not steal borrowers from other bad banks.

I now show that an incumbent bad bank has no incentive to increase its investment in monitoring technology to $\tilde{\nu}_B \gg \nu_B^*$ to hold on to its borrower when competing with a good bank. If a bad bank chooses $\tilde{\nu}_B$, one has (use (3.7)),

$$\tilde{V}_B = \frac{1-q}{N}[-k + \frac{\tilde{\nu}_B X}{\rho}] + \frac{q}{\rho N}\gamma\{S + [\tilde{\nu}_B - \nu_G^*]X\} + 2\frac{q}{\rho N}[1-\gamma][\tilde{\nu}_B - \nu_B^*]X - c\frac{\tilde{\nu}_B^2}{2}. \tag{3.27}$$

Maximizing (3.27) w.r.t. $\tilde{\nu}_B$ gives

$$\tilde{\nu}_B^* = \{1 + q[1-\gamma]\}X/c\rho N + \underline{\nu}_B. \tag{3.28}$$

Insert $\tilde{\nu}_B = \tilde{\nu}_B^*$ from (3.28) into (3.27) and use (3.8) and (3.9) to obtain

$$\tilde{V}_B^* = \frac{1-q}{N}\{-k + \frac{\underline{\nu}_B X}{\rho}\} + q\gamma\frac{S - [\underline{\nu}_G - \underline{\nu}_B]X}{\rho N} + \frac{X^2}{2c[\rho N]^2}\{[1-q]^2 - q^2\gamma^2\}. \tag{3.29}$$

Use (3.16) and (3.29) to see that

$$V_B^* - \tilde{V}_B^* = q\gamma \frac{[\underline{\nu}_G - \underline{\nu}_B]X - S}{\rho N} + q[1-\gamma]\frac{S}{\rho N} - \frac{X^2}{2c[\rho N]^2}q^2[1-2\gamma]. \tag{3.30}$$

Because $\frac{X^2}{c\rho N} < S$ and $S < [\underline{\nu}_G - \underline{\nu}_B]X$ (see Assumption 3.1), one can see that (3.30) is positive, and hence an incumbent bad bank will not try to hold on to its borrower when competing with a good bank.

Finally, note from (3.8) and (3.9) that the following condition guarantees that $\nu_G^*$ and $\nu_B^*$ are in the interior for all $q, \gamma \in [0, 1]$:

$$2X/c\rho N + \underline{\nu}_G < 1. \tag{3.31}$$

This condition, the restriction (3.13), and Assumption 3.1 are easily simultaneously satisfied (e.g., choose $X$ high enough to satisfy (3.13), and then choose sufficiently high $N$ to satisfy Assumption 3.1 and (3.31)). This completes the proof. ■

*Proof of Corollary 3.1*

Differentiate (3.8) and (3.9) w.r.t. $k$ and recall that $X \equiv Y - [1-k]r_D$, to obtain

$$\frac{\partial \nu_B^*}{\partial k} = \frac{[1-q\gamma]r_D}{c\rho N} \quad \text{and} \quad \frac{\partial \nu_G^*}{\partial k} = \frac{\{1+q[1-\gamma]\}r_D}{c\rho N}, \tag{3.32}$$

which are both positive. ■

*Proof of Corollary 3.2*

Differentiate (3.8) and (3.9) w.r.t. $q$, to obtain

$$\frac{\partial \nu_B^*}{\partial q} = -\frac{\gamma}{c\rho N}X < 0 \quad \text{and} \quad \frac{\partial \nu_G^*}{\partial q} = \frac{1-\gamma}{c\rho N}X > 0. \tag{3.33}$$

Thus, competition increases the investment in monitoring technology for a good bank, but not for a bad bank. ■

*Proof of Proposition 3.2*

Differentiate both expressions in (3.32) w.r.t. $q$ to get

$$\frac{\partial^2 \nu_B^*}{\partial q \partial k} = -\frac{\gamma}{c\rho N}r_D < 0 \quad \text{and} \quad \frac{\partial^2 \nu_G^*}{\partial q \partial k} = \frac{1-\gamma}{c\rho N}r_D > 0.$$

Hence, competition elevates the effectiveness of capital regulation for a good bank, but not for a bad bank. ■

*Proof of Proposition 3.3*

Differentiating (3.16) w.r.t. $k$ and rearranging yields

$$\frac{\partial V_B^*}{\partial k} \quad = \quad [\underline{\nu}_G - \underline{\nu}_B]\frac{r_D}{\rho N}\{-[1-q]\alpha - \zeta[1-q^2[2-\gamma]\gamma]\}, \tag{3.34}$$

where I have used the following definitions

$$\alpha \equiv \frac{1 - r_D\underline{\nu}_B/\rho}{[\underline{\nu}_G - \underline{\nu}_B]r_D/\rho} - \frac{2X}{c\rho N[\underline{\nu}_G - \underline{\nu}_B]} \quad \text{and} \quad \zeta \equiv \frac{X}{c\rho N[\underline{\nu}_G - \underline{\nu}_B]} > 0. \tag{3.35}$$

Rewrite $\alpha$ as $\alpha = 1 + \frac{1 - r_D\underline{\nu}_G/\rho - 2Xr_D/c\rho^2 N}{[\underline{\nu}_G - \underline{\nu}_B]r_D/\rho}$. Substitute for $\underline{\nu}_G$ from (3.31) to obtain

$$\alpha > 1 + \frac{1 - \frac{r_D}{\rho}[1 - 2X/c\rho N] - \frac{2Xr_D}{c\rho^2 N}}{[\underline{\nu}_G - \underline{\nu}_B]r_D/\rho}. \tag{3.36}$$

Rearrange (3.36) to get $\alpha > 1 + \frac{1 - r_D/\rho}{[\underline{\nu}_G - \underline{\nu}_B]r_D/\rho} > 1$. Note from the definition of $\zeta$ in (3.35) and the fact that $\frac{X^2}{c\rho N} < [\underline{\nu}_G - \underline{\nu}_B]X$ (see Assumption 3.1) that $\zeta < 1$. Thus,

$$\alpha > 1 \quad \text{and} \quad 0 < \zeta < 1. \tag{3.37}$$

Note that $\gamma[2-\gamma]$ is maximized for $\gamma = 1$. Use this and (3.37) in (3.34) to see that $-[1-q]\alpha - \zeta[1-q^2[2-\gamma]\gamma] \leq -[1-q]\alpha - \zeta[1-q^2] < 0$. This implies that $\frac{\partial V_B^*}{\partial k} < 0$, and proves that the value of a bad bank is always negatively affected by stricter capital requirements.

For a good bank, use (3.17) to see that

$$\frac{\partial V_G^*}{\partial k} \quad = \quad [\underline{\nu}_G - \underline{\nu}_B]\frac{r_D}{\rho N}\{-[1-q]\alpha + 1 + q[1-2\gamma] - \zeta[1-q^2[1-\gamma^2]]\}. \tag{3.38}$$

Observe that for $q = 0$, the expression (3.38) simplifies to $\left.\frac{\partial V_G^*}{\partial k}\right|_{q=0} = [\underline{\nu}_G - \underline{\nu}_B]\frac{r_D}{\rho N}[-\alpha + 1 - \zeta]$, which (using (3.37)) is always negative. In addition, note that

$$\left.\frac{\partial V_G^*}{\partial k}\right|_{q=1,\gamma=0} = 2[\underline{\nu}_G - \underline{\nu}_B]\frac{r_D}{\rho} > 0.$$

Observe that $\frac{\partial V_G^*}{\partial k}$ is monotonically increasing in $q$ and decreasing in $\gamma$. Hence by continuity one can see that capital regulation increases the value of a good bank for $q$ sufficiently high and $\gamma$ sufficiently low. This completes the proof. ■

*Proof of Lemma 3.2*

I need to show that $\nu_B^*$ and $\nu_G^*$ are decreasing in $N$. Differentiate (3.8) and (3.9) w.r.t. $N$ to get

$$\frac{\partial \nu_B^*}{\partial N} = -\frac{[1-q\gamma]X}{c\rho N^2} - \frac{\gamma X}{c\rho N^2}\frac{\partial q}{\partial N}, \tag{3.39}$$

$$\frac{\partial \nu_G^*}{\partial N} = -\frac{\{1+q[1-\gamma]\}X}{c\rho N^2} - \frac{1-\gamma}{c\rho N}[\frac{q}{N} - \frac{\partial q}{\partial N}]. \tag{3.40}$$

Note that the ratio $q/N$ is subject to the regularity condition, $\frac{\partial[q/N]}{\partial N} < 0$, implying that the expected number of other borrowers that the incumbent bank can make an offer to is decreasing in $N$. This should hold because, while $q$ is increasing in $N$, the market – with a higher $N$ – must be shared among more competing banks, reducing each bank's share. Transform $\frac{\partial[q/N]}{\partial N} < 0$ to get $\frac{q}{N} - \frac{\partial q}{\partial N} > 0$. Use this and $\frac{\partial q}{\partial N} > 0$ together with (3.39) and (3.40) to see that $\frac{\partial \nu_B^*}{\partial N} < 0$ and $\frac{\partial \nu_G^*}{\partial N} < 0$.

Now I prove that $V_B^*$ and $V_G^*$ are decreasing in $N$. Differentiate (3.16) and (3.17) w.r.t. $N$ to obtain

$$\frac{\partial V_B^*}{\partial N} = -\frac{1-q}{N^2}[-k + \frac{\underline{\nu}_B X - S}{\rho}] - [1-q\gamma]\frac{S}{\rho N^2} - \frac{X^2}{2c\rho^2 N^3}\{1-2q+q^2[2-\gamma]\gamma\}, \tag{3.41}$$

$$\frac{\partial V_G^*}{\partial N} = -\frac{1-q}{N^2}[-k + \frac{\underline{\nu}_B X - S}{\rho}] - \{1-q[1-\gamma]\}\frac{S}{\rho N^2} - \frac{\{1+q[1-2\gamma]\}}{\rho N^2}[\underline{\nu}_G - \underline{\nu}_B]X$$
$$-\frac{X^2}{2c\rho^2 N^3}\{1-2q+q^2[1-\gamma^2]\}. \tag{3.42}$$

I now make the following substitutions. First, recall from (3.13) that $-k + [\underline{\nu}_B X - S]/\rho > 0$. I will use the substitution $S < -\rho k + \underline{\nu}_B X$. Second, use Assumption 3.1, in particular substitute for $S$ and $[\underline{\nu}_G - \underline{\nu}_B]X$ the expression $\frac{X^2}{c\rho N}$. All these substitutions in (3.41) and (3.42) give

$$\frac{\partial V_B^*}{\partial N} < -\frac{X^2}{2c\rho^2 N^3}\{2[1-q\gamma] + 1 - 2q + q^2[2-\gamma]\gamma,$$

$$\frac{\partial V_G^*}{\partial N} < -\frac{X^2}{2c\rho^2 N^3}\{2[1-q[1-\gamma]] + 1 + q[1-2\gamma] + 1 - 2q + q^2[1-\gamma^2]\}.$$

This can be further rearranged to

$$\frac{\partial V_B^*}{\partial N} < -\frac{X^2}{2c\rho^2 N^3}\{1-q\gamma + [1-q][2-q\gamma] + q^2\gamma[1-\gamma]\}, \tag{3.43}$$

$$\frac{\partial V_G^*}{\partial N} < -\frac{X^2}{2c\rho^2 N^3}\{1-q[1-\gamma] + 1 - q\gamma + 2[1-q] + q^2[1-\gamma^2]\}. \tag{3.44}$$

Because $q$ and $\gamma$ are limited to the interval $[0,1]$, the expressions in (3.43) and (3.44) are always negative. This concludes the proof. ■

*Proof of Proposition 3.4*

Differentiating (3.10) w.r.t. $k$ one can obtain

$$-\{[1-\gamma]\frac{\partial \bar{V}_B^*}{\partial N}+\gamma\frac{\partial \bar{V}_G^*}{\partial N}\}\frac{\partial N}{\partial k}=[1-\gamma]\frac{\partial \bar{V}_B^*}{\partial k}+\gamma\frac{\partial \bar{V}_G^*}{\partial k}. \tag{3.45}$$

I know from Lemma 3.2 that $\frac{\partial \bar{V}_B^*}{\partial N}<0$ and $\frac{\partial \bar{V}_G^*}{\partial N}<0$. Hence, the sign of $\frac{\partial N}{\partial k}$ equals the sign of the right side of (3.45); that is, higher capital induces more entry iff

$$[1-\gamma]\frac{\partial \bar{V}_B^*}{\partial k}+\gamma\frac{\partial \bar{V}_G^*}{\partial k}>0. \tag{3.46}$$

Use (3.34) and (3.38) to simplify (3.46) to get that higher capital induces more entry iff

$$DV(\gamma,q)>0, \tag{3.47}$$

where

$$DV(\gamma,q)\equiv -[1-q]\alpha+\gamma\{1+q[1-2\gamma]\}+\zeta[-1+3q^2\gamma[1-\gamma]]\}, \tag{3.48}$$

$\alpha$ and $\zeta$ as defined in (3.35), and conditions in (3.37).

I first observe what impact higher capital has on entry at a fixed $q$. Observe that for a fixed $q$ the function $DV(\gamma,q)$ for $\gamma$ is an inverse parabola. Note that $DV(\gamma=0,q)=-[1-q]\alpha-\zeta<0$. In addition, I have $DV(\gamma=1,q)=-[1-q][\alpha-\gamma]-\zeta<0$. This means that higher capital always reduces entry at $\gamma=0$ and $\gamma=1$. This and the parabolic shape of the function $DV(\gamma,q)$ implies the following for the intermediate values of $\gamma$. There exist solutions to the equation $DV(\gamma,q)=0$ denoted by $\gamma_1(q)\in[0,1]$ and $\gamma_2(q)\in[0,1]$ iff $DV(\gamma,q)>0$ for at least one $\gamma\in[0,1]$.

Now I show that $DV(\gamma,q)>0$ for at least one $\gamma\in[0,1]$ iff competition $q$ is high enough; that is, $q\geq\bar{q}$. First, note that

$$DV(\gamma,q=0)<-\alpha+\gamma-\zeta, \tag{3.49}$$

which is negative for all $\gamma\in[0,1]$. Second, observe that

$$DV(\gamma=1/2,q=1)=1/2+\zeta[-1+3/4]>0. \tag{3.50}$$

These two facts and the monotonicity of $DV(\gamma,q)$; that is,

$$\frac{\partial DV(\gamma,q)}{\partial q}=\alpha+\gamma[1-2\gamma]+6\zeta q\gamma[1-\gamma]>0, \tag{3.51}$$

imply that there exist a certain $\bar{q}$ such that $DV(\gamma,q)<0$, that is, higher capital discourages entry, for all $\gamma\in[0,1]$ if $q<\bar{q}$. For high competition (i.e., $q\geq\bar{q}$), I have two regions of $\gamma$. In the first region (i.e., $\gamma\in[0,\gamma_1(q))\cup(\gamma_2(q),1]$), I have $DV(\gamma,q)<0$ and higher capital discourages entry. In the second region (i.e., $\gamma\in[\gamma_1(q),\gamma_2(q)]$), I have $DV(\gamma,q)\geq 0$ and higher capital induces more entry. ■

*Proof of Corollary 3.3*

First, I compute the impact of entry on monitoring of bad banks. Partially differentiate (3.8) w.r.t. $k$ to obtain

$$\frac{\partial \nu_B^*}{\partial k} = -\frac{X}{c\rho N}[\gamma \frac{\partial q}{\partial N} + \frac{1-\gamma q}{N}]\frac{\partial N}{\partial k} + \frac{1-q\gamma}{c\rho N} r_D. \tag{3.52}$$

Observe that (3.52) equals (3.32), except for the additional term

$$-\frac{X}{c\rho N}[\gamma \frac{\partial q}{\partial N} + \frac{1-\gamma q}{N}]\frac{\partial N}{\partial k}. \tag{3.53}$$

Observe that $\gamma \frac{\partial q}{\partial N} + \frac{1-q\gamma}{N} > \gamma \left[\frac{\partial q}{\partial N} - \frac{q}{N}\right]$, which is always positive because $\frac{\partial [q/N]}{\partial N} < 0$ (see Lemma 3.2). This means that (3.53) is positive as long as $\frac{\partial N}{\partial k} < 0$ and negative if $\frac{\partial N}{\partial k} > 0$. Thus, the monitoring incentives induced by additional capital are strengthened if capital discourages entry, and weakened if capital encourages entry.

I proceed similarly for good banks. Differentiate (3.9) w.r.t. $k$ to obtain

$$\frac{\partial \nu_G^*}{\partial k} = \frac{X}{c\rho N^2}\{[N\frac{\partial q}{\partial N} - q][1-\gamma] - 1\}\frac{\partial N}{\partial k} + \frac{1+q[1-\gamma]}{c\rho N} r_D. \tag{3.54}$$

The expression in (3.54) is equal to (3.32), except for the additional term

$$\frac{X}{c\rho N^2}\{[N\frac{\partial q}{\partial N} - q][1-\gamma] - 1\}\frac{\partial N}{\partial k}. \tag{3.55}$$

As in the proof of Lemma 3.2, $\frac{\partial q/N}{\partial N} < 0$ implies $\frac{q}{N} - \frac{\partial q}{\partial N} > 0$, hence $N\frac{\partial q}{\partial N} - q < 0$. Thus, (3.55) is positive as long as $\frac{\partial N}{\partial k} < 0$ and negative if $\frac{\partial N}{\partial k} > 0$. This means that the effectiveness of capital requirements increases (or decreases) when capital requirements induce less (or more) entry. ■

*Proof of Proposition 3.5*

Assume that Assumption 3.1 holds for any $k$ and also the condition (3.31) for an interior optimum. Following Proposition 3.4, positive capital is desired in two cases. In the first case, $DV(\gamma, q|k=0) > 0$, where $DV(\gamma, q|k)$ is defined as $DV(\gamma, q)$ in (3.48) and constants $\alpha$ and $\zeta$ are computed at specific $k$ such that the conditions in (3.37) hold. In the second case, $DV(\gamma, q|k=0) < 0$, yet positive at higher values of $k$. Because (3.16) and (3.17) are quadratic functions of $k$, the value of a bank in this case has a maximum either at $k=0$ or at $k=1$. In fact, by symmetry of the quadratic functions the value of a bank at $k=1$ exceeds that at $k=0$ if $DV(\gamma, q|k=\frac{1}{2}) > 0$. In sum, banks want to have positive capital requirements as long as $DV(\gamma, q|k=0) > 0$ or $DV(\gamma, q|k=\frac{1}{2}) > 0$. This is the case for a sufficiently high $q$ (i.e., $q > \hat{q}$) and for an intermediate $\gamma$ (see Proposition 3.4). ■

*Proof of Proposition 3.6*

First, I show that the welfare optimal level of capital is strictly less than 1. I compute the welfare contribution of a bank by summing together the bank and borrower surplus, the expected loss to the deposit insurance fund, and entry costs. One has

$$W_B(\nu_B) = \frac{1-q\gamma}{N}[-k+\frac{\nu_B X}{\rho}] - c\frac{[\nu_B-\underline{\nu}_B]^2}{2} - \frac{[1-\nu_B][1-k]r_D[1-q\gamma]}{\rho N} - F, \quad (3.56)$$

$$W_G(\nu_G) = \frac{1+q[1-\gamma]}{N}[-k+\frac{\nu_G X}{\rho}] - c\frac{[\nu_G-\underline{\nu}_G]^2}{2} - \frac{[1-\nu_G][1-k]r_D\{1+[1-q\gamma]\}}{\rho N}$$
$$-F - \frac{q[1-\gamma]}{\rho N}S. \quad (3.57)$$

Social welfare is the sum of the welfare contributions of the individual banks; that is, $TW(\nu_B,\nu_G) = N\{\gamma W_G(\nu_B) + [1-\gamma]W_B(\nu_G)\}$. Use (3.56) and (3.57) and maximize w.r.t. $k$ to obtain

$$\frac{\partial TW}{\partial k}(k=k_{reg}) = \frac{1}{N}[-1+\frac{r_D}{\rho}] + \frac{[1-k_{reg}]r_D^2}{c\rho^2 N} = 0.$$

Note that $\frac{\partial^2 TW}{\partial k^2} < 0$. Solve for the welfare optimal capital requirement $k_{reg}$ to obtain $k_{reg} = \max(0, 1 - \frac{c\rho^2 N[1-\frac{r_D}{\rho}]}{\{1+\gamma[1-\gamma]q^2\}r_D^2})$, which is strictly less than 1.

I now show that the privately optimal capital requirements are increasing in $q$ and reach 1 for a sufficiently high $q$ and intermediate $\gamma$. Again Assumption 3.1 and (3.31) hold for any $k$. First, if $DV(\gamma,q|k=0) > 0$, where $DV(\gamma,q|k)$ is defined in (3.48), optimal capital requirements are positive. Note from (3.50) that $DV(\gamma=1/2,q=1|k) > 0$ and from (3.51) that $\frac{\partial DV(\gamma,q|k)}{\partial q} > 0$ for all $k$. This guarantees that bank values are maximized at capital requirements equal 1, for sufficiently high $q$ and intermediate $\gamma$. Because $k_{reg} < 1$, there exists a $\bar{q} \geq \hat{q}$ such that for $q > \bar{q}$ and intermediate $\gamma$ banks want the regulator to choose capital requirements equal 1, which is above the welfare optimal level. Second, if $DV(\gamma,q|k=0) < 0$ and $DV(\gamma,q|k=\frac{1}{2}) > 0$ banks want the regulator to set capital requirements equal to 1 (see Proposition 3.5) and $\bar{q} = \hat{q}$. ■

## 3.9 Appendix B: Social Welfare Analysis

In this appendix I analyze in more detail what kind of capital regulation is optimal from the social welfare maximization point of view. First, I consider a situation in which banks are of equal type. Second, I extend this analysis to good and bad banks.

### 3.9.1 Banks of equal type

I first assume that banks are of the same type; that is, $\underline{\nu} = \underline{\nu}_B = \underline{\nu}_G$. Social welfare consists of four parts; namely, the values of banks, consumer surplus, cost to the deposit insurance fund and costs of entry. First, I compute the values of equal type banks. Banks cannot take each other's market share (see Proposition 3.1). I can write the value of a bank with monitoring level $\nu$ competing with a bank with monitoring $\nu^*$ as (see (3.7))

$$V(\nu) = \frac{1-q}{N}[-k+\frac{\nu X}{\rho}] + \frac{q}{\rho N}\{[\nu-\nu^*]X] + S\} - c\frac{[\nu-\underline{\nu}]^2}{2}. \quad (3.58)$$

Second, I evaluate the consumer surplus. In the case of no competition, the incumbent bank leaves borrowers with no profits. However, in the case of competition borrowers receive the total profits of the project net of funding costs (i.e., $\nu X$) lowered by the profits of their bank $[\nu - \nu^*]X + S$. Hence, the consumer surplus of borrowers of one bank equals

$$CS(\nu) = [1-q] \times 0 + \frac{q}{\rho N}\{\nu X - [\nu - \nu^*]X - S\}. \tag{3.59}$$

Third, in expectations the deposit insurance fund carries the following cost. The bank fails with probability $1-\nu$. In this case, the deposit insurance fund repays deposits in the total value of $\frac{[1-k]r_D}{\rho N}$. The expected loss to the deposit insurance fund $L$ per bank is

$$L(\nu) = [1-\nu]\frac{[1-k]r_D}{\rho N}. \tag{3.60}$$

Fourth, each bank when entering incurs entering costs $F$. In sum, the social welfare that one bank generates is

$$W(\nu) = V(\nu) + CS(\nu) - L(\nu) - F. \tag{3.61}$$

Insert (3.58), (3.59) and (3.60) into (3.61) to compute the welfare contribution of one bank

$$W(\nu) = \frac{1}{N}[-k + \frac{\nu X}{\rho}] - c\frac{[\nu - \underline{\nu}]^2}{2} - \frac{[1-\nu][1-k]r_D}{\rho N} - F. \tag{3.62}$$

Social welfare is $TW = N \times W$; that is,

$$TW(\nu) = [-k + \frac{\nu X}{\rho}] - Nc\frac{[\nu - \underline{\nu}]^2}{2} - \frac{[1-\nu][1-k]r_D}{\rho} - NF. \tag{3.63}$$

Banks invest in monitoring to maximize their values in (3.58), which gives the optimal monitoring levels[29]

$$\nu^* = \frac{X}{c\rho N} + \underline{\nu}. \tag{3.64}$$

I can show the following lemma.

**Lemma 3.3.** *The individually chosen level of monitoring $\nu^*$ is lower than the welfare optimal level of monitoring $\nu_{reg}$.*

*Proof:* Maximize (3.63) w.r.t. $\nu$ to obtain

$$\frac{\partial TW}{\partial \nu}(\nu_{reg}) = \frac{Y}{c\rho N} - c[\nu_{reg} - \underline{\nu}] = 0,$$

which yields

$$\nu_{reg} = \frac{Y}{c\rho N} + \underline{\nu}. \tag{3.65}$$

[29]Note that (3.64) can be obtained from the heterogeneous model by saying that one type of banks prevails. Set $\gamma = 0$ and $\underline{\nu}_B = \underline{\nu}$ in (3.8) or set $\gamma = 1$ and $\underline{\nu}_G = \underline{\nu}$ in (3.9). In the case of $\gamma = 0$, there are only bad banks and in the case of $\gamma = 1$, only good banks exist.

Note that $\frac{\partial^2 TW}{\partial \nu^2} < 0$. Compare (3.65) to (3.64) to note that $\nu_{reg} > \nu^*$ because $Y > X$. ■

The intuition for Lemma 3.3 is that the profit maximization of each bank differs from the welfare maximization of the regulator. In addition to maximizing profits of banks and borrowers' surplus, the regulator also strives to minimize the expected losses to the deposit insurance fund. Increasing monitoring of a bank above the privately optimal level augments a bank's stability and social welfare.

The regulator, however, can have difficulties in directly controlling a bank's monitoring levels. Although it is in the bank's interest to expose its monitoring capabilities to its borrowers, it can hide the level of monitoring from the regulator. In this case, the regulator cannot directly control banks' monitoring.

The regulator can influence banks' monitoring levels indirectly through capital regulation. Setting sufficiently high capital requirements induces banks to invest more in monitoring. The regulator solves the following maximization problem in order to maximize social welfare.

$$k_{reg} = \arg\max_{k} TW(\nu^*). \tag{3.66}$$

I show the following result.

**Proposition 3.7.** *With one type of bank the welfare optimal capital requirement $k_{reg}$ is given by*

$$k_{reg} = \max(0, 1 - \frac{c\rho^2 N}{r_D^2}[1 - \frac{r_D}{\rho}]). \tag{3.67}$$

*Proof:* Insert (3.64) for $\nu$ into (3.63) to obtain

$$TW = [-k + \frac{\nu X}{\rho}] + \frac{X^2}{2c\rho^2 N^2} - \frac{[1-\nu][1-k]r_D}{\rho N} + \frac{X}{c\rho^2 N^2}[1-k]r_D - NF.$$

Rearrange to obtain

$$TW = [-k + \frac{\nu Y - [1-k]r_D}{\rho}] + \frac{Y^2 - r_D^2[1-k]^2}{2c\rho^2 N} - NF. \tag{3.68}$$

Maximize (3.68) w.r.t. $k$ to obtain

$$\frac{\partial TW}{\partial k}(k = k_{reg}) = \frac{1}{N}[-1 + \frac{r_D}{\rho}] + \frac{[1-k_{reg}]r_D^2}{c\rho^2 N} = 0.$$

Solve for $k_{reg}$ to obtain (3.67). Note also that the $\frac{\partial^2 TW}{\partial k^2} < 0$. ■

The intuition for Proposition 3.7 is the following. The presence of deposit insurance induces a bank to underestimate risk and invest in monitoring less than is welfare optimal. The regulator, which also considers possible losses to the deposit insurance fund, wants banks to monitor more and to be safer. Setting capital requirements above zero achieves this goal.

The welfare optimal capital requirement is greater than zero if capital is a relatively cheap source of financing; that is, if the cost of capital approaches the cost of debt, $\rho <$

$\frac{r_D}{2}[1+\sqrt{\frac{4}{cN}+1}]$. Only in this case, the benefits of higher monitoring prevail over the additional costs of financing.

I can show the following.

**Corollary 3.4.** *If banks are of the same type, the welfare optimal capital requirement $k_{reg}$ does not change with competition $q$.*

*Proof:* The welfare optimal capital requirement is given by (see (3.67))

$$k_{reg} = \max(0, 1 - \frac{c\rho^2 N}{r_D^2}[1 - \frac{r_D}{\rho}]). \tag{3.69}$$

Note that $k_{reg}$ is not a function of competition $q$. ■

Social welfare is independent of interbank competition $q$ for banks of equal quality. This is because competition has no other role but to transfer rents from banks to their borrowers which has no effect on a social welfare. Despite high competition, banks hold on to their market share and competition has no influence on the monitoring levels (see (3.64)). Recall that capital works as a mechanism to increase monitoring levels. Because monitoring levels are unchanged, the regulator has no incentives to change its level of optimal capital requirements. Technically speaking, the welfare function does not change with competition; consequently, the level of capital that maximizes this function is independent of competition.

I now observe the individual choice of capitalization of banks. Banks first commit to capital; and after commitments are observable, they invest in monitoring.

**Corollary 3.5.** *If banks could individually choose their level of capital, they would have chosen zero capital.*

*Proof:* Assume that all banks have the conjectured level of capital $k_0$. I show that any bank will deviate to the lowest level of capital. First, I show that banks choose zero capital under the conjecture that they cannot take new borrowers. I compute the value of a bank that deviates from capital $k_0$ to $\hat{k}$ whereas all other banks stick to $k_0$. The deviating bank chooses monitoring $\hat{\nu}$ whereas all others choose $\nu^*$ as given in (3.64).

A value that the deviating bank derives from its $1/N$ initial borrowers, conditional on having no competing offers, equals

$$\hat{V}(\text{no competition}) = \frac{1}{N}[-\hat{k} + \hat{\nu}^* \frac{\hat{X}}{\rho}], \tag{3.70}$$

where I have used $\hat{X} = Y - [1 - \hat{k}]r_D$. If there is competition, rewrite (3.4) to obtain

$$\hat{\nu}[Y - R^{\max}(\hat{\nu}|\nu^*)] = \nu^*[Y - R^{\min}(\nu^*)] - S, \tag{3.71}$$

where $R^{\min}(\nu^*)$ is defined by (compare with (3.3))

$$-k_0 + \frac{\nu^*}{\rho}\{R^{\min}(\nu^*) - [1 - k_0]r_D\} = 0. \tag{3.72}$$

I compute the incumbent bank's profits as

$$\hat{V}(\text{competition}) = \frac{1}{N}[-k + \hat{\nu}\frac{R^{\max}(\hat{\nu}|\nu^*) - [1-k]r_D}{\rho}]. \tag{3.73}$$

Use (3.71) and (3.72) to compute the value that the deviating bank derives from its borrowers conditional on having competing offers:

$$\hat{V}(\text{competition}) = \frac{1}{N}[-\hat{k} + k_0] + \frac{1}{\rho N}\{\hat{\nu}\hat{X} - \nu^* X_0 + \hat{\nu} S\}, \tag{3.74}$$

where I have used $X_0 = Y - [1-k_0]r_D$.

In sum, the value of the deviating bank is

$$\hat{V}^* = [1-q]V(\text{no competition}) + qV(\text{competition}) - \frac{c[\hat{\nu} - \underline{\nu}]^2}{2}. \tag{3.75}$$

Insert (3.70) and (3.74) into (3.75) to obtain

$$\hat{V}^* = \frac{1-q}{N}[-\hat{k} + \hat{\nu}\frac{\hat{X}}{\rho}] + \frac{q}{N}[-\hat{k} + k_0 + \frac{1}{\rho}\{\hat{\nu}\hat{X} - \nu^* X_0 + S\}] - \frac{c[\hat{\nu} - \underline{\nu}]^2}{2}. \tag{3.76}$$

Note that the deviating bank maximizes its value in (3.76) by investing

$$\hat{\nu}^* = \frac{\hat{X}}{c\rho N} + \underline{\nu} \tag{3.77}$$

in monitoring technology. If condition (3.31) holds for all $k$; that is,

$$2Y/c\rho N + \underline{\nu} < 1, \tag{3.78}$$

the level of monitoring as given in (3.77) cannot be higher than 1. Insert (3.77) and (3.64) into (3.76) into obtain

$$\hat{V}^* = \frac{1}{N}[-\hat{k} + \underline{\nu}\frac{\hat{X}}{\rho} + \frac{\hat{X}^2}{2c\rho^2 N}] - \frac{q}{N}[-k_0 + \frac{\nu^* X_0}{\rho}]. \tag{3.79}$$

Differentiate (3.79) w.r.t. $\hat{k}$ to obtain

$$\frac{\partial \hat{V}^*}{\partial \hat{k}} = \frac{1}{N}\{-1 + \frac{r_D}{\rho}[\underline{\nu} + \frac{\hat{X}}{c\rho N}]\}. \tag{3.80}$$

Note that condition (3.78) guarantees that $\underline{\nu} + \frac{\hat{X}}{c\rho N} < 1$, hence (3.80) is negative for any $\hat{k}$. In sum, the bank would deviate to $\hat{k} = 0$.

I now show that the conjecture that banks hold on to their market share is correct as long as Assumption 3.1 holds for any $k$; that is, as long as

$$\frac{Y^2}{c\rho N} < S < [\underline{\nu}_G - \underline{\nu}_B][Y - r_D]. \tag{3.81}$$

If the deviating bank takes additional borrowers, its value is

$$\hat{V} = \frac{1-q}{N}[-\hat{k} + \hat{\nu}^* \frac{\hat{X}_0}{\rho}] + \frac{2q}{N}[-\hat{k} + \frac{1}{\rho N}\{\hat{\nu}^* \hat{X} - \nu X_0\}] - \frac{c\hat{\nu}^{*2}}{2},$$

and its optimal monitoring level is $\hat{\nu}^* = \frac{[1+q]\hat{X}}{c\rho N} + \underline{\nu}$. The value that the deviating bank extracts from the new borrower is

$$\hat{V}(\text{new borrower}) = \frac{1}{N}[-\hat{k} + k_0] + \frac{1}{\rho N}[\underline{\nu}[\hat{X} - X_0] + \hat{\nu}^* \hat{X} - \nu^* X_0 - S].$$

Insert expressions for $\hat{\nu}$ and $\nu$ to obtain

$$\begin{aligned}\hat{V}(\text{new borrower}) \quad = \quad & \frac{-\hat{k} + k_0}{N} + \frac{1}{\rho N}\{\underline{\nu}[\hat{k} - k_0]r_D + \frac{\{Y - [1-\hat{k}]r_D\}^2}{c\rho N} - \\ & \frac{\{Y - [1-k_0]r_D\}^2}{c\rho N} + \frac{q\hat{X}^2}{c\rho N} - S\}.\end{aligned}$$

Rearrange to obtain

$$\hat{V}(\text{new borrower}) = \frac{\hat{k} - k_0}{N}[-1 + \frac{1}{\rho}\{\underline{\nu} r_D + r_D[\hat{X} + X_0] + \frac{q\hat{X}^2}{c\rho N} - S\}]. \tag{3.82}$$

Note that (3.78) guarantees that $-1 + \frac{r_D}{\rho}[\underline{\nu} + \frac{2\hat{X}}{c\rho N}] < 0$ and (3.81) guarantees that $\frac{q\hat{X}^2}{c\rho N} < S$. This shows that (3.82) is negative.

In sum, each bank finds it profitable to lower its level of capital to zero regardless of the capitalization levels of competing banks. Hence, in equilibrium all banks set zero level of capital. ■

The intuition for Corollary 3.5 is the following. Capital is expensive. Because each bank's choice of monitoring is observable to the borrower deciding on which loan to take, a bank does not need costly capital to commit to a high(er) monitoring choice. That is, banks individually find it profitable to deviate to as low a level of capital as possible regardless of the level of capital of their competing banks. Hence, in equilibrium banks choose the lowest level of capital possible, this being zero or the minimum capital requirements set by the regulator.

Corollary 3.5 together with Proposition 3.7 provides a role for capital regulation. Although banks privately strive toward zero capital, the regulator follows positive welfare optimal capital requirements. This is different in Allen, Carletti, and Marquez (2007). They show that banks may privately commit to higher capital requirements than demanded by the regulator. Two differences in the model create the discrepancies in the results. First, in Allen, Carletti, and Marquez (2007) monitoring is unobservable. Banks commit to high capital to guarantee high monitoring to their borrowers. In my case, monitoring is observable and there is no need for capital from this perspective. Second, in Allen, Carletti, and Marquez (2007) banks choose high capital only if competition is high. High capital becomes important only if banks are at risk of losing their market share. In my case, Assumption 3.1 prevents a shift of the market share between equal type banks. Having high capital does not help a bank obtain new borrowers. Hence, individual incentives toward high capital are diminished.

A potentially different question is what level of capital banks would like the regulator to impose on the industry. In this formulation, banks could not deviate from the regulatory set capital requirements. In this case, banks choose $k$ to maximize their values in (3.58). That is, banks solve the following maximization problem.

$$k_{opt} = \arg\max_{k} V(\nu^*).$$

I can show the following result.

**Corollary 3.6.** *The banking industry wants the regulator to set zero level of capital requirements.*

*Proof:* Insert the equilibrium level of monitoring $\nu = \nu^*$ from (3.64) into (3.58) to obtain

$$V^* = \frac{1-q}{N}[-k + \frac{\underline{\nu} X}{\rho} + \frac{X^2}{c\rho^2 N}] + q\frac{S}{\rho N} - \frac{X^2}{2c[\rho N]^2}.$$

Rearrange to obtain

$$V^* = \frac{1-q}{N}[-k + \frac{\underline{\nu} X}{\rho}] + q\frac{S}{\rho N} + \frac{X^2}{2c[\rho N]^2}[1-2q]. \tag{3.83}$$

Differentiate (3.83) w.r.t. $k$ to obtain

$$\frac{\partial V^*}{\partial k} = \frac{1-q}{N}\{-1 + \frac{r_D}{\rho}[\underline{\nu} + \frac{X}{c\rho N}]\} - \frac{qX^2}{2c[\rho N]^2}. \tag{3.84}$$

Observe that $\nu^* = \underline{\nu} + \frac{X}{c\rho N}$. Condition (3.78) guarantees that $\nu^* < 1$, hence (3.84) is always negative. ■

Setting higher capital requirements imposes a higher cost on the industry and this is not in the interest of banks. Corollary 3.6 is even stronger than Corollary 3.5. Deviation to high levels of capital is not profitable even if all banks deviate together.

### 3.9.2 Good and bad banks

Now I extend the analysis to the heterogeneous banks; that is, $\underline{\nu}_G > \underline{\nu}_B$. Social welfare consists of social welfare that pertains to bad banks and social welfare arising from good banks.

I now compute the social welfare of bad banks. The social welfare consists of four parts; namely, the values of banks, consumer surplus, cost to the deposit insurance fund, and costs of entry. First, I can write the expected value of a bad bank with monitoring level $\nu_B$ as (see (3.7))

$$V_B(\nu_B) = \frac{1-q}{N}[-k + \frac{\nu_B X}{\rho}] + \frac{q[1-\gamma]}{\rho N}\{[\nu_B - \nu_B^*]X] + S\} - c\frac{[\nu_B - \underline{\nu_B}]^2}{2}. \qquad (3.85)$$

The first expression in (3.85) denotes the profits of a bad bank if its $\frac{1}{N}$ borrowers find no other offer, which occurs with probability $[1-q]$. The second expression in (3.85) denotes the profits of a bad bank if its borrowers find a competing bad bank, which occurs with a probability $q[1-\gamma]$. The monitoring level of a competing bank is $\nu_B^*$. A bad bank loses a proportion of its market share and its profits are zero when required to compete against a good bank. The last expression represents costs of investing in monitoring. Note that in total a bad bank having initially $\frac{1}{N}$ borrower keeps the following market share

$$\frac{1}{N}\{1 - q + q[1-\gamma]\} = \frac{1}{N}[1 - q\gamma].$$

Second, I evaluate the consumer surplus. In the case of no competition, the incumbent bad bank leaves borrowers with no profits. However, in the case of competition borrowers receive the total profits of the project net of funding costs $\nu_B X$ lowered by the profits of their bank $[\nu_B - \nu_B^*]X + S$. Hence, the consumer surplus of borrowers of one bank equals

$$CS_B(\nu_B) = [1-q] \times 0 + \frac{q[1-\gamma]}{\rho N}\{\nu_B X - [\nu_B - \nu_B^*]X - S\}. \qquad (3.86)$$

Third, in expectations the deposit insurance fund carries the following cost. The bank fails with probability $1 - \nu_B$. In this case, the deposit insurance fund repays all deposits in the total value of $\frac{[1-k]r_D}{\rho N}[1-q\gamma]$. The expected loss to the deposit insurance fund $L$ per bank is

$$L_B(\nu_B) = [1-\nu_B]\frac{[1-k]r_D}{\rho N}[1-q\gamma]. \qquad (3.87)$$

Fourth, each bad bank when entering incurs entering costs of $F$. In sum, the welfare per bad bank is

$$W_B(\nu_B) = V_B(\nu_B) + CS_B(\nu_B) - L_B(\nu_B) - F. \qquad (3.88)$$

Insert (3.85), (3.86) and (3.87) into (3.88) to compute the welfare contribution of one bank

$$W_B(\nu_B) = \frac{1-q\gamma}{N}[-k + \frac{\nu_B X}{\rho}] - c\frac{[\nu_B - \underline{\nu_B}]^2}{2} - \frac{[1-\nu_B][1-k]r_D[1-q\gamma]}{\rho N} - F. \qquad (3.89)$$

The welfare effect of good banks differs from (3.89) in two ways. First, good banks take new borrowers in competition from bad banks, which occurs with probability $q[1-\gamma]$.

Hence, each good bank obtains in total $\frac{1+q[1-\gamma]}{N}$ borrowers. A bank's profits and the cost to the deposit insurance fund are increased by a factor $1+q[1-\gamma]$ due to additional market share. More specifically, the factor $1-q\gamma$ in (3.89) is replaced by $1+q[1-\gamma]$. Second, the shift of the market share incurs switching costs $\frac{q[1-\gamma]}{\rho N}S$. Expressions for the cost of investment in monitoring technology and entry costs remain the same as in (3.89), except that a good bank monitors with $\nu_G$. The welfare effect of a good bank is

$$\begin{aligned} W_G(\nu_G) &= \frac{1+q[1-\gamma]}{N}[-k+\frac{\nu_G X}{\rho}] - c\frac{[\nu_G-\underline{\nu}_G]^2}{2} - \frac{[1-\nu_G][1-k]r_D\{1+[1-q\gamma]\}}{\rho N} \\ &\quad -F - \frac{q[1-\gamma]}{\rho N}S. \end{aligned} \tag{3.90}$$

The social welfare is the sum of the welfare effects of all banks:

$$TW(\nu_B,\nu_G) = N\{\gamma W_G(\nu_B) + [1-\gamma]W_B(\nu_G)\}. \tag{3.91}$$

I can obtain the following preliminary result.

**Lemma 3.4.** *Social welfare increases in q.*

*Proof:* Insert (3.8) into (3.89) to obtain

$$W_B = \frac{1-q\gamma}{N}[-k+\frac{\underline{\nu}_B Y - r_D[1-k]}{\rho} + \frac{[1-q\gamma]^2}{2c\rho^2N^2}\{Y^2 - r_D^2[1-k]^2\}. \tag{3.92}$$

Insert (3.9) into (3.90) to compute

$$W_G = \frac{1+q[1-\gamma]}{N}[-k+\frac{\underline{\nu}_B Y - r_D[1-k]}{\rho} + \frac{\{1+q[1-\gamma]\}^2}{2c\rho^2N^2}\{Y^2 - r_D^2[1-k]^2\}. \tag{3.93}$$

Use (3.91) to obtain

$$\begin{aligned} TW &= \frac{1}{N}[-k+\frac{\underline{\nu}_B Y - [1-k]r_D}{\rho}] + \gamma\{1+q[1-\gamma]\}[\underline{\nu}_G - \underline{\nu}_B]\frac{Y}{\rho N} + \\ &\quad + \frac{1+\gamma[1-\gamma]q^2}{2c\rho^2N^2}\{Y^2-[1-k]^2r_D^2\} - \frac{\gamma[1-\gamma]q}{\rho N}S - NF. \end{aligned} \tag{3.94}$$

Now differentiate (3.94) w.r.t. $q$ to obtain

$$\frac{\partial TW}{\partial q} = \frac{\gamma[1-\gamma]}{\rho N}\{[\underline{\nu}_G - \underline{\nu}_B]Y - S\} + + \frac{2\gamma[1-\gamma]q}{N^2}\{Y^2 - [1-k]^2r_D^2\}. \tag{3.95}$$

Use Assumption 3.1 to see that $[\underline{\nu}_G - \underline{\nu}_B]Y > S$ and (3.13) to see that $Y > [1-k]r_D$. This yields $\frac{\partial TW}{\partial q} > 0$. ■

This lemma points to the distribution effect of competition. Higher comptition $q$ means that more borrowers obtain financing at good banks, which increases the efficiency of a banking system raising social welfare.

The regulator should strive to increase interbank competition $q$ as much as possible. It could do this by several means. Banks' activities should be under constant antitrust surveillance. Also transparency on the loan market should be as high as possible.

Now I observe whether the individually chosen monitoring levels correspond to the welfare optimal ones.

**Lemma 3.5.** *Individually chosen levels of monitoring are lower than the welfare optimal ones, i.e.* $\nu_B^* < \nu_{B,reg}$ *and* $\nu_G^* < \nu_{G,reg}$.

*Proof:* Maximize (3.89) w.r.t. $\nu_B$ to obtain

$$\frac{\partial W_B}{\partial \nu_B}(\nu_{B,reg}) = \frac{1-q\gamma}{N}[\frac{Y}{\rho}] - c[\nu_{B,reg} - \underline{\nu}_B] = 0,$$

which yields $\nu_{B,reg} = \frac{[1-q\gamma]Y}{c\rho N}$. Observe from (3.8) that $\nu_{B,reg} > \nu_B^*$.

Maximize (3.90) w.r.t. $\nu_G$ to obtain

$$\frac{\partial W_G}{\partial \nu_G}(\nu_{G,reg}) = \frac{1-q\gamma}{N}[\frac{Y}{\rho}] - c[\nu_{G,reg} - \underline{\nu}_G] = 0,$$

which yields $\nu_{G,reg} = \frac{[1-q\gamma]Y}{c\rho N}$. Observe from (3.9) that $\nu_{G,reg} > \nu_G^*$. ■

The intuition for Lemma 3.5 replicates that of Lemma 3.3. The regulator considers the expected losses to the deposit insurance fund. Consequently, the regulator would like banks to increase their levels of monitoring above their privately optimal levels.

Indirectly, the regulator can impose higher monitoring by setting capital requirements above zero. That is, well capitalized banks invest in monitoring more (see Corollary 3.1). The regulator maximizes social welfare in (3.91) by setting capital requirements to $k_{reg}$. That is, the regulator solves the following maximization problem:

$$k_{reg} = \arg\max_k TW(\nu_B^*, \nu_G^*). \tag{3.96}$$

I can now show the following result.

**Proposition 3.8.** *The welfare maximizing capital requirement is given by*

$$k_{reg} = \max(0, 1 - \frac{c\rho^2 N[1-\frac{r_D}{\rho}]}{\{1+\gamma[1-\gamma]q^2\}r_D^2}). \tag{3.97}$$

*Proof:* Differentiate (3.94) w.r.t. $k$ to obtain

$$\frac{\partial TW}{\partial k}(k = k_{reg}) = \frac{1}{N}[-1 + \frac{r_D}{\rho}] + \{1+\gamma[1-\gamma]q^2\}\frac{[1-k_{reg}]r_D^2}{c\rho^2 N^2} = 0. \tag{3.98}$$

Solve this for $k_{reg}$ to obtain

$$1 - k_{reg} = \frac{[1-\frac{r_D}{\rho}]c\rho^2 N}{\{1+\gamma[1-\gamma]q^2\}r_D^2}, \tag{3.99}$$

which yields (3.97). Note also that $\frac{\partial^2 TW}{\partial k^2} < 0$. ■

The intuition for Proposition 3.8 matches that of Proposition 3.7. Due to deposit insurance, banks underestimate risk and insufficiently invest in monitoring. The regulator increases banks' investments in monitoring and social welfare by setting positive capital requirements.

The comparative statics with respect to the optimal capital requirements are as follows.

**Corollary 3.7.** *The welfare maximizing capital requirement $k_{reg}$ is decreasing in the number of banks $N$.*

*Proof:* Differentiate part of (3.97) as follows.

$$\frac{\partial}{\partial N}[\frac{1}{N} + \gamma[1-\gamma]\frac{q^2}{N}] = -\frac{1}{N^2} + \frac{2\gamma[1-\gamma]}{N}\frac{\partial q}{\partial N} - \gamma[1-\gamma]\frac{q^2}{N^2}. \tag{3.100}$$

Note that

$$\frac{\partial}{\partial N}[\frac{1}{N} + \gamma[1-\gamma]\frac{q^2}{N}] < \gamma[1-\gamma][\frac{2}{N}\frac{\partial q}{\partial N} - \frac{q^2}{N^2} - \frac{4}{N^2}] < 2\gamma[1-\gamma]\frac{\partial[q/N]}{\partial N} < 0. \tag{3.101}$$

Thus, $\frac{\partial k_{reg}}{\partial N} < 0$. ■

Lowering the number of banks in the banking industry augments the size of each bank. However, capital regulation is especially effective for big banks. To see this, note that the fixed cost of investment in monitoring makes it optimal for small banks to invest a little in monitoring technology both from their individual profit maximization view and from the welfare optimal point of view. Hence, capital regulation does not play a large role for small banks. For large banks, the investment in monitoring technology is high, as is the discrepancy between privately optimal and welfare optimal levels of monitoring. This makes capital regulation effective. In sum, the regulator sets higher capital requirements if banks are large; that is, for low $N$.

I can now show the following result.

**Corollary 3.8.** *The welfare maximizing capital requirement $k_{reg}$ is increasing in competition parameter $q$.*

*Proof:* Differentiate (3.97) w.r.t. $q$ to obtain $\frac{\partial k_{reg}}{\partial q} > 0$. ■

Raising capital requirements augments the differences between banks through limiting the distortions of deposit insurance (see the intuition for Proposition 3.3). Increasing the differences between banks enhances social welfare through the effect of competition. In particular, competition shifts the market share from bad to good banks. The stronger competition is, the more positive its redistribution effect is and the more important the differences are between banks. That is, the positive effect of capital requirements is the greatest if competition is high. Because the cost of capital is fixed, the welfare optimal capital requirement $k_{reg}$ increases with competition $q$.

Corollary 3.8 contrasts with Corollary 3.4. The competition parameter $q$ affects the welfare optimal capital requirement $k_{reg}$ in a heterogeneous banking system because of the shift of the market share. In a homogeneous banking system, banks hold on to their borrowers. Consequently, in a homogeneous banking system the competition parameter $q$ has no effect on the welfare optimal level of capital requirements.

A different question is what level of capital banks would want the regulator to set. The preferences of good and bad banks toward capital requirements are in conflict with those of the regulator, who strives toward the welfare optimal level of capital requirements. The group of good banks would like the regulator to set the capital requirement $k_{G,opt}$ to maximize $V_G(\nu_G^*)$, see (3.7). On the other hand, the group of bad banks would lobby the regulator to set the capital requirement $k_{B,opt}$ in order to maximize $V_B(\nu_B^*)$, see (3.7),

$$k_{G,opt} = \arg\max_k V_G(\nu_G^*) \quad \text{and} \quad k_{B,opt} = \arg\max_k V_B(\nu_B^*). \tag{3.102}$$

The maximization problem in (3.102) differs from the welfare maximization in (3.96) due to two main reasons. First, banks do not consider expected losses to the deposit insurance fund, whereas the regulator does. Second, banks in (3.102) optimize the profits of banks of their own type. That is, good banks try to compel the regulator to set the capital requirements that are the best for good banks but not necessarily for bad banks.

The following result extends Proposition 3.3.

**Proposition 3.9.** *Bad banks want the regulator to set the capital requirement to zero; that is, $k_{B,opt} = 0$. Good banks want the regulator to increase the level of capital requirements as long as competition is sufficiently strong (high $q$) and the quality of the banking industry is sufficiently low (low $\gamma$).*

*Proof:* The proof is similar to the proof of Proposition 3.3 and is therefore omitted. ■

High capital requirements not only result in a higher cost of financing, but also hamper the competitive ability of bad banks. Consequently, bad banks would opt for zero capital requirements. However, high capital requirements, despite their higher cost, bring the enhanced position of good banks into competition with bad banks. In a highly competitive and low-quality banking industry, good banks often find themselves competing with bad banks. Thus, for high $q$ and low $\gamma$, good banks suggest that the regulator increases capital requirements.

**Corollary 3.9.** *For sufficiently high $q$ and low $\gamma$, good banks would want the regulator to increase capital requirements above the welfare optimal level.*

*Proof:* I assume that Assumption 3.1 holds for all $k$. In this case (3.81) holds. Note that (3.81) guarantees that (3.37) holds for all $k$, hence Proposition 3.3 holds for all $k$. That is, because bad banks want to lower capital requirements for all $k$, their optimal capital requirements are zero. If $q$ is sufficiently high and $\gamma$ sufficiently low, good banks want to increase capital requirements for all $k$, hence they would like to have $k = 1$, which is above the welfare optimal capital requirement $k_{reg}$ (observe from (3.97) that $k_{reg} < 1$). ■

The reason for this corollary is that good banks would like to enhance the quality difference with respect to bad banks so that they can extract higher rents from their borrowers when competing with bad banks. Higher capital achieves precisely this. It increases the quality difference between good and bad banks because bad banks can no longer obtain cheap funding from insured deposits. In the expected value sense, higher differences allow good banks to extract higher rents from the borrowers, hence they want to have high capital requirements. However, the regulator maximizes total welfare, including the consumer surplus. Thus, the regulator prefers to set lower capital requirements due to higher costs of capital financing.

In our basic model, an entering bank does not yet know its type. However, it knows the cross-sectional distribution $\{\gamma, 1-\gamma\}$ of being good or bad. I analyze what level of capital an entering bank would impose on the regulator.

**Proposition 3.10.** *An entering bank wants the regulator to increase the level of capital requirements if competition is sufficiently high (i.e., $q \geq \hat{q}$), and the quality of the banking industry is intermediate (i.e., $\gamma \in [\hat{\gamma}_1(q), \hat{\gamma}_2(q)]$).*

*Proof:* The proof is similar to the proof of Proposition 3.4 and is therefore omitted. ■

Proposition 3.10 extends Proposition 3.4. Higher capital requirements increase the differences between good and bad banks. Whereas bad banks earn zero profits due to high competition for all levels of capital, good banks earn higher rents in competition with bad banks if capital requirements are high. For the intermediate qualities of a banking industry good banks often compete with bad banks. This makes capital requirements beneficial for banks on average.

**Corollary 3.10.** *For sufficiently high $q$ and intermediate $\gamma$, an entering bank would want the regulator to increase capital requirements above the welfare optimal level.*

*Proof:* I assume that Assumption 3.1 holds for all $k$. In this case (3.81) holds. Note that (3.81) guarantees that (3.37) holds for all $k$, hence Proposition 3.4 holds for all $k$. That is, if $q < \bar{q}$ and $\gamma \in [\gamma_1(q), \gamma_2(q)]$, entering banks' values increase at the level of capital requirements $k$ for all values of $k$, hence entering banks would like to set $k = 1$, which is above the welfare optimal capital requirement $k_{reg}$ (observe from (3.97) that $k_{reg} < 1$). ■

The reason for this corollary is that banks want to increase the quality differences between themselves to extract higher rents from borrowers. Higher capital augments the difference in quality between good and bad banks. That is, good banks take higher rents from borrowers when competing with bad banks. The expected profit of an entering bank goes up to the extent that the borrowers are worse off. Although increasing capital requirements above the welfare optimal level might positively affect the expected value of entering banks, it hampers social welfare through the negative effect on the consumer surplus. In this case the regulator, which maximizes social welfare, wants to set capital requirements below the level that is optimal for entering banks.

### 3.9.3 Welfare with endogenous entry

I now analyze the welfare effects of entry.

**Lemma 3.6.** *Social welfare is decreasing in $N$.*

*Proof:* I assume that the condition in (3.13) holds for all $k$. That is,

$$-1+\frac{\nu_B Y}{\rho}-\frac{S}{\rho}>0. \tag{3.103}$$

This guarantees that a bank that is a monopolist always offers loans to its borrowers even though it is financed completely by capital.

Differentiate (3.94) w.r.t. $N$ to obtain

$$\begin{aligned}\frac{\mathrm{d}TW}{\mathrm{d}N} &= -\frac{1}{N^2}[-k+\frac{\nu_B Y-[1-k]r_D}{\rho}]-\frac{1}{\rho N^2}\{[\underline{\nu}_G-\underline{\nu}_B]Y-S\}-\frac{2}{N^3}\{Y^2-[1-k]^2r_D^2\}\\ &+\frac{\partial[q/N]}{\partial N}\frac{\gamma[1-\gamma]}{\rho N}\{[\underline{\nu}_G-\underline{\nu}_B]Y-S\}+\frac{\partial[q/N]}{\partial N}\frac{2\gamma[1-\gamma]q}{N}\{Y^2-[1-k]^2r_D^2\}-F. \end{aligned}\tag{3.104}$$

Use $[\underline{\nu}_G-\underline{\nu}_B]Y>S$ from Assumption 3.1, $\frac{\partial[q/N]}{\partial N}<0$ and the condition in (3.103) to see that $\frac{\partial TW}{\partial N}<0$. ■

The intuition for this corollary stems from the fixed cost of monitoring, which creates economies of scale. With low $N$, banks are bigger but the costs of monitoring are fixed. Thus, the cost per borrower is lower and the banks' efficiency is increased. In addition, fewer banks have to pay entry cost $F$.

The welfare implications of opening up a banking system are different in good and bad banking systems. Social welfare decreases with the number of banks in the banking system. Proposition 6.1 shows that opening up borders induces more entry in bad banking systems but no additional entry in good banking systems. Together with Lemma 3.6, this means that opening up a bad banking system hampers social welfare. In contrast, opening up a good banking system has no direct effect on social welfare. In fact, the indirect effect of opening up a good banking system might be an increase in interbank competition $q$, which leads to an improvement in social welfare.

Note also that two parameters of competition $q$ and $N$ act in an opposite way (compare Lemma 3.6 with Lemma 3.4). Regulators should strive to increase interbank competition $q$ keeping the number of banks $N$ fixed.

I can now show the following result.

**Proposition 3.11.** *If increasing capital requirements augments entry, the regulator sets lower capital requirements when entry is possible.*

*If increasing capital requirements lowers entry, the regulator sets higher capital requirements when entry is possible.*

*Proof:* The proof directly follows from Corollary 3.7. ■

When entry is possible, higher capital requirements affect the number of entering firms. If capital requirements increase entry (Part 2a of Proposition 3.4), banks on average are of lower size. Consequently, capital regulation is less effective. That is, the marginal benefits of higher capital decrease and the marginal cost stays the same. Hence, the regulator sets lower capital requirements.

If capital requirements decrease entry, (Part 1 and Part 2b of Proposition 3.4) banks on average are of bigger size. Consequently, capital regulation is more effective and the regulator sets higher capital requirements.

### 3.9.4 Influence on the real economy

So far the social welfare/private optimum comparison has only involved the expected losses to the deposit insurance fund, which is taken into account from the social welfare point of view, but not from a bank's private optimum determination. However, social welfare is more than just focusing on the losses to the deposit insurance fund. As I have highlighted in the main text, it also involves stability considerations; that is, externalities to the real economy and other financial institutions. One way of capturing these externalities is looking at the average success probability of banks across the economy. This is not different from focusing on the average monitoring intensity of banks per borrower, which is also the average success probability of the borrower. I can show that competition in general improves the average success probability in the economy. That is, I can prove the following proposition.

**Proposition 3.12.** *Increasing competition via elevating $q$ augments the average success probability of the borrowers.*

*Proof:* The average monitoring per borrower is defined as

$$E(\nu) = [1-\gamma][1-q\gamma]\nu_B^* + \gamma\{1+q[1-\gamma]\}.$$

Insert (3.8) and (3.9) into obtain

$$E(\nu) = \{1+q^2\gamma[1-\gamma]\}\frac{X}{c\rho N} + [1-\gamma][1-q\gamma]\underline{\nu}_B + \gamma\{1+q[1-\gamma]\}\underline{\nu}_G. \tag{3.105}$$

Now differentiate (3.105) to obtain

$$\frac{\partial E(\nu)}{\partial q} = 2q\gamma[1-\gamma]\frac{X}{c\rho N} + \gamma[1-\gamma][\underline{\nu}_G - \underline{\nu}_B], \tag{3.106}$$

which is always positive. ■

The intuition for Proposition 3.12 is the following. Competition has two effects. First, competition increases the levels of monitoring technologies of good banks but decreases monitoring of bad banks (see Corollary 3.2). Interestingly, an increase in monitoring of good banks is exactly compensated with a decrease of monitoring of bad banks. Hence, the average monitoring *per bank* does not change with competition parameter $q$. However, higher competition $q$ also means that good banks acquire additional borrowers. Good banks become more important than bad banks due to their larger size. More specifically, more borrowers are monitored by good banks with high levels of monitoring and fewer by bad banks with low levels of monitoring. Consequently, the average monitoring level *per borrower* augments with $q$.

Proposition 3.12 provides a very positive view on competition. It is true that high competition lowers the levels of monitoring of bad banks and makes them less stable. However, competition also shifts the market share from bad to good banks. That is, bad banks lose their size and become less important in the average stability sense.

I can observe the effect of capital on the average success probability.

**Corollary 3.11.** *Higher capital requirements increase the average success probability the most if competition is high (high $q$) and the quality of the banking industry is intermediate $\gamma = 1/2$.*

*Proof:* Differentiate (3.105) w.r.t. $k$ to obtain

$$\frac{\partial E(\nu)}{\partial k} = \{1 + q^2\gamma[1-\gamma]\}\frac{r_D}{c\rho N}. \tag{3.107}$$

Note that (3.107) increases in $q$ and is the highest for $\gamma = \frac{1}{2}$. ■

The first part of Corollary 3.11 is explained as follows. High competition shifts a greater proportion of the market share from bad to good banks. Proposition (3.2) shows that capital regulation is more effective for good banks. Because high competition increases the share of good banks, it also increases the effectiveness of capital regulation.

The second part of Corollary 3.11 dwells on the observation that the shift of the market from bad to good banks is the highest in the intermediate quality banking industry. This first of all increases the share of good banks the most. Second, it also allows good banks to grow in size the most. Both effects increase the effectiveness of capital regulation.

# Chapter 4

# Regulation of a Competitive Banking System

**Abstract**

The regulation of a highly competitive and rapidly changing banking system presents daunting challenges. Competition typically increases efficiency, yet, is traditionally considered a threat to stability. Recent research, however, argues that competition and stability could (sometimes) also go hand in hand. This chapter discuses direct and indirect barriers to competition stemming from stability-oriented regulation. Despite the potential beneficial effects of competition, regulators should only cautiously liberalize their banking sectors. Liberalization, and the period of transformation that it implies, can induce opportunistic behavior suggesting the need for precaution. Moreover, competition might lower the effectiveness of existing regulatory tools. I also consider the efficiency and effectiveness of current regulatory practices in light of the recent consolidation in the banking industry.

Keywords: Bank Regulation, Competition, Consolidation
JEL CLASSIFICATION: G21, G28

## 4.1 Introduction

Under the pressure of fierce competition, the banking industry is rapidly reshaping its structure (see Berger and Mester (1996), Stiroh and Strahan (2003) and Boot (2003)). On the one hand, banks appear to be scrambling to become bigger, whether through organic growth or through mergers and acquisitions. On the other hand, banks need to decide whether to specialize or be broad (all-purpose) financial institutions. For bank regulators, this offers challenges in that they must follow the developments and ask themselves whether the past regulatory framework is still suitable under the quickly evolving, dynamic banking environment. In this chapter I provide an overview of the insights that the literature so far has produced on the interaction between competition and stability. In particular, I focus on the impact of the more competitive environment on the effectiveness of bank regulation.

The main objective of prudential regulation is clear. The regulator stands for the stability of the banking system, yet also needs to take into account the efficiency of the industry. In light of long-lasting efforts of regulators, the implementation of the stability objective appears to be a difficult task. Banks may intrinsically lack stability, having on one side of their balance sheets liquid demand deposits and, on the other side, illiquid loans. A weak position of a bank might induce depositors to withdraw their funds. In the worst case, such a bank run could induce a confidence crisis and spread to other banks and create a systemic banking crisis with negative repercussions for the entire economy. Indeed, the negative externalities of a bank failure justify the existence of bank regulation in the first place.

The question appears whether the stability and efficiency (and competitiveness) objectives are compatible with each other. More specifically, higher competition may improve the efficiency of the banking industry and that could be good for stability; yet, in the extreme case, it may also produce cut-throat competition that leads to bank failures. Older studies argue that higher competition induces banks to behave more riskily and creates instability (see Keeley (1990)). This influence might be especially strong in developing countries with weak institutional and legal structures. In developed countries, however, the link between stability and competition could be more positive. Although older studies emphasize the potential negative impact of competition, recent studies are more positive (see Boyd and De Nicolo (2005)).

Even though competition does not have a clear-cut effect on stability, it might distort stability-oriented regulatory practices and make them ineffective. I argue that this does not justify limiting competition. The regulator should respond by reshaping stability-oriented regulation to suit the more competitive environment.

The chapter is organized as follows. In Section 4.2, I review how bank regulation aims to assure stability of the banking system. Section 4.3 analyzes the direct impact of competition on bank stability and efficiency. Section 4.4 reviews regulatory practices that might (intentionally or unintentionally) limit competition. In Section 4.5, I discuss the impact of competition on the effectiveness of stability-oriented regulation, with a particular focus on deposit insurance and capital regulation. Section 4.6 assesses the effectiveness of regulation in light of the recent consolidation and conglomeration of the banking industry. Section 4.7 concludes the chapter.

## 4.2 Rationales for Bank Regulation

The reasons for bank regulation stem from the functions that banks perform in the economy. In particular, banks can be considered a lubricant for the economy at large, and bank instability might impose substantial externalities. In this section, I seek to uncover the main source of externalities in banking. This helps in understanding the optimal regulatory and supervisory design.

Banks are active on the asset side as well as on the liability side of their balance sheets. On the asset side, banks have loans that need monitoring. These assets, often rather opaque, reflect the idiosyncracies of bank borrowers that may not have access to financial markets. In particular, information asymmetries could make capital markets inefficient in financing informationally opaque companies (e.g., small and medium enterprises). Banks aim to resolve this market failure by screening and monitoring borrowers (e.g., business that need financing). This allows them to intermediate between investors (depositors) and businesses that need financing.[1] The additional information that banks possess inevitably makes them opaque (see Morgan (2002)).

On the liability side, banks create liquidity. That is, banks finance illiquid loans with liquid demand deposits.[2] Demand deposits, however, present an underlying threat for a bank. Depositors can withdraw their funds at any time. This could create instabilities, but it also exerts pressure on banks to behave prudently (see Calomiris and Kahn (1991)).

The opaqueness of banks on the asset side together with liquidity provision on the liability side make banks inherently unstable institutions. More specifically, if many depositors unexpectedly withdraw funds, the bank involved incurs high liquidation costs and even a previously solvent bank might be forced in liquidation. A broader, systemic bank crisis could result because the banking industry is highly interconnected and the failure of one bank can have negative repercussions for other banks and hence for the economy at large. Empirical evidence points at a cost of a bank crisis ranging from 5 to 20% of annual GDP (see Hoggarth, Reis, and Saporta (2002) and Bordo, Eichengreen, Klingebiel, and Martinez-Peria (2001)).

These huge negative externalities could provide a justification for bank regulation. More specifically, due to externalities, banks on their own may not fully internalize the costs of failure; that is, the social cost of failure exceeds the private cost. A widely accepted form of regulation is the implementation of deposit insurance. Deposit insurance prevents bank runs because depositors know that their money will be repaid anyway. However, it creates other distortions. Depositors do not carefully examine the stability of their banks anymore. The

[1] Banks screen borrowers, prevent opportunistic behavior by borrowers by monitoring, and punish and/or audit them (see Freixas and Rochet (1999)).

[2] For a further discussion of liquidity provision, see Von Thadden (1999) and Bhattacharya, Boot, and Thakor (1998).

regulator may now have to take full responsibility for the prudent behavior of each bank.

Several regulatory tools have been created to deal with the banks' potential excessive risk taking. Intrusive regulation in the past included restrictions on activities, geographical limitations, and various limitations on the banks' prices (e.g., deposit interest rate ceiling). An example of intrusive regulation was the Glass-Steagall Act in the U.S. that kept investment banking and commercial banking separated.[3] Nowadays, rather than being "brute forced" into desired behavior, prevailing regulation is more of an indirect nature, in which banks are indirectly "compensated" for being prudent or alternatively charged for risk taking. An example is capital regulation, in which banks are obliged to hold a level of equity capital that corresponds to the riskiness of their activities.

Basel I was the first widely implemented form of capital regulation. Despite several criticisms, its aims are largely fulfilled. It has raised capital across the banking industry across the world. While Basel I had some calibration to risk taking, it was rather rudimentary. The new Basel II framework aims to improve on this. It also has a greater emphasis on market discipline.

In summary, banks provide for an efficient allocation of funds in the economy, mainly by resolving informational problems between firms and their financiers and providing liquidity transformation. As argued, banks' role in resolving information asymmetries and liquidity transformation makes them prone to bank runs and other potential instabilities. To safeguard the banking system, deposit insurance and prudential regulation is needed.

Before focusing on the connection between regulation and competition, I first analyze the effects of competition in light of the stability and efficiency objectives of the regulator.

## 4.3 Competition, Efficiency, and Bank Stability

In this section I first analyze the link between competition and the efficiency of the financial system. Following this, I evaluate the impact of competition on bank stability.

### 4.3.1 Competition and efficiency

An efficient financial system should minimize transaction costs, where these costs should be interpreted broadly as the resources that dissipate or evaporate in the process of allocating resources. This generally necessitates a certain degree of competitiveness. Indeed, competition is generally found to be of critical importance for cost efficiency (including productive efficiency) and for the optimal allocation of resources (allocative efficiency).

Stiroh and Strahan (2003) analyze the link between growth and the performance of the banking industry after the deregulation in the U.S. in the 1980s. They show that deregulation shifted the market share from less to more efficient banks. Similarly Carlson and Mitchener (2006) present evidence on the positive effect of competition on stability. They argue that

[3] See Boot, Milbourn, and Dezelan (2001) on the distinction between direct and indirect regulation.

the expansion of branching in the U.S. in the 1920s increased competition, which weeded out inefficient banks and effectively made the banking system stronger. This is related to a study by Berger and Hannan (1998), in which the efficiency of banks is linked to the market structure. In particular, more monopolistic banks tend to be less efficient than banks subjected to more competition. Evanoff and Ors (2002) evaluate the level of cost efficiency after competition increases. They find that the incumbent banks respond by increasing their level of cost efficiency if a new bank enters or an existing competitor becomes stronger. Similarly, Košak and Zajc (2006) show that opening up the European banking system decreased the efficiency gap between Western banks (in the old EU member countries) and Eastern banks (in the new EU member countries).

Jayaratne and Strahan (1998) analyze increased competition in the U.S. due to relaxation of bank branch restrictions. They show that per-capita growth in income and output increases significantly following deregulation of the financial industry. Hence, higher competition in the banking industry has a positive effect not only on the efficiency of the banking industry but also on the productivity of the real economy.

Competition might have a somewhat different effect on the optimal allocation of resources in banking due to the presence of information asymmetries. Petersen and Rajan (1995), for example, show that competition might hamper the intrinsic role of banks as a monitor of borrowers. In particular, in Petersen and Rajan's view, banks may invest less in relationship building and are less willing to grant credit because they expect lower future relationship rents due to the more competitive environment.

However, several studies give a more positive view on the role of competition in enhancing bank monitoring ability and consequently in credit allocation. Boot and Thakor (2000) suggest that competition drives banks to focus on activities that are less prone to price competition, such as relationship lending. Competition undermines rents more in transaction lending, such as mortgage loans, than in more information-based lending, as in relationship banking. Contrary to Petersen and Rajan, they predict that competition could lead to more relationship lending. Similarly, Dell'Ariccia and Marquez (2004) suggest that competition induces more sector specialization. That is, banks may choose to escape price competition by investing more in knowing their borrowers and thus offer more tailored services. Sector specialization helps in doing this. The positive effects of competition in their analysis are twofold. First, competition assures lower interest rates for borrowers and, second, informationally opaque borrowers obtain financing more easily.[4]

In short, findings in the literature point to a favorable effect of competition on efficiency. It typically improves the cost efficiency of banks and capital allocation across the economy. Competition may also intensify bank-borrower relationships despite popular notions to the contrary. Now I turn to the mixed evidence on the relationship between competition and stability.

[4]Beck, Demirgüç-Kunt, and Levine (2003) empirically confirm that bank concentration increases obstacles to obtaining finance.

### 4.3.2 Competition and stability

In this section I analyze the interaction between competition and bank stability. Competition and stability are often seen as conflicting rather than complementary objectives in banking, thus possibly presenting regulators with a difficult trade-off. I show that this might especially be true in developing countries, whereas in developed countries competition and stability often go hand in hand.

Historically speaking, up to the 1980s, competition was generally strictly contained in the banking system. Globalization helped to force policy makers to open up banking systems for competition. In this more dynamic environment concerns about stability became more prevalent.

Keeley (1990) theoretically justifies the connection between competition and instability. Keeley argues that banks behave prudently because they want to preserve high future rents, which are lost if banks fail. High competition erodes future rents, making banks' risky behavior more attractive. In a similar spirit, Dell'Ariccia and Marquez (2006) show that deregulation that reduces information asymmetry between banks might increase competition. In order to obtain additional market share, banks loosen their credit standards, which might result in a lending boom and potentially in a banking crisis.

Although competition could make the banking system more fragile, it is not the cause for fragility in the banking system (see Vives (2001a)). As argued earlier, banking has inherent features that make it potentially fragile. In this spirit, Boyd, De Nicolo, and Smith (2004) show that banking crises are not only confined to competitive banking systems, but can also occur in a monopolistic system. More recently, people have argued that competition might even have a positive effect on bank stability. The analysis of competition and efficiency in Section 4.3.1 already pointed to this. Competition helps strong banks gain market share, and this very survival of strong banks could lead to a more stable banking system. Boyd and De Nicolo (2005) provide a different argument. They argue that competition forces banks to charge lower interest rates to their borrowers, which could induce borrowers to behave more prudently.

Often it is too easy to blame competition for banking crises. Looking back at many recent crises (e.g., the S&L crisis and the Swedish banking crisis), acute regulatory problems including lax enforcement, weak internal controls, poor supervision, poor transparency, and lack of market pressure were pervasive (see Barth, Trimbath, and Yago (2004) and Englund (1999)).

The empirical study by Beck, Demirgüç-Kunt, and Levine (2003) confirms the positive role of competition on bank stability. Their cross-country analysis shows that crises are less likely in banking systems with fewer restrictions on bank competition and activities. In addition, stronger national institutional arrangements that encourage competition lower the probability of bank failure.

Demirgüç-Kunt and Detragiache (2001) are more cautious in presenting beneficial effects of liberalization. Although they control for several macroeconomic variables, they show that liberalization is negatively related to the stability of the banking system. This points to an important insight. Competition per se might not be bad for stability, but *changes* in the (competitive) environment might cause instabilities because one needs to adjust to this new environment. Thus, the path towards a new equilibrium entails risks (see also Chapter 3). The negative effect of liberalization on stability in the Demirgüç-Kunt and Detragiache (2001) study is weaker where the institutional environment is stronger. In particular, countries with low respect for rule of law, where corruption is high and contract enforcement is low, should be very careful when introducing financial liberalization. These circumstances particularly apply to developing countries.

Honohan and Stiglitz (2001) further defend the view that financial liberalization should be approached carefully in developing countries. In developing countries, agency and information problems are severe. That is, banks might be managed to the benefit of corrupt owners.[5] In addition, reliable information about a bank's behavior such as its risk taking and its economic value (e.g., the true value of loans and collateral) is more difficult to obtain. In light of these problems, prudent regulation might be difficult. Competition may then exacerbate this problem. In their view, a lower level of competition might then be desirable and help support sound regulatory practice to contain risky bank behavior.

Altogether, while competition clearly enhances efficiency, the link between competition and stability is rather inconclusive. Foremost, structural changes in banking systems entail risk. Thus, any liberalization should be managed carefully during the transformation path. From a more steady-state perspective, the evidence points to the following tentative conclusion. Competition seems beneficial in more developed banking systems, whereas in underdeveloped systems it poses a severe threat to stability.

## 4.4 Regulatory Barriers to Competition

What can policy makers do to foster competition? In this section I focus on the impact that prudential regulation has on competition in the banking system. The key conclusion is that one needs to carefully reexamine existing stability-oriented regulatory practices that may implicitly entail barriers to competition. I primarily focus on regulatory induced barriers to competition in the context of the U.S. and EU banking systems.

As a response to the Great Depression, the U.S. implemented the Glass-Steagall Act, which enforced tight control over banking activities. The Glass-Steagall Act separated the activities of commercial banks and securities firms. The U.S. also imposed Regulation Q to restrict the level of competition with the aim of increasing bank stability. Regulation Q prohibited paying interest rates on demand deposits and capped the interest rate on savings deposits.[6] Interestingly, regulation Q had an unexpected outcome. It hampered banks'

[5]For instance, the Russian banking sector, though growing rapidly, remains fragmented, with a large population of very small banks. Many of these small banks are banks in name only and are used by their owners for such purposes as tax "optimization" or money laundering (see Tompson (2004)).

[6]The Glass-Steagall Act also introduced deposit insurance by the Federal Deposit Insurance Corporation

ability to compete against money market mutual funds. The micro-created instability forced its removal in the 1980s. U.S. banks were also geographically restricted; for example, by limitations on opening branches in different U.S. states (the McFadden Act). The Riegle-Neal Act of 1994 removed most barriers to interstate branching. The Gramm-Leach-Bliley Act of 1999 lifted restrictions on the activities that banks could undertake (i.e., combining securities and insurance with commercial banking). In doing so, it allowed for the creation of a financial services holding company.

Deregulation had a huge effect on the structure of the U.S. banking industry. Whereas as in 1984, 14,407 U.S. banks operated 43,250 branches, by 2005 the number of banks had fallen to 7,527 and the number of branches had increased to 71,716.[7] Deregulation of restrictions on activities together with intensifying competition forced banks to search for rents in their non-traditional business and identify their true comparative competitive advantages. The rapid transformation of the banking industry indirectly points to previously established inefficiencies, which in part were induced by excessive regulation. Jayaratne and Strahan (1996) and Clarke (2004) provide empirical evidence supporting the conclusion that the interstate bank deregulation (and also the removal of intrastate branching restrictions) enhanced economic growth. The consolidation also appears to have led to a more efficient provision of banking services. Claessens and Laeven (2004) find no evidence that concentration negatively affects the level of competition in the banking system. The level of competition is largely driven by allowing for foreign entry and reducing activity restrictions.

The diversity of the European Union presents a different challenge for policy makers. Policy makers are trying to overcome the geographic segmentation of banking markets but are only very slowly succeeding. The Single Market Program of 1992 stipulating home country control on regulating the cross-border activities of banks is only slowly showing positive effects. This Single Market Program of 1992 makes banks accountable to their home regulators, without (extensive) intrusion of the host regulator for their branches abroad. Bank consolidation is largely a domestic phenomenon and only recently are cross-border mergers being observed.[8] A variety of reasons can be suggested for this. For most, cultural differences between countries and implicit barriers involving domestic "favoritism" come to mind.

One of the direct barriers is the reluctance of national regulators to allow foreign banks in. Examples include the takeover of two Italian banks, Banco Nazionale del Lavoro and Banca Antonveneta, by respectively, the Spanish bank Banco Bilbao Vizcaya Argentaria (BBVA) and the Dutch bank ABN-AMRO; respectively. In these cases, pressure from the EU helped overcome the intrusion of the domestic regulator. Both domestic banks were ultimately taken over.[9] It may well be that these developments establish a "turning point"

---

and increased the ability of the Federal Reserve to increase money supply.

[7]See FDIC Statistics on Banking, A Statistical Profile of the United States Banking Industry, March 2006.

[8]Cross-border mergers are observed at the regional level; for example, between the Netherlands and Belgium, and within the Scandinavian countries.

[9]Also in the U.S., the interests of more powerful players affect regulation. Kroszner and Strahan (1999)

and that cross-border activity may accelerate.

In a recent initiative, the European Commission proposed an amendment to EU legislation that would restrain the discretion of national regulators in preventing a cross-border takeover. Current EU rules allow national supervisory authorities to block proposed merger if they consider that the "sound and prudent management" of the target company could be put at risk. The proposed new directive would in particular clarify the criteria against which supervisors should assess possible merger operations. This would improve clarity and transparency in supervisory assessment and help ensure the consistent handling of merger requests across the EU (see European Commission (2006b)).

Chapter 3 gives another arguably pertinent reason for domestic mergers. It shows that especially weak domestic banks could "delay the inevitable" by merging to gain extra economies of scale and market power to defend against entry of foreign competitors.[10] In the short run this seems viable. However, in the long run inherent weaknesses of these banks would put their (independent) survival at risk and, even worse, the potential distress of these institutions may cause substantial systemic problems. This would suggest that opening such a domestic market should allow for takeovers of weak domestic institutions by foreign entrants. The previously mentioned directive of the European Commission works in this way.

Potential barriers to competition also come from stability-oriented regulation. For example, the costs and complexity of implementing Basel II standards might put small banks at a competitive disadvantage. The dispersed EU regulatory and supervisory structure could also lead to high costs and complexity discouraging entry. For example, in the EU, regulatory costs are increased because cross-border operating banks often have to report to several regulators. Although EU directives and other international standards leave limited scope for typical national arrangements, national rules are still not fully harmonized, and also regulatory and supervisory arrangements differ. This is particularly relevant for banks that have their foreign operations in subsidiaries. These are, contrary to branches, regulated by the host-country regulator (see Section 4.6.1 for further discussion).

The answer to this puzzle might be that local regulation suits local players. Foreign banks may then be at a disadvantage. Regulators might also be biased toward the needs of their domestic banks, especially if the domestic banks are large (see Boot and Thakor (1993)). Such sizable domestic banks may also find themselves implicitly protected by too-big-to-fail policies. The failure of such a bank might simply be too costly to the economy at large, and/or to the popular support of domestic politicians.

Another disadvantage of foreign banks might be a "home" bias of the local clientele. That is, customers (especially in retail banking) seem to prefer domestic banks over foreign banks simply because they are domestic. The existence of "home" bias can in part be attributed

find that the process of branching deregulation in the U.S. was mainly driven by the relative strength of interest groups of potential winners (large banks) and losers (small banks). Their analysis confirms that the same factors influenced congressional voting on branching deregulation.

[10]One could speculate that the recent merger between two Italian banks, Capitalia and Unicredito, could partially be a strategic response to the entry of ABN-AMRO in the Italian banking market.

to differences in national consumer-protection laws, which make it difficult for banks to standardize their products (see Dermine (2006)). In any case, even if home bias were not present, inertia on the part of customers together with real switching costs make it difficult for new banks to enter.

Another important barrier to entering a foreign country is the existence of asymmetric information. Particularly in small-business lending, domestic banks have more information about the business conditions in their country than their foreign competitors. Consequently, foreign competitors are initially at a disadvantage. Degryse and Ongena (2004) predict that asymmetric information induces more entry by mergers and acquisitions than by de novo entry, in which a bank opens branches in the foreign country on its own. Merging with a domestic bank could allow a foreign entrant to overcome the lack of domestic expertise and to gain local borrowers and local knowledge. Although entry via cross-border mergers has until now been fairly limited, Degryse and Ongena (2004) predict that eventually it will occur as targets become larger and their quality more verifiable (the competition authorities will also play a role).

In conclusion, *explicit* barriers to cross-border mergers do not seem excessive. Implicit barriers are still quite prevalent. The picture is, however, less clear due to the special nature of banking activities. Contrary to other industries, stability concerns may give domestic regulators an excuse to limit competition.

## 4.5 The Effect of Competition on the Effectiveness of Regulation

In order to understand how regulatory practices should be adapted to a more competitive environment, this section analyzes the *interaction* between competition and the effectiveness of regulation. The special nature of banking, and particularly the concerns about instability and its potential consequences for the rest of the economy, seems to continue to warrant regulatory scrutinity. However, competition poses unique challenges to the effectiveness and viability of regulation. Problems of regulatory arbitrage and level-playing-field issues come to mind. To this end, I review first the effect of competition on the effectiveness of deposit insurance. Next, I assess the influence of competition on capital regulation.

### 4.5.1 Competition and deposit insurance

As highlighted in Section 4.2, the justification for heavy regulatory interference in the banking industry is primarily to prevent costly systemic crises. Individual failures as a natural outcome of a Darwinian survival process without system-wide adverse consequences could be tolerated. With this in mind, I analyze how competition affects the effectiveness of deposit insurance and the distortions that deposit insurance induces. I show that the *design* of deposit insurance is particularly complicated in a more competitive environment. In particular, the coverage of deposit insurance needs to be carefully specified. Also, competition exacerbates the distortions in bank behavior that deposit insurance creates. To contain the distortions, deposit insurance could be designed more closely to reflect the risk of a bank.

It is true that competition promotes efficiency, but through which mechanism? Stiroh and Strahan (2003) show that high competition shifts the market share from low-to high-quality banks. Eventually, low-quality banks are forced to exit the market unless they choose to restructure themselves. In general, the regulator should endorse such efficient evolution of the industry. More importantly, as said, a possible failure of an individual bad bank can be tolerated if no adverse consequences spread to the entire banking system. Taking this into account, regulation and supervision should ideally go hand in hand with market discipline. Market discipline refers to borrowers, depositors, and other banks that should be cautious enough to spot possible failure and put pressure on banks to behave prudently.

Deposit insurance, however, shifts the burden of bank surveillance almost exclusively to the deposit insurer. Because the deposit insurer bears the potential losses of failure, depositors are no longer interested in the stability of the bank. However, this does not mean that deposit insurance must be eliminated. Without deposit insurance, costly bank runs may occur with a high cost for the entire economy, and indeed precisely the threat of costly bank runs was the reason d'être for deposit insurance.

Even without deposit insurance, implicit guarantees are prevalent. That is, high perceived costs of a systemic bank crisis may force the regulator to bail out failed banks and their creditors even in the absence of deposit insurance. An example for this is the rescue of the Bank of Crete by the Greek central bank in 1989 even though at that time Greece did not have any deposit insurance scheme. In the case of the Swedish banking crisis, formal deposit insurance was also not present, yet the government felt compelled to provide a blanket guarantee.

Gropp and Vesala (2004) claim that a bank might behave more safely in the presence of carefully designed explicit deposit insurance. They argue that deposit insurance tha is limited only to depositors commits the regulator not to bail out other creditors, such as large creditors or holders of subordinated debt. An example is Germany, where several small banks failed with explicit deposit insurance present (a recent example is the failure of Gontard and Metallbank AG in 2002). Only depositors were protected whereas other creditors incurred full losses.

What this may point to is that in a competitive environment the coverage of deposit insurance should be strictly contained and well specified, and should not cover all creditors. Some market discipline is then preserved. Some ambiguity might be desirable to encourage market discipline. This could be consistent with UK regulatory policy, in which the regulator issued a general policy that the regulator would step in only in case of failures leading to systemic bank fragility (see Financial Service Authority (2003)). By being ambiguous on when systemic fragility was deemed to be present, some ambiguity was created.

A related negative effect of deposit insurance is that it potentially changes bank behavior towards a more risky behavior. Competition could further encourage this. A potential solution is to make deposit insurance premiums risk sensitive. However, Chan, Greenbaum, and Thakor (1992) show that deposit insurance cannot be correctly priced to the bank's risk in a perfectly competitive environment due to informational problems. A fairly risk-priced deposit insurance can only be implemented if banks are subsidized through a deposit insurance premium. Freixas and Rochet (1998) show that a certain level of subsidization might be optimal to separate inefficient banks from efficient ones. However, subsidization may hinder competitive processes such as banks entry decisions.

Pennacchi (2005) demonstrates that risk-based deposit insurance also generates procyclical effects. He suggests that in order to minimize procyclicality deposit insurance should be structured as a moving average of contracts. In the U.S. the Federal Deposit Insurance Corporation (FDIC) is designing deposit insurance to more closely reflect the risks of the financial institution. The FDIC computes deposit insurance premium using capital levels, supervisory ratings, and also financial ratios and issuer ratings.[11]

Chapter 3 provides a more complete analysis of true effect of deposit insurance in an industrial organization model of banking in which banks are of a heterogeneous quality. It shows that, as long as the deposit insurance premium cannot be made fully type (and/or risk) dependent, deposit insurance effectively subsidizes low-quality banks relative to high(er)-quality banks. This makes low-quality banks more competitive than they would otherwise be, and makes it more difficult for good banks to gain market share at their expense.[12] The consequence of this is that lending rates are pushed down by the over-competitive low-quality banks, and this discourages entry. Increasing capital requirements mitigate this by reducing the deposit insurance subsidy for lower-quality banks, thereby reducing their competitive strength and encouraging entry.

In short, there is a need to reassess the design of deposit insurance in the highly competitive banking system. The easiest modification of deposit insurance is to introduce strictly contained coverage such that some of the creditors are left exposed. Theoretical studies show that the implementation of risk-based deposit insurance may be difficult, although some regulators have made attempts in this direction. The next section discusses how capital regulation can mitigate the distortions created by deposit insurance.

### 4.5.2 Competition and capital regulation

In order to limit distortions of deposit insurance, regulators have predominantly turned to capital regulation. Capital regulation demands that each bank has enough capital at stake, and this helps counter the tendency of deposit insurance to gamble. First, I focus on banking

[11] See Federal Register, Vol. 71, No. 230, p. 69282.

[12] Although lack of contractability generally makes it infeasible to have deposit insurance premiums fully risk-based (i.e., type and risk dependent) and effectively introduces cross-subsidies, systemic concerns in the banking industry create all kinds of other cross-subsidies and interdependencies. For example, many agree that the functioning of the banking sector depends crucially on the confidence that the public has in the financial system at large.

systems in developing countries, where competition may decrease the responsiveness of banks to capital regulation. Second, I analyze capital regulation in developed countries. There banks respond quickly to capital regulation and the design of capital regulation becomes important.

In Chapter 3 I give a positive view on the rather rudimentary Basel I capital framework. I show that the positive effect of Basel I was to increase the level of capital throughout the banking industry. As stated in Section 4.5.1, I argue that fixed-price deposit insurance effectively subsidizes weak banks; capital regulation helps counter this. It mitigates the artificially inflated competitive strength of weak banks, and in doing so helps mitigate the distortions associated with deposit insurance. When endogenous entry is allowed for, high capital requirements "cleanse" the banking system by reducing the competitive strength of weak banks, and in doing so can encourage entry. This positive effect of capital requirements is especially strong if competition is fierce and the distortions created by deposit insurance are high.

Chapter 3 also points to the critical aspect of capital regulation in connection with bank stability and interbank competition. It analyzes how effective capital requirements are at inducing banks to behave prudently. It shows that increasing interbank competition typically compromises the effectiveness of capital requirements precisely for low-quality banks (such that low-quality banks respond less to higher capital requirements). That is, competition makes capital requirements inefficient especially for the banks that need them most – for low-quality, risky banks. In Chapter 3, I also show that competition makes capital regulation less effective in weak banking systems whereas it strengthens the effectiveness of capital regulation in high-quality banking systems. These results point to the difficulty of introducing more competition when the banking system is of low quality.

Hellmann, Murdock, and Stiglitz (2000) go even further. They claim that high competition can make capital requirements counterproductive. They argue that costly capital decreases banks' franchise values. Banks that expect lower profits in the future care less about their future existence and their stability. Consequently, higher capital requirements might induce banks to behave more riskily.[13] Repullo (2004) mitigates their concerns. He shows that banks pass the losses due to costlier capital requirements to depositors. That is, an increase in capital requirements reduces equilibrium deposit rates in such a way that banks' franchise values do not change. He shows that capital regulation stays effective despite high competition. He also shows also that risk-based capital regulation is more efficient than flat-rate capital regulation.

Particularly in the developing countries, where corporate governance and institutional

[13]Several other contributions argue that capital requirements may enhance bank risk rather than lower it. Blum (1999) shows in an intertemporal setting that a bank may want to increase risk today if it anticipates an increase in capital requirements in the future. He argues that a unit of capital is worth more if capital requirements are tight. One way of increasing capital in the future is to boost profits by immediately taking riskier strategies. Besanko and Kanatas (1996) show that increasing capital requirements augments the proportion of outside ownership and lowers the proportion of inside (managerial) ownership. Consequently, managers employ less effort in assuring bank stability and banks might become less stable.

control are weak, the effectiveness of capital requirements might also be hampered due to agency problems. A bank manager that maximizes its own wealth instead of the shareholders' value might be inclined to take on considerable amount of risk despite a high level of external capital financing. Designing complex risk-adjusted capital formulas such as Basel II may not substantially improve the accuracy of capital regulation, yet it should result in higher regulatory costs. Particularly in the U.S., for these very reasons a different approach to capital is advocated. What is called a leverage ratio is defended as providing a good measure for timely intervention. Capital requirements then need to be robust, such that the regulator can correctly compute them and take action if capital is too low. Boot, Milbourn, and Deželan (2001) and Honohan and Stiglitz (2001) argue that in developing countries capital regulation may have to be combined with direct regulation that includes strong regulatory supervision, deposit interest rate ceilings, and limits on bank activities.

Whereas in developing countries with closed banking systems banks might be insufficiently responsive to capital regulation, in developed countries with fierce competition banks may be affected by capital regulation in an excessive and/or unforeseen way. For example, the imprecision of the rather rudimentary capital requirements used in the Basel I framework could spur regulatory arbitrage. That is, banks may try to avoid capital requirements by engaging in off–balance-sheet financing, such as securitization, and/or more risky activities that in the rudimentary Basel I standards do not require extra capital.[14] To prevent this, the response of regulators was to move to better calibrated risk-based capital requirements also involving the possibility for banks to use their own internal models to assess necessary capital. This led to the Basel II capital regulation framework.

However, competition might also render Basel II incomplete. In particular, the Basel II capital framework allows banks to opt for the more advanced, internal-rating-based approach (IRB), whereby banks develop their own empirical models to quantify required capital and the regulator limits its role to certifying those models. It is far from clear what the effects of the Basel II framework will be. First, it is not without risk to have capital regulation depend on the banks' own models. The necessary certification of those models might induce their standardization and the potential invalidity of such models might lead to a newly induced systemic risk. Second, the implementation of the IRB framework is demanding and suitable only for large banks and it might give them a comparative advantage against small banks. This might deter new entry and lower competition. Alternatively, banks might use the "standardized" approach although they have already implemented the IRB framework for their own use.

Moreover, at a more fundamental level, one could question whether capital regulation should aim at precisely calibrating optimal capital levels, or rather enforce a strict lower limit essentially as a certification requirement. In the latter interpretation, one could even

[14]Greenbaum and Thakor (1987) show that banks tend to securitize safe loans whereas they typically keep risky loans on their balance sheets . Calomiris and Mason (2004) empirically confirm that through securitization banks may seek to avoid minimum capital requirements. Jones (2000) reviews the principal techniques for regulatory capital arbitrage invoked by the Basel I standards.

envision banks voluntarily choosing higher levels of capital because being well capitalized may also offer a distinct competitive advantage.

An additional negative effect of risk-based capital regulation is that it might act procyclically (see Kashyap and Stein (2004)). That is, in a recession banks must hold higher capital requirements due to greater risk of their borrowers. Higher capital requirements may then limit access to bank loans, which might deepen the recession.

However, Basel II also has positive features. It attempts to better match pricing in the capital markets by limiting (costly) capital to risks. This may help enhance transparency and market discipline. Indeed one of the objectives of Basel II is to augment market discipline. Its third pillar on disclosure and transparency alo explicitly aims to facilitate market discipline. Particularly in the more competitive environment of a developed banking system, the benefits of correct alignment of incentives in Basel II together with augmented market discipline might be welcomed.

Summing up, competition exacerbates the distortions of bank regulation. In particular, deposit insurance has two distortionary effects. First, it puts banks on an equal footing and helps low-quality banks. Second, deposit insurance may increase banks' risk taking. A careful design of deposit insurance (i.e., limited, well-specified coverage of deposit insurance) could partially contain its distortions and at the same time prevent a costly bank run. Capital regulation may additionally mitigate the distortions of deposit insurance. If competition is high, capital regulation is less effective in the developing banking system and it needs to be complemented with direct regulation; that is, regulatory supervision and intervention. Capital regulation is more effective in a developed banking system. However, the competitive environment makes its implementation far from trivial.

## 4.6 Regulation in the Presence of Consolidation

Regulators need also to adapt to the observed consolidation in banking, which has elevated the complexity of these financial institutions. In this section, I assess the viability of regulation in light of this elevated complexity. I focus first on the causes and effects of consolidation. Next, I evaluate the effectiveness of regulatory tools for these more complex institutions.

### 4.6.1 The effectiveness of current regulation in the consolidated banking industry

The efficiency of the current regulatory design should be reconsidered in light of the ongoing consolidation in the banking industry. First, I show that consolidation may positively affect bank efficiency and stability. Second, based on the situation in the EU banking system, I highlight several drawbacks of consolidation. To account for this, I propose a few potential improvements in the regulatory design.

Recent research points to the presence of scale economies in many bank activities. However, mergers typically involve not only scale economies but also increased organizational complexity. Not surprisingly many studies find little overall value gains in mergers, with often overall negative effects. More recently, mergers up to $10 billion of total assets appear to offer some potential value creation. Typically however, the potential for value gains is increasing with the geographical and product specialization of the institutions involved. Penas and Unal (2004) and Focarelli and Panetta (2003) point to the efficiency gains from bank mergers that stem mostly from diversification and synergies between merging banks. In the short run, merged banks obtain higher market power; however, in the long run efficiency gains may dominate market power and consumer surplus could increase. The recent development of IT technology facilitating cross-selling of new products and running complex risk-measurement systems have somewhat improved the value gains in large mergers.

Empirical research describes a positive connection between concentration and stability. Beck, Demirgüç-Kunt, and Levine (2003) confirm that in more concentrated banking systems crises occur less often. They suggest that concentration should be viewed separately from competition. That is, the most stable banking system is one with strong competition between a few large players.

However, consolidation of the banking industry should still be considered with caution. Consolidation could have a negative influence on the effectiveness of bank regulation – particularly timely intervention. Increasing complexity might make this much more difficult; moreover, too-big-to-fail concerns may lead to regulatory paralysis. Large financial institutions might have a strong influence on the action of regulators through the political powers that they control.[15] This "regulatory capture" can be prevented with enhanced transparency of regulatory procedures combined with strong market pressure.[16]

Consolidation might also create a lack of accountability due to the involvement of too many regulators. For example, some European banks must report to more than 20 European regulatory bodies. Who should be responsible for resolving the failure of such a bank? The regulator of the institution's home country might have incentives not completely aligned with other regulators. It could postpone the liquidation of an insolvent institution for too long in fear that such liquidation could trigger instability in the domestic banking industry; the "too-big-to-fail" problem.[17] Cross border activities with potential different systemic concerns in host versus home countries may exacerbate this problem. This makes it even more important that regulatory tools and practices such as closure policy be carefully designed. Only then can regulators be held accountable for their actions.[18]

The question is also who will act as a lender of last resort (LoLR). That is, who will

[15]See Kroszner (1999) for more on the political economy of banking regulation.

[16]Regulators should put special focus on the stability of large institutions because they have the greatest effect on the stability of the entire banking system (see Mishkin (1999)).

[17]Boot and Thakor (1993) show that the regulator's privately optimal bank closure policy is more lax than the socially optimal closure policy. They argue that the regulator concerned about its reputation (its monitoring ability) might postpone liquidation of an insolvent financial institution.

[18]Heinemann and Illing (2002) show that increased transparency of government policy reduces the likelihood of speculative attacks.

provide liquidity in the case of a liquidity crisis? In the EU the national regulator is legally responsible for the provision of liquidity. Again, optimal liquidity provision for one country might not be optimal for foreign operations.[19]

Despite possible political problems, the EU should think in the direction of at least partially centralizing the regulatory powers in European institutions such that the effective procedures in the case of failure of an international bank can be followed. Optimal synchronization across regulators and supervisors is crucial (see Vives (2001b) and Boot (2006)). Dell'Ariccia and Marquez (2006) show that, if regulation is decentralized, the regulators in competition select sub-optimally low standards. Although centralized regulation overcomes this problem, it entails a loss of flexibility in designing optimal regulation that suits the needs of each banking system separately. If banking systems are homogeneous enough, centralized regulation prevails over decentralized regulation.

Another problem is that consolidation complicates the provision of deposit insurance. In the EU, a bank is regulated by the home country regulator, including foreign branches. Accordingly, the home country regulator also insures deposits (in the home and host countries). This creates two problems. First, deposit insurance schemes are not fully harmonized across countries in the EU. Banks from different countries therefore compete with each other under different conditions.[20] More importantly, in small countries with large banks and sizable foreign operations the credibility of such deposit insurance schemes comes into question because those countries may not want to bear the burden of potential losses to sizable foreign depositors.

Regulation of a giant financial institution is not only a problem for the EU. The formation of globaly operating financial institutions such as Citygroup poses identical problems. It is important that financial institutions face similar regulatory and supervisory requirements in different countries. Regulatory arbitrage might otherwise induce regulators to compete for banks by possibly lowering standards compromising the effectiveness of regulation (see Dell'Ariccia and Marquez (2006)). This also has some implications for the switch from Basel I to Basel II. A positive effect of the Basel I framework was that it created an equal level playing field for banks in different countries. The Basel II framework contains not only more discretional elements but may also not be introduced in all countries and, even if introduced in a country, it may not apply to all banks.[21] These differences elevate regulatory compliance costs and could induce regulatory arbitrage.

[19] In current EU regulation, the home country regulator is responsible for regulating domestic banks *and* its branches in other countries. It also provides deposit insurance for depositors of the bank in home and host countries. However, a branch might be important for the stability of the banking system of the host country. Hence, reporting to home and host country regulators is important.

[20] Due to the Deposit Guarantee Schemes (DGS) Directive (94/19/EC), EU countries are required to maintain a minimum coverage of deposit insurance of €20,000; however, they are free to set coverage above the minimum level. Further harmonization is needed, including fine tuning "topping up" arrangements (in which a bank branch in another member state voluntarily joins the host country's deposit guarantee system); shortening the time it takes for schemes to pay out to depositors after a bank failure and improving exchange of information between schemes (see European Commission (2006a)).

[21] The EU has already introduced this in its directives 2000/12/ES and 93/6/EES. The U.S. is cautiously lagging behind, and has announced that it will only apply it to the largest banks.

In brief, although consolidation may make banks more efficient and more stable, it may impinge the existing regulatory design. In particular, consolidation may hamper the power of the regulator to act against an insolvent (or excessively risky) bank. In the EU, the coexistence of several regulators with dimmed accountability may become critical in the case of the failure of a large cross-country bank. Further centralization of regulatory powers might be necessary. The existing design of deposit insurance may also turn out to be inadequate.

### 4.6.2 Current regulatory tools and the increasing scope of financial conglomerates

The removal of regulatory barriers may have induced banks to expand scope. Therefore not only the size is increasing, but also the scope. I analyze what drives conglomeration and its impact on regulation. First, I review potential efficiency gains and diversification advantages of conglomeration. Second, with respect to the optimal regulatory design, I stress the difficulties of regulatory supervision and point to the role of market discipline.

The empirical literature shows that diseconomies of scope are quite prevalent (see Saunders (2000)). Diversification in general destroys value (for studies on diversification discount, see Berger and Ofek (1995)). The reason behind conglomeration might well be managers' eagerness to grow large and powerful. Managers might also defend themselves from being acquired by acquiring other companies first.

One of the reasons that banks enter other activities is to diversify the risks of the banking business. It is difficult for the regulator to judge the risks stemming from such complex institutions. Moreover, banks increasingly engage in financial market activities such as securitization. Although this may reduce risks (assets might be removed from the balance sheet), it often also adds risks because banks take positions in the financial markets. This may introduce risks that originate from the interconnections between banks and financial markets. In addition, securitization has partially transferred credit risk from banking to the insurance industry. As Allen and Carletti (2005) argue, an idiosyncratic shock to the insurance industry can potentially be propelled back to the banking system, endangering its stability.

Herring (2002) argues that the regulators are afraid of closing financial conglomerates because of the complexity of such institutions; for example, because it is too complex to assess the costs of liquidation and its potential effects on other financial institutions. Hence, regulators might be inclined to rescue them. A potentially similar argument points to the difficulties in closing a financial holding company with many subsidiaries in different countries. In that case, the complexity is further elevated by uncertainty about which country's bankruptcy law should apply in closing such an institution.[22]

[22]In closing a bank with branches in several EU countries, the bankruptcy law of the home country applies (the directive on Winding up of Credit Institutions, Official Journal 125, 5 May 2001). In the case of a subsidiary, this is different. However, even in the case of branches the host country may try to ring fence assets in order to protect its local constituency.

Regulation of financial conglomerates may call for a greater role of market discipline. However, opaqueness is a serious problem. This already applies to traditional banking business. Banks are inherently opaque institutions. Conglomeration exacerbates opaqueness, which might limit the scope for market discipline. Nevertheless, market discipline together with increased transparency deserves to be one of the cornerstones of regulation in a more competitive banking industry.[23] Without market discipline, the regulator would have to carry the entire burden of controlling financial stability, which could lead to the previously mentioned problems of regulatory capture and regulatory arbitrage.[24]

To summarize, some evidence shows that consolidation in banking leads to efficiency gains and to a more stable banking system. However, consolidation presents several potential dangers for bank stability. First, a large financial institution may have powers to influence the regulator, which results in regulatory capture. Second, its failure could trigger as system crisis; that is, it may become too large (or too complex) to fail, and the regulator could be forced to bail it out in the case of failure. In dealing with these dangers, the regulatory arrangements are important. In this respect, the situation in the EU is not encouraging. Regulation in the EU is decentralized between many regulators that lack clear accountability. Potential further cross-border integration between EU banks may call for regulatory powers to be at least partially centralized in European institutions.

## 4.7 Conclusions

Competition challenges both banks and regulators. Banks will seek to optimally position themselves to deal with a more competitive environment; regulators need to critically assess the suitability of regulating practices in this kind of more competitive environment. Although regulators should not work against these competitive forces, the stability of the banking system needs to be safeguarded. In Section 4.3 I show that particularly in developing countries the regulator may have difficulties in preventing banks from taking excessive risks if competition heats up. In developed countries, competition may generally help both efficiency and stability. Accordingly, regulators may aim to reduce direct and indirect barriers to competition. In Section 4.4 I have described the process towards a more competitive banking system, using lessons from the U.S. and EU situations.

Section 4.5 shows that, together with promoting competition, regulators must adopt stability-oriented regulation. To prevent regulatory distortions, regulators need to fine tune regulation, yet seek to limit regulatory interference. More risk-based regulatory interventions may help. Examples include risk-adjusted deposit insurance and risk-adjusted capital regulation. However, competition changes the banking environment quickly and even fine-tuned regulation is at risk if it runs counter to market forces. This also points to critically assessing the ambition of regulation. It is crucial that players fulfill some minimum requirements to

[23] Market discipline is also one of the cornerstones of Basel II capital regulation, as stated earlier.

[24] Barth, Caprio Jr., and Levine (2004) show that banking systems are most stable when the biggest role is given to market mechanisms that help enforce prudent bank behavior.

operate, such as minimum capital requirements. The exact fine-tuning might be of lesser importance.

Bank regulation will continue to play a critical role. The potential costs of banking crises are high and substantial externalities exist. In Section 4.6 I have highlighted the importance of clear accountability of regulators in their supervisory review process in light of the consolidation of the banking industry, and the increasingly prevalent cross-border and cross-sector operations of banks. Regulators also need to increasingly use market discipline and enhance transparency to alleviate the burden that falls on supervisors and regulators. Market discipline should operate in conjunction with supervision.

In future work, the impact of competition on the efficiency of capital regulation and deposit insurance deserves further study. Although I have provided some indications of what the structure of regulation should be, for the time being we know little about the optimal design of regulation in a competitive and dynamic banking sector.

# Chapter 5

# Competition and Entry in a Spatial Model with Applications to Banking

**Abstract**

This chapter recasts the analysis of Chapter 3 in a Hotelling framework. It analyzes entry in a spatial model in which firms are heterogeneous in their production costs. The key result is that intensifying competition by lowering transportation costs can augment expected profits and entry, essentially by shifting the market share from high-cost to low-cost firms. It is also shown that increasing the differences in firm costs augments firm profits, entry, and social welfare. Lowering production cost of high-cost firms augments entry if competition is low (i.e., for high transportation costs) but hampers entry if competition is high (i.e., for low transportation costs). When specifically applied to banking, the results of the analysis show that *i*) deposit insurance may lead to less entry, *ii*) capital regulation may lead to more entry in banking, and *iii*) bailout policies may reduce bank profits and entry, even if they come at no additional cost for banks.

Keywords: Entry, Spatial Competition, Transportation Costs, Banking
JEL CLASSIFICATION: L11, G21

## 5.1 Introduction

The perceived distances between firms and consumers has decreased substantially due to developments in IT and the lifting of regulatory barriers. Firms operate in increasingly competitive and dynamic environments in which impediments to cross-border entry are low. In this chapter I pose two questions: *i)* how does the level of interfirm competition in the market (measured by the level of transportation costs) affect a firm's entry decision in such a market; and *ii)* how do the changes in firm heterogeneity affect firm profits, entry, and social welfare.

I analyze these issues in a spatial monopolistic competition setting which distinguishes two types of firms, that differ in their cost of production. In particular, high-cost firms incur higher per unit production costs than low-cost firms. Firms are allowed to compete for consumers (located uniformly on a unit circle). The level of competition is determined by the size of transportation costs that consumers incur when traveling to the firms. This chapter recasts the analysis in Chapter 3 in a Hotelling framework. While the analysis has generic interest in most of the application, I will look at the banking industry; see Degryse and Ongena (2005) for the relevance of "distance" to bank competition.

The key result is that increasing competition in the market through lowering transportation costs might be beneficial for entering firms and could bring more entry essentially by redistributing a market share from high-cost to low-cost firms. The shift in the market share could augment the expected profits of entering firms. Although the redistribution of the market share that lowering transaction costs entails is beneficial for entering firms, the net effect on entry could go either way because competition also transfers rents from firms to consumers. I establish conditions for which competition encourages entry. Roughly speaking, higher competition in terms of lower transportation costs has a positive effect on entry provided there are sufficiently many (but not too many) low-cost firms in the economy *and* competition is already high (i.e., entry costs should be sufficiently low). I show that for such intermediate quality industries increasing competition reduces the competitive strength of high-cost firms and also encourages entry.

Salop (1979) argues that lowering transportation costs increases competition and hampers firm profits. This then results in a lower number of entering firms. However, Salop shows that this intuitive result reverses for high transportation costs (i.e., at lower levels of competition). In particular, a "perverse" equilibrium may exist, in which lowering transportation costs leads to higher profits and more entry. This is because (almost) monopolistic firms spend less to compensate consumers on transportation costs. In particular, high transportation costs prevent most of consumers to travel to a competing firm. The firm may act (almost) as a monopolist and may charge to consumers (most of) what they are willing to pay for the products. If transportation costs decrease, consumers are willing to pay more for the same product knowing that they incur lower traveling costs when buying it. This underscores that entry is a non-monotonic function of transportation costs in homogeneous industries (i.e., industries in which firms have equal production costs). More specifically, entry increases in transportation costs for low transportation costs but decreases in transportation costs for high transportation costs.

This approach deviates from the extant spatial competition literature in that it recognizes differences in cost efficiency among firms.[1] Allowing for heterogeneity (i.e., different production costs) among firms may again reverse the standard result in Salop (1979). Another "perverse" equilibrium may appear, this time for low transportation costs (for high levels of competition). More specifically, if competition is fierce, a decrease in transportation costs may augment entry. The rationale for this is the following. Higher competition improves the total efficiency of the industry through the shift of market share from high-cost to low-cost firms. This then results in higher firm profits and higher entry.[2] If competition is low, however, efficiency increases only slightly and the standard Salop effect is at work; namely, profits go down due to the negative effect of competition on firm rents. In sum, entry is a U-shaped function of the level of transportation costs; that is, entry decreases in transportation costs for low transportation costs and increases in transportation costs for high transportation costs.

Melitz (2003) analyzes the intra-industry effects of international trade using a Dixit and Stiglitz (1977) model of competition extended for firm heterogeneity. He shows that opening up borders for trade increases the total productivity of firms, essentially by a reallocation of resources from low-quality to high-quality firms. In his model, firms first enter the industry and later enter export markets. He shows that a decrease in trade costs generates entry of new firms into the export markets. At the same time, firms exit the industry. He does not offer an answer for how trade costs influence the total number of firms in the industry. The main analysis here complements Melitz's research by showing that *initial entry* into the industry may increase and consequently the total number of firms may go up with lower costs of trade.

I also analyze the effect of firm heterogeneity on firm profits, entry, and social welfare. I show that increasing the differences in costs between firms may be valuable for firms even though it lowers their cost efficiency. The reason for this is that higher differences between firms lower the level of competition. A firm's profits may depend only on its relative advantage to its competitor. More specifically, the bigger the differences in firm costs, the higher rents firms can extract from their consumers. Whereas lower cost efficiency may be passed on to consumers, increased differences between firms augment firm profits and entry. Even more, increasing the cost differences between firms may also improve social welfare. That is, competition increases the efficiency of the industry through the shift of the market share from high-cost to low-cost firms. For more heterogeneous firms, a redistribution of market share is more pronounced and, consequently, social welfare increases.

I further analyze the role of cost differences to show that increasing the competitive

[1] Empirical work shows that the magnitude of variation in cost efficiency at the plant-level is enormous and persistent even within narrowly defined industries (see Bartelsman and Doms (2000)).

[2] Recent literature that discusses the positive relation between competition and productivity includes Aghion and Howitt (1992), Nickell (1996), and Raith (2003).

ability of low-cost firms always augments entry, whereas enhancing the competitive ability of high-cost firms might hamper entry. That is, improving the cost efficiency of a low-cost firm augments the average firm rents and also increases the differences between firms. Both effects augment profits and entry. However, enhancing the cost efficiency of a high-cost firm lowers the differences in costs between firms and hampers the shift of market share from high-cost to low-cost firms. If competition is high, and the industry is of high quality (i.e., there are many low-cost firms), this results in decreased profits and entry.

I apply the results to banking. More specifically, I provide a spatial setting for the analysis of Chapter 3, which was derived in the context of capital regulation in banking. Chapter 3 models bank competition in a simplified search model. This chapter shows that similar results can also be derived in a spatial context.[3] Whereas Chapter 3 exclusively focuses on the competition effects of capital regulation, I focus here on a broader set of regulatory practices in banking. I first explore the impact of branching deregulation on entry in banking. Interestingly, banks on average may have incentives to push for branching deregulation. Branching deregulation brings beneficial redistribution of market share to better banks. In spite of higher competitive pressure, this could improve profitability of those banks.

I also show that policies enhancing market disclosure and market pressure should be the most valuable in highly competitive banking industries. That is, market disclosure and market pressure enhance the differences between banks. This is especially valuable if competition between banks is high. That is, competition shifts the market share to the most efficient banks. This leads to higher profits and augments social welfare.

Deposit insurance decreases the differences between banks as it mostly helps weak banks. I show that in highly competitive, high-quality banking systems banks on average may lobby for limits on deposit insurance. To the contrary, in less competitive banking markets banks may prefer to have more generous deposit insurance.

I also argue that extensive bailout policies augment entry in banking if competition is low but hamper entry if competition is high. Bailout policies essentially lower the costs of financing for banks. Especially, high-cost (riskier) banks obtain funds at better terms than without bailout policies. If competition is low, this increases bank profits and augments entry. However, in the highly competitive banking industry increasing the competitive strength of riskier banks has a negative effect on entry. More specifically, bailout policies unfairly help high-cost (risky) banks and this is detrimental for low-cost (safe) banks. Low-cost banks cannot compete on a fair ground with high-cost banks, which lowers expected profits and hampers entry. Consequently, the banking industry as a whole may push for more stringent bailout policies.

The simplicity of this model does not come without some concessions. First, in this model

[3]A spatial model may also literally apply to banking, where geographical distances between banks and borrowers could play a role. For example, Degryse and Ongena (2005) show that distances between borrowers and their banks, (as well as distances between borrowers and competing banks) play an important role in the Belgian banking system.

entering firms are not aware of their types. In reality one would expect entering firms to have at least some prior knowledge about their costs. Second, firms in this model incur an entry cost only when entering. Later, this cost is sunk and high-cost firms do not exit the market. However, this aligns with empirical research that indicates that newly entering firms have higher failure rates than incumbents (see Dunne, Roberts, and Samuelson (1989)). This suggests that entering firms first make irreversible investments in production and only later discover their productivity (their type).

The chapter is organized as follows. In Section 5.2 I develop the model. Section 5.3 contains some basic analyses. In Section 5.4 I present the main findings. I first discuss the model with homogeneous firms. Later I extend it to heterogeneous firms. Section 5.5 contains the application to banking. Section 5.6 concludes this chapter.

## 5.2 The Model

### 5.2.1 Preliminaries

There are two players in the model: firms and consumers. Firms produce products that consumers buy. Firms differ in the variable cost of production. In particular, low-cost firms have low per-unit costs of production whereas high-cost firms incur high production costs.

I describe competition with a spatial model of price discrimination. Consumers are located on a unit circle. Firms compete for each consumer by setting a price that depends on the consumer's location. That is, firms can perfectly price discriminate between differently located consumers which is a typical situation in banking (see Degryse and Ongena (2004) (2006), Petersen and Rajan (2002) and Agarwal and Hauswald (2007)). The results could also be replicated in the uniform pricing model (as used by Salop (1979)) in which firms would set one price for all consumers regardless of their location.

The consumer decides on a firm by comparing price offers and firm locations (and corresponding transportation costs). In particular, a consumer located further away from the firm is prepared to pay less for its product because of two reasons. First, he has to overcome higher transportation costs to reach the firm. Second, he is located closer to a competing firm, which makes an offer of a competing firm more attractive.

The level of competition is determined by the level of transportation costs. If transportation costs are high, traveling to the distant firm might be prohibitively expensive and consumers might be locked in to the closest firm. For lower transportation costs, consumers are more willing to consider an offer of a more distant firm. Consequently, firms extract lower rents from consumers. This horizontal differentiation of firms due to transportation costs is combined with vertical differentiation determined by differences in firms' production costs.

$t = 0$:

♠ Firms enter the industry at a cost $F$.

♠ Firms distribute themselves uniformly on the unit circle.

$t = 1$:

♠ Firms discover their types.

♠ Each firm offers a distinct price to each consumer.

♠ Each consumer chooses a firm to purchase a product from.

♠ Firms produce and consumers make their purchases.

Figure 5.1: Timeline

### 5.2.2 Model details

*Timing and information structure:* There is universal risk neutrality, with a discounting factor normalized to 1. There are two dates, $t = 0$ and $t = 1$. At $t = 0$, firms enter the industry at a cost $F$. Upon entering, a firm does not know its type. Firms locate symmetrically on the unit circle. At $t = 1$, firms and consumers discover other firms' types. Consumers then decide from which firm they will purchase a product. The chosen firm commences with production. Finally, each consumer travels to the chosen firm to collect the product. Figure 5.1 summarizes the sequence of events.

*Firms:* Firms choose to enter the industry at $t = 0$. All firms are initially (perceived as) identical. Entering $N$ firms distribute symmetrically on the unit circle.[4] At $t = 1$, firms learn whether they are high-cost ($B$) or low-cost ($G$) firms. Low-cost firms have a lower variable cost of production $c_G$ than high-cost firms $c_B$; that is, $\Delta c \equiv c_B - c_G > 0$. The cross-sectional probabilities of having high costs ($\gamma$) or low costs ($1 - \gamma$) are known to all.

*Consumers:* Consumers are uniformly distributed on the unit circle. Their total mass is normalized to one unit. A consumer has inelastic demand and purchases 1 unit of a product if its price is lower than the reservation price $Y$. The effective price for a product equals the factory door price set by the producer plus a transportation cost of $\tau$ per unit of distance that the consumer incurs when purchasing the product.

*Competitive Environment:* Competition between firms occurs at $t = 1$ when the types of the firms are already public knowledge. Each firm makes a (potentially different) price offer to each consumer. That is, firms can price discriminate between consumers on the basis of their locations. Each consumer compares price offers of firms with their distances and accepts the best offer. The chosen firm then starts producing. Consumers travel to the chosen firm, incurring transportation costs, and buy the product (see also Figure 5.1).

[4] Because all firms are perceived as identical at that moment, this even distribution of consumers over all firms is quite natural. Bhaskar and To (2004) show that firms locate symmetrically in the circular model with perfect price discrimination. In the uniform pricing model, Lederer and Hurter (1986) show that, although firms never locate coincidentally in the Hotelling model, they locate symmetrically on the circular model.

## 5.3 Some Preliminaries

I solve the model through backward induction. First, I compute a firm's pricing decision at $t = 1$, holding the number of competing firms fixed at $N$. Subsequently, I allow for entry. Last, I compute social welfare.

### 5.3.1 Price offers and profits

Firms are located symmetrically with a distance $\frac{1}{N}$ between adjacent firms. Competition between two adjacent firms of type $i$ and type $j$ located at 0 and $\frac{1}{N}$ that set prices $p_i$ and $p_j$ evolves as follows. A consumer that is separated at distance $x$ from a firm of type $i$ and $\frac{1}{N} - x$ from an adjacent firm of type $j$ is indifferent between the offers of both firms if

$$Y - p_i - \tau x = Y - p_j - \tau[\frac{1}{N} - x]. \tag{5.1}$$

The lowest price that firms are prepared to offer equals their variable costs. The most distant consumer that a firm of type $i$ can obtain when bidding against a firm of type $j$ is at a distance $\bar{x}$, where $\bar{x}$ follows from (5.1); that is, substitute $p_i = c_i$ and $p_j = c_j$ in (5.1) to obtain

$$\bar{x} = \frac{c_j - c_i}{2\tau} + \frac{1}{2N}. \tag{5.2}$$

I now make the first assumption (recall that $\Delta c = c_B - c_G$).

**Assumption 5.1.**: $\Delta c < \frac{\tau}{N}$.

This assumption guarantees that the difference in cost efficiency between firms is small enough such that even high-cost firms obtain some market share. More specifically, a high-cost firm offers at most $Y - c_B$, whereas an adjacent low-cost firm could offer more, that is, $Y - c_G$. However, Assumption 5.1 guarantees that a high-cost firm can attract at least the closest consumer because of its proximity advantage; that is, $Y - c_B > Y - c_G - \frac{\tau}{N}$. Hence, the consumer in the location of a high-cost firm does not switch to an adjacent low-cost firm. This also implies that each firm competes only for the consumers located between itself and an adjacent firm, but not for consumers on the other side of an adjacent firm.

Now I compute optimal price schedules. I separate two cases. First, if competition is high (i.e., $\frac{\tau}{N} \leq Y - c_j$), a firm of type $j$ can reach all consumers between itself and an adjacent firm. That is, a firm of type $i$ can make an acceptable offer conditional on the adjacent firm not bidding. In particular, the right side of (5.1) is positive for $p_j = c_j$ and for $x \in (0, \frac{1}{N})$. A firm of type $i$ uses its advantage in transportation costs such that it charges price $p_i(x)$ for consumers located at the distance $x \in (0, \bar{x})$. I then insert $p_j = c_j$ into (5.1) and compute for $p_i$ to obtain

$$p_i(x) \quad = \quad c_j + \frac{\tau}{N} - 2\tau x. \tag{5.3}$$

A consumer at $x$ is indifferent between a firm of type $i$ offering $p(x)$ and the best offer that an adjacent firm of type $j$ can make, (in such a case, the consumer decides on a firm of type $i$).

Second, if competition is low (i.e., $\frac{\tau}{N} > Y - c_j$), an adjacent firm of type $j$ cannot reach the most distant consumers located closest to a firm of type $i$. We define

$$\underline{x} = \frac{1}{N} - \frac{Y - c_j}{\tau}. \tag{5.4}$$

Consumers located very close to a firm of type $i$ (i.e., $x < \underline{x}$) are located too far from an adjacent firm of type $j$ to buy its product regardless of its price. That is, the transportation cost $[\frac{1}{N} - x]\tau$ to reach a firm of type $j$ is greater than the consumer's rent $Y - p_j$ even if it charges the lowest price $p_j = c_j$. Those consumers cannot switch firms; hence they accept the monopolistic price of the firm of type $i$. In particular, the firm charges the reservation price of the consumer $Y$ lowered by $\tau x$ to compensate consumers for their transportation costs. The price schedule is

$$\hat{p}_i(x) = \begin{cases} c_j + \frac{\tau}{N} - 2\tau x & \text{if } x > \underline{x}, \\ Y - \tau x & \text{if } x \leq \underline{x}. \end{cases} \tag{5.5}$$

I now make the second assumption.

**Assumption 5.2.**: $Y - c_B - \frac{\tau}{2N} > 0.$

Assumption 5.2 guarantees that the reservation value of the product is high enough such that the consumer in the middle between two adjacent high-cost firms is prepared to purchase a product. This guarantees that the entire market is covered and that two adjacent firms compete for at least one consumer in the middle between them.

Now I compute firm profits. I again consider two cases. For $\frac{\tau}{N} \leq Y - c_j$, the price schedule in (5.3) applies and the expected profit of a firm of type $i$, competing with two adjacent firms of types $j$, is

$$\Pi_{i,j}(N) = \int_{-\bar{x}}^{\bar{x}} [p_i(x) - c_i] \mathrm{d}x, \tag{5.6}$$

where $\bar{x}$ is as given in (5.2). For $\frac{\tau}{N} > Y - c_j$, the price schedule in (5.5) applies and the expected profit is

$$\Pi_{i,j}(N) = \int_{-\bar{x}}^{\bar{x}} [\hat{p}_i(x) - c_i] \mathrm{d}x. \tag{5.7}$$

Inserting (5.3) into (5.6) and (5.5) into (5.7) yields[5]

$$\Pi_{i,j}(N) = \begin{cases} \frac{\{\tau+N[c_j-c_i]\}^2}{2\tau N^2} & \text{if } \frac{\tau}{N} \leq Y - c_j, \\ \frac{2Y-c_j-c_i}{N} - \frac{\tau}{2N^2} + \frac{[c_j-c_i]^2-2[Y-c_j]^2}{2\tau} & \text{if } \frac{\tau}{N} > Y - c_j. \end{cases} \tag{5.8}$$

### 5.3.2 Endogeneous entry

We now endogenize $N$, and hence allow for entry. The entry decision is made at $t = 0$. At that moment, each prospective firm does not yet know its own (future) type, but assesses its expected production costs based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$ of being a low-cost or high-cost firm. Each firm computes whether its expected profits from entering exceed the cost of entry F, anticipating the competitive environment (including the number of firms already present).[6]

The expected firm's profit after entering (i.e., after the entry cost $F$ has already been incurred) is

$$E(\Pi(N)) = \gamma^2\Pi_{G,G} + \gamma[1-\gamma][\Pi_{G,B} + \Pi_{B,G}] + [1-\gamma]^2\Pi_{B,B}. \tag{5.9}$$

I can now prove the following lemma (insert (5.8) into (5.9)).

**Lemma 5.1.** *The expected profit of an entering firm is*

$$E(\Pi) = \begin{cases} \frac{\tau}{2N^2} + \frac{\gamma[1-\gamma]\Delta c^2}{\tau} & \textit{if } \frac{\tau}{N} \leq Y - c_B, \\ \frac{2[Y-E(c)]}{N} - \frac{\tau}{2N^2} - \frac{[Y-E(c)]^2}{\tau} & \textit{if } \frac{\tau}{N} > Y - c_G, \\ \gamma\frac{\tau}{2N^2} + [1-\gamma]\{\frac{2[Y-c_B]}{N} - \frac{\tau}{2N^2} + \frac{\gamma\Delta c^2-[Y-c_B]^2}{\tau}\} & \textit{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G, \end{cases} \tag{5.10}$$

*where $E(c)$ is defined as the expected production cost, that is $E(c) = \gamma c_G + [1-\gamma]c_B$.*

Lemma 5.1 shows that three regions of competition are distinguished. In the most competitive region, ($\tau \leq [Y - c_B]N$), adjacent firms can attract all consumers between them, even if firms are of high-cost type. In the least competitive region ($\tau > [Y - c_G]N$), adjacent firms act as monopolists for at least some of the closest consumers even if they are of low-cost type. The region with intermediate competition ($Y - c_B < \frac{\tau}{N} \leq Y - c_G$) provides the most fruitful results. Here low-cost firms can reach all consumers, but high-cost firms cannot reach the most distant ones. That is, if a firm competes with a low-cost firm, it must compete for

[5]Thisse and Vives (1988) similarly observe two different price schedules. In my model, the demand for each consumer is inelastic and the price always declines with distance. Thisse and Vives (1988) allow for elastic consumer demand and show that prices may initially rise but could subsequently decline with distance. Firms then charge low prices to consumers in order to increase demand. High transportation costs make the demand from more distant consumers less valuable, and therefore firms charge them higher prices. For consumers located at even further distances, prices may start declining. This is because firms need to compete for distant consumers, which bids down the price levels.

[6]I consider only the following simple entry procedure: firms decide on entering sequentially in random order. Note that the order does not matter because all prospective entering firms are identical and assess their production costs based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$.

all consumers. Yet, if it competes with a high-cost firm, it behaves as a local monopolist for the closest consumers but competes for more distant ones.

To prevent complexity due to discreteness in the number of firms, I allow $N$ to be a continuous variable, such that $N^*$ is determined by the equilibrium condition:

$$E(\Pi(N^*)) = F. \tag{5.11}$$

This provides

**Lemma 5.2.** *The expected profits are decreasing in the number of entering firms.*

Intuitively, a higher number of firms reduces the market share of each firm, diminishing firm profits. This lemma guarantees that there exists a unique equilibrium number of entering firms $N^*$.

### 5.3.3 Social welfare

I compute social welfare as the sum of consumer and producer profits. This equals the reservation price of the products minus transportation costs, production costs, and entry costs. Social welfare per one firm is

$$W_{i,j} = \int_{-\bar{x}}^{\bar{x}} [Y - c_i - \tau x]\mathrm{d}x - F. \tag{5.12}$$

Hence, the total social welfare is

$$TW = N \times E(W_{i,j}) = N\{\gamma^2 W_{G,G} + \gamma[1-\gamma][W_{G,B} + W_{B,G}] + [1-\gamma]^2 W_{B,B}\},$$

which yields

$$TW = -\frac{\tau}{4N} + \frac{\gamma[1-\gamma]\Delta c^2 N}{2\tau} + Y - E(c) - FN. \tag{5.13}$$

## 5.4 Main Analysis

Now I analyze the effect of competition in terms of transportation costs on firms' entry decisions. I first focus on homogeneous firms. Second, I allow for firm heterogeneity. I analyze the impact of firm heterogeneity on firm profits, entry, and social welfare.

### 5.4.1 Homogeneous firms

I now analyze how the strength of competition in the market affects the market structure in the case of homogeneous firms (i.e., $c_G = c_B = c$). The expected firm profits are (rewrite (5.10))

$$E(\Pi) = \begin{cases} \frac{\tau}{2N^2} & \text{if } \tau \leq [Y-c]N, \\ \frac{2[Y-c]}{N} - \frac{\tau}{2N^2} - \frac{[Y-c]^2}{\tau} & \text{if } \tau > [Y-c]N. \end{cases} \tag{5.14}$$

Lowering production costs has the following effect on entry.

**Proposition 5.1.** *Improving cost efficiency has no impact on profits and entry at low entry costs (i.e., $F < \hat{F}$) but augments profits and entry if entry costs are high (i.e., $F \geq \hat{F}$).*

Proposition 5.1 points to two regions of competition. With low entry costs, competition is high. Improving cost efficiency has no effect on firms' profitability because the benefits are passed on to consumers. With high entry costs, competition is low. Firms behave as monopolists for closely located consumers and from them they extract bigger rents due to lower costs. Consequently, improving cost efficiency augments firm profits and entry.

I can show the following relation between transportation costs and entry.

**Proposition 5.2.** *The level of entry is a bell-shaped function of transportation costs; that is, there exists a $\bar{\tau}$ such that for $\tau < \bar{\tau}$, $\frac{\partial N}{\partial \tau} > 0$ and for $\tau \geq \bar{\tau}$, $\frac{\partial N}{\partial \tau} \leq 0$.*

The intuition for this result is the following. With low transportation costs (i.e., at $\tau < \bar{\tau}$), competition is high. Further reduction of transportation costs deprives firms of their locational advantages. That is, consumers more easily travel to a competing firm. This lowers profits and hampers entry. However, for large transportation costs an additional effect is at work. Now each firm acts as a monopolist for its closest consumers. A firm must compensate its consumers for the transportation costs that they incur. Decreasing the level of transportation costs allows firms to augment rents obtained from these consumers. The second effect prevails for a sufficiently large interval in which firms behave as a monopolists; that is, for sufficiently large transportation costs. In this case, lowering transportation costs augments profits and entry. This result replicates the analysis by Salop (1979) in the model of perfect price discrimination.

One can compute the following comparative statics with respect to $\bar{\tau}$.

**Corollary 5.1.** *The region in which transportation costs are positively related to entry is decreasing in firms' entry costs (i.e., $\frac{\partial \bar{\tau}}{\partial F} < 0$) and in production costs (i.e., $\frac{\partial \bar{\tau}}{\partial c} < 0$).*

The intuition for the first part of Corollary 5.1 is as follows. Lowering entry costs increases the level of competition, which makes transportation costs more often positively related to entry (see Proposition 5.2). That is, lowering entry costs increases the region in which transportation costs and entry are positively related.

The second part of Corollary 5.1 stems from the following effect. Lower firm production costs augment competition because firms can compete for more distant consumers. However, Proposition 5.2 showed that higher competition makes transportation costs more often positively related to entry. This shows that lower production costs augment the region in which transportation costs and entry are positively related.

### 5.4.2 Heterogeneous firms

I now allow for heterogeneity in the production costs of firms (i.e., $c_B > c_G$). First, I analyze the impact of transportation costs on entry (and the related question of the connection between firm size and demand density). Second, I focus on the interconnection between heterogeneity and competition between firms. In particular, I ask whether more heterogeneity is beneficial for firms and for social welfare.

The following proposition shows that firm heterogeneity can change the relation between competition and entry as described in Proposition 5.2.

**Proposition 5.3.** *In the case of high heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$) and for the intermediate quality of the industry (i.e., $\gamma \in (\gamma_1, \gamma_2)$), the level of entry is a $U$-shaped function of transportation costs; that is, there exists a $\tilde{\tau}$ such that for $\tau < \tilde{\tau}$, $\frac{\partial N}{\partial \tau} < 0$ and for $\tau \geq \tilde{\tau}$, $\frac{\partial N}{\partial \tau} \geq 0$.*

The intuition for Proposition 5.3 is the following. For low values of transportation costs, increasing competition through a reduction of transportation costs allows low-cost firms to take more consumers from high-cost firms. This redistribution of market share from high-cost to low-cost firms augments the efficiency of the industry. Consequently, the expected firm profits increase and entry goes up.

To understand this further, we reexamine the conditions on $\gamma$ that are necessary for the positive relation between competition (in terms of lower transportation costs) and entry. A lower bound on $\gamma$ exists because high-cost firms only lose with higher competition. They lose both their customers and rents. However, an upper bound on $\gamma$ exists as well. Low-cost firms gain with competition only if they compete with high-cost firms. Hence, there must be sufficiently many high-cost firms in the industry.

For high transportation costs, the shift of the market share is small and higher competition only lowers rents without decisively improving the efficiency. Consequently, competition lowers profits and hampers entry. In total, the level of entry is a $U$-shaped function of transportation costs.

Interestingly, involving heterogeneity in the industry may reverse the result from homogeneous industry (compare Proposition 5.3 with Proposition 5.2). In particular, in homogeneous industry lowering transportation costs results in higher competition between firms, lower expected firm profits, and lower entry. However, for high transportation costs (i.e., for $\tau > \bar{\tau}$), a perverse equilibrium exists, in which lowering transportation costs actually increases firm (almost monopolistic) profits and leads to more entry. Proposition 5.3 shows that the standard result of Salop (1979) again reverses in heterogeneous industries. In particular, lowering transportation costs may again increase expected firm profit and lead to more entry. This, now happens for low transportation costs (i.e., for $\tau < \tilde{\tau}$).

Now I analyze how increasing heterogeneity affects firm profits, their entry decisions, and social welfare. I consider the change in the heterogeneity between firms as exogeneously driven.[7] The heterogeneity in production costs extends Proposition 5.1 to the following result.

[7]In Section 5.5.2, I apply the results to the banking industry, in which heterogeneity plays an important role because of stability considerations and may be affected in part exogenously by the bank regulator.

**Lemma 5.3.** *Increasing the heterogeneity $\Delta c$ at constant average production costs $E(c)$ augments profits at low entry costs (at low $F$) but does not change profits when entry costs are high (at high $F$). The positive effect is the greatest for industries of intermediate quality (i.e., $\gamma = \frac{1}{2}$).*

At low entry costs $F$, firms fiercely compete for the consumers. This transfers almost all the rents from firms to consumers. Firms can only keep the rents if they have a cost advantage over competitors. High heterogeneity means that low-cost firms can keep higher rents when in competition with high-cost firms. This raises expected profits. The positive effect is higher if low-cost firms often compete with high-cost firms, which is the case for intermediate quality of the industry (i.e., $\gamma = \frac{1}{2}$).

However, with high entry costs, competition is low and firms largely behave as local monopolists. In this case, changing the differences in costs does not matter as long as the average reservation price $E(c)$ stays constant.

**Lemma 5.4.** *Increasing the average production cost $E(c)$ has no effect on entry with low entry costs (at low $F$) but results in lower entry when entry costs are high (at high $F$).*

At high competition (at low $F$), firms earn very small profits. Profits are mostly transfered to consumers. Increasing production costs does not affect firm profits, but results in lower consumer profits. Hence, higher production costs do not affect firm profits nor the level of entry. If competition is low, however, firms take most of the profits. Increasing production costs now hampers the profits of the firms and leads to lower entry.

Combining Lemma 5.3 with Lemma 5.4, I can derive the following result.

**Proposition 5.4.** *If increasing heterogeneity $\Delta c$ comes with higher average production costs (higher $E(c)$), entry is augmented at low entry costs $F < \hat{F}$; however, entry is hampered when entry costs are high (i.e., $F \geq \hat{F}$). A parameter $\hat{F}$ is increasing in $\gamma$ (i.e., $\frac{\partial \hat{F}}{\partial \gamma} > 0$).*

At low entry costs, competition is fierce. This makes higher heterogeneity valuable because it results in a larger shift of the market share from high-cost to low-cost firms. The efficiency of the industry is increased. Higher production costs, on the other hand, have no effect on firm profits. To see this, observe that with low entry costs competition is fierce and firms mostly pass the profits to consumers. Therefore increasing the average production cost mostly hampers consumers and does not affect firm profits. Summarizing, firm profits and entry increase.

With high entry costs, competition is low and higher heterogeneity has a limited effect on firm profits. That is, low competition confines the positive role of heterogeneity essentially by limiting a shift of the market share from high-cost to low-cost firms. Furthermore, with low competition, firms keep most of their profits. Hence, increasing production costs negatively affects firm profits. Hence, firm profits and entry decrease.

Increasing the quality of the industry $\gamma$ augments the proportion of low-cost firms in the industry. Low-cost firms are fiercer competitors than high-cost firms if everything else is equal. That is, low-cost firms can reach more distant consumers, which makes transportation costs less important (see also Corollary 5.1). Higher quality $\gamma$ then increases competition. This makes higher heterogeneity more important compared to a decrease in production costs. Summarizing, increasing heterogeneity is more valuable if the industry is of high quality.

Proposition 5.4 brings policy implications for regulated industries. Costly regulatory policies that exacerbate the differences among firms in their production costs are effective in competitive and high quality industries. They augment firm profits and entry through redistribution of market share towards more efficient firms. Yet, such policies may not be valuable in non-competitive, low quality industries. If competition is low, the shift of the market share is low and the costs of such policies decrease profits and entry.

The following corollary is intuitive.

**Corollary 5.2.** *Improving the cost efficiency of low-cost firms always augments entry.*

Lower production costs of low-cost firms increase expected profits due to two reasons. First, firms may earn higher profits due to lower production costs. Second, improving the cost efficiency of low-cost firms increases the differences in production costs between low-cost and high-cost firms. This allows low-cost firms to earn extra profits when competing with high-cost firms. Consequently, expected profits and entry increase.

**Corollary 5.3.** *Improving the cost efficiency of high-cost firms hampers entry with low entry costs (i.e., $F < \hat{F}$), but augments entry when entry costs are high (i.e., $F \geq \hat{F}$). Also $\frac{\partial \hat{F}}{\partial \gamma} > 0$.*

This counterintuitive result can be explained as follows. With low entry costs, competition between firms is fierce. The benefits of a lower production cost of high-cost firms $c_B$ are passed on to consumers (see Proposition 5.1). However, lowering the production cost of high-cost firms also decreases the differences between low-cost and high-cost firms. Lower differences increase the competitiveness of the industry. That is, firms with equal production costs can extract lower rents from consumers. In addition, lower differences between firms limit the beneficial market shift from high-cost to low-cost firms. Hence, profits and entry decrease (see Lemma 5.3). This is different with high entry costs, where firms largely behave as local monopolists. Lowering the production costs of high-cost firms now results in higher firm profits. Hence, there is more entry.

Corollary 5.2 and Corollary 5.3 point to potential differences in directing policies towards low-cost vs. high-cost firms. Whereas policies that increase the productivity of low-cost firms always benefit the industry and lead to higher entry, policies that help high-cost firms may be detrimental to the industry as a whole, especially if competition is high and the industry is of high quality. Differently said, in this case policies that further hurt high-cost firms may improve firm profits and entry.

The impact of higher differences between firms on social welfare is the following.

**Proposition 5.5.** *Increasing heterogeneity $\Delta c$ always augments social welfare.*

Proposition 5.5 is driven by a welfare-positive effect of competition. That is, competition shifts the market share from high-cost to low-cost firms. Higher heterogeneity has two positive effects. First, the shift of the market share is higher. Second, the benefits of the shift of the market share are higher. That is, a shift of one consumer from a high-cost firm to a low-cost firm is more beneficial if the differences in costs are high. Both effects increase social welfare.

Summarizing, this section first connects entry in the industry with the level of competition in the industry; namely, with the level of transportation costs. I show that for sufficiently heterogeneous firms, lowering transportation costs leads to a beneficial shift of the market share that increases firm profits and entry. Second, I show that increasing the heterogeneity of the industry augments firm profits, entry in the industry, and social welfare.

## 5.5 Extensions and Applications

In this section I first extend the analysis and connect it with the literature on optimal firm size. Second, I apply the analysis to illuminate some pertinent issues in bank regulation. I start with deposit insurance and capital regulation and continue with the role of market transparency and bail out policies.

### 5.5.1 Extension: Firm size and demand density

This analysis can also be interpreted to predict the relation between demand density and the size of firms. I show that in a sufficiently competitive and heterogeneous industry firms in denser markets may be smaller. Higher demand density affects firm size through two different channels. First, higher density creates larger demand. Larger demand induces more entry which leads to fiercer competition and lower rents. To compensate for lower rents, firms must be larger (see Syverson (2004)). However, if competition and heterogeneity of industry is high, the second channel might prevail. In particular, higher density increases competition, which allows for a greater shift of the market share from high-cost to low-cost firms and enhances the cost efficiency of the industry. The gains in cost efficiency boost profits and entry which leads to smaller firms.

The model is now expanded such that a total mass of consumers on the unit circle is given by $D$. For homogeneous firms, I can show the following proposition.

**Corollary 5.4.** *Average firm size is a U-shaped function of market density; that is, there exists $\bar{D}$ such that for $D < \bar{D}$, $\frac{\partial[D/N]}{\partial D} < 0$ and for $D \geq \bar{D}$, $\frac{\partial[D/N]}{\partial D} \geq 0$.*

The intuition is the following. Increasing demand density augments entry due to two effects. First, higher demand density increases demand for products, which directly augments profits and entry. An indirect effect is at work as well. That is, a higher number of entering firms lowers the distances between firms and consumers. For low levels of demand density, this actually improves profits because firms that largely behave as monopolists spend less to compensate consumers for their transportation costs. Thus, the indirect effect creates additional entry. If the indirect effect is sufficiently strong (which happens for low demand density; that is, for $D < \bar{D}$), entry in the market is so high that the number of firms grows faster than the market itself and the firm size shrinks.

At high demand density, a direct effect is still present. That is, even higher density improves firm profits which results in more entry. However, an indirect effect works in the opposite direction. In particular, a higher number of entering firms strengthens competition. Firms more fiercely compete for consumers, which leads to lower profits and entry. Consequently, higher density increases firm size.

In a similar spirit, the impact of demand density on firm size is different in a heterogeneous industry (compare with Corollary 5.4).

**Corollary 5.5.** *In the case of heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$), and for the intermediate quality of the industry (e.i., $\gamma \in (\gamma_1, \gamma_2)$), the average firm size is a bell-shaped function of demand density; that is, there exists $\bar{D}$ such that for $D < \bar{D}$, $\frac{\partial[D/N]}{\partial D} > 0$ and for $D \geq \bar{D}$, $\frac{\partial[D/N]}{\partial D} \leq 0$.*

The reason for the bell-shaped function is the following. For high demand density the relationship between demand density and firm size may be negative. This is because increasing density brings new firms into the industry, which intensifies competition. In a sufficiently heterogeneous industry, intensifying competition augments firm profits. In particular, high competition shifts the market share from high-cost to low-cost firms, which leads to more efficient industry, bigger rents, and additional entry. In total, the number of firms goes up more than the market size increases. Consequently, firms become smaller.

For low demand density, the relation between demand density and firm size is positive. This is because higher density attracts additional entry, which augments competition. However, the positive effect of competition – namely, the redistribution of the market share from high-cost to low-cost firms – is limited. Competition then aggravates firm profits, which hampers entry. In sum, an increase in demand density makes firms larger.

In an empirical analysis of the cement industry, Syverson (2004) confirms a positive relation between demand density and firm size. Syverson (2004) may account for the relation between firm size and demand density that I predict for relatively low levels of demand density. Arguably, the cement industry is geographically rather segmented. Corollary 5.5 predicts that this relation may reverse for some less segmented industry.

Now I turn to the applications of the results.

### 5.5.2 Application: Bank regulation

Heterogeneity in banking plays an important role due to its stability implications. In particular, the bank regulator may contain the heterogeneity (in the qualities) of banks to limit potentially devastating bank failures.[8] The regulator may use several regulatory tools to

[8]See Bhattacharya, Boot, and Thakor (1998) for a further rationale for bank regulation.

impact the stability of the banking system. However, such policies may also exogenously affect heterogeneity and the level of competition in the banking industry. This makes this analysis especially applicable to bank regulation. The analysis highlights the side effects of stability-oriented bank regulation that come from its impact on heterogeneity in the banking industry.[9]

Evidence from branching deregulation of the U.S. banking system identifies how intra-industry competition measured by the transportation cost affects entry in banking. Before 1970, branching in the U.S., was effectively prohibited. Beginning in the late 1970s, the U.S. gradually removed the restrictions to within-state and inter-state branching. This has substantially lowered bank-borrower distances and lowered transportation costs. Proposition 5.3 predicts that lowering the transportation costs should increase entry if the shift in the market share from low- to high-quality banks is sufficiently high (i.e., the industry is sufficiently heterogeneous and competitive). Stiroh and Strahan (2003) show that deregulation led to a substantial reallocation of market share toward better banks. Aligned with Proposition 5.3, Berger et al. (2004) confirm that entry was higher in states with newly liberalized branching rules where shifts in market share became feasible.

Proposition 5.3 also identifies a potential rationale for deregulation. Banks on their own might pressure the regulator to lower barriers to branching and, in doing so, induce more interbank competition. This applies to better quality banks that as such would gain market share. That is, such deregulation would allow for a shift in the market share toward good banks. Even with higher competitive pressure, this could elevate the better bank profitability. Kroszner and Strahan (1999) provide some evidence to this point. They show that the private-interest and the strength of winners (large banks) versus losers (small banks) affected interstate branching deregulation through influence on congressional voting.

Deposit insurance is one of the most widely used forms of bank regulation (see Demirgüç-Kunt, Karacaovali, and Laeven (2005)). Deposit insurance guarantees depositors that their claims will be repaid even if their bank fails. This may contain bank runs. However, insured depositors disregard bank risk knowing that the deposit insurer carries potential losses. Hence, depositors are willing to invest in risky and in safe banks at exactly the same interest rates. This shows that deposit insurance helps risky banks obtain a cheap source of funds: insured deposits. That is, deposit insurance puts banks on an equal footing.

In terms of this analysis, deposit insurance lowers the differences in funding costs between banks. Proposition 5.5 then presents a rather negative view on deposit insurance. In particular, deposit insurance limits an efficient redistribution of market share from high-cost (high-risk) to low-cost (low-risk) banks. This may negatively affect social welfare. Further more, Proposition 5.4 predicts that more generous deposit insurance may lead to less entry in banking, especially if competition is high and if banks are on average safe.

This shows that regulators of highly competitive and high quality banking systems should

[9]This analysis does not serve to criticize bank regulation as such. The importance of bank regulation in containing costly bank failures is unquestionable. I merely point to some of its potentially unexpected effects.

carefully consider the effect of deposit insurance. Correct implementation seems to be crucial. Although theoretical studies (see Chan, Greenbaum, and Thakor (1992)) show that a deposit insurance premium can hardly be adjusted to a bank's risk, some regulators are nevertheless moving in this direction.[10] Alternatively, the coverage of deposit insurance may be limited (see Chen (1999) and Gropp and Vesala (2004)).

Proposition 5.4 could also predict that the banking industry as a whole would lobby for limits on deposit insurance in highly competitive, high-quality banking systems. Lower deposit insurance would induce a redistribution of market share to better banks. However, in less competitive banking markets, meaning those with little competition between existing banks, the redistribution effect is limited and banks may prefer to have more generous deposit insurance.

In Chapter 3 I show that capital regulation limits the distortion of deposit insurance. In particular, capital regulation forces banks to raise capital and therefore rely less on insured deposits. Hence, risky banks benefit less due to deposit insurance. Capital regulation limits the distortions that flat-rate deposit insurance introduces to banking. Implicitly, such deposit insurance benefits lower quality banks the most, and makes them fiercer competitors than they otherwise would have been. Capital requirements are an effective regulatory tool that mitigates this distortion, and in doing so increases the value of entry.

In Chapter 3 I show that costly capital regulation may lead to higher bank profits and more entry if competition is sufficiently high. This analysis rederives this result in the framework of a spatial model. In terms of this model, increasing capital requirements augments the average cost of financing for banks ($E(c)$ increases). However, higher capital requirements at the same time increase the cost differences between banks (an increase in $\Delta c$) by limiting the effect of deposit insurance. Proposition 5.4 then shows that higher costly capital requirements may result in more entry if competition is sufficiently high; that is, if transportation costs for bank clients are sufficiently low.

This analysis also offers some implications for market-based regulation. Bank supervisors increasingly try to learn from market information when it comes to analyzing bank stability (KMV-models, Basel II, etc.). Proposition 5.4 shows that market disclosure and market pressure policies, while highlight the differences between banks, are valuable for banks only in highly competitive banking industries. That is, market-based regulation can better discriminate between low-risk and high-risk (low-cost and high-cost) banks. This increases the level of heterogeneity in the industry, leading to a shift of the market share from high-risk to low risk banks. If competition is sufficiently high, bank profits and entry are augmented. That is, the banking industry as a whole may benefit from market-based regulation. However, if the level of competition is low, the shift of the market share is limited. Hence the additional costs of market based regulation may overcome its benefits.

[10]In the U.S., the Federal Deposit Insurance Corporation (FDIC) is designing deposit insurance to reflect the risks of the financial institution more closely. The FDIC computes the deposit insurance premium using capital levels, supervisory ratings, and financial ratios and issuer ratings. See FDIC Statistics on Banking, A Statistical Profile of the United States Banking Industry, March 2006.

Corollary 5.3 applied to regulatory bailout policies points to the potentially negative effect of bailout policies even if they come at no additional cost to the banking industry. Generous bailout policies essentially help risky (and high-cost) banks obtain cheaper financing. Stable banks benefit from bailout policies substantially less. This shows that bailout policies decrease the differences between banks. If competition is high, this hampers redistribution of market share to more stable (and potentially more efficient) banks, lowering profits and entry. That is, the banking industry as a whole may be hampered by too lax bailout policies. However, if competition is low, the redistribution of the market share is limited. In this case, costless bailout policies positively affect bank stability and increase bank profits and entry.

## 5.6 Conclusions

How does higher competition in terms of lower transportation costs affect firm entry in the industry? I counter the general intuition that higher inter-industry competition results in lower profits and less entry. I show that intensifying inter-industry competition could encourage entry if firms are heterogeneous. This result comes from the efficiency gain that competition creates. In particular, higher competition induces a redistribution of consumers from high-cost to low-cost firms, augmenting firm profits and entry. I show that entry is a bell-shaped function of transportation costs if firms are sufficiently heterogeneous. Yet, in a homogeneous industry entry is a $U$-shaped function of transportation costs.

I also point to the effect of differences in production costs between firms at the level of entry. I show that costly policies that increase differences in production costs may augment entry. In particular, the shift of the market share from high-cost to low-cost firms increases the efficiency of the industry. This augments expected firm profits and entry. Interestingly, enhancing cost differences by making high-cost firms even worse may augment average firm profits and entry if the level of competition is sufficiently high.

The banking industry seems to be ideally suited for application of these results. The banking industry is highly heterogeneous and the level of competition varies substantially across countries. In addition, regulatory policies aiming at bank stability exogenously affect the level of heterogeneity of banks. I show that deposit insurance may hamper entry while capital regulation limits its negative impact on entry. Even more, higher capital requirements may lead to more entry. Surprisingly, bailout policies may hamper bank profits and entry even though they come at no additional cost for banks. As a general conclusion, the regulator should be cautious when implementing the stability-oriented regulation, especially if the banking system is rather competitive and of high quality. In this case, the distortionary effects of regulatory policies may prevail and may be harmful for the banking industry.

This analysis presents a basic attempt to study how firms' entry decisions are affected by heterogeneity and competition in a spatial model. Both heterogeneity and competition in the industry (given by the level of transportation costs) were taken as exogenously given. In Chapter 3 I make firm heterogeneity endogenous. In particular, firms (i.e., banks) could invest and improve their quality. One could also make transportation costs endogenous. For example, firms could invest and lower transportation costs for their consumers. This extension still awaits exploration.

## 5.7 Appendix

*Proof of Lemma 5.1*

I distinguish three cases. First, for $\frac{\tau}{N} \leq Y - c_B$ the expected profit of an entering bank is

$$E(\Pi) = \gamma^2 \frac{\tau}{4N^2} + \gamma[1-\gamma][\frac{\{\tau + N[c_B - c_G]\}^2}{4\tau N^2} + \frac{\{\tau + N[c_G - c_B]\}^2}{4\tau N^2}] + [1-\gamma]^2 \frac{\tau}{4N^2}. \quad (5.15)$$

Simplify (5.15) to obtain the first line in (5.10).

Second, for $\frac{\tau}{N} > Y - c_G$, one obtains

$$\begin{aligned} E(\Pi) &= \gamma^2[-\frac{\tau}{4N^2} - \frac{[Y-c_G]^2}{\tau} + \frac{Y-c_G}{N}] + \gamma[1-\gamma]\{\frac{2Y-c_G-c_B}{2N} - \frac{\tau}{4N^2} \\ &\quad + \frac{[c_B-c_G]^2 - 2[Y-c_B]^2}{4\tau} + \frac{2Y-c_B-c_G}{2N} - \frac{\tau}{4N^2} + \frac{[c_G-c_B]^2 - 2[Y-c_G]^2}{4\tau}\} \\ &\quad + [1-\gamma]^2[-\frac{\tau}{4N^2} - \frac{[Y-c_B]^2}{\tau} + \frac{Y-c_B}{N}]. \end{aligned} \quad (5.16)$$

Simplify (5.16) to obtain the second line in (5.10).

Third, for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one computes

$$\begin{aligned} \Pi_{G,B} + \Pi_{B,G} &= \frac{[Y-c_B] + [Y-c_G]}{2N} - \frac{\tau}{4N^2} + \frac{[[Y-c_G]-[Y-c_B]]^2 - 2[Y-c_G]^2}{4\tau} + \\ &\quad \frac{\{\tau + N[[Y-c_G]-[Y-c_B]]\}^2}{4\tau N^2}. \end{aligned} \quad (5.17)$$

Simplify (5.17) to obtain the third line in (5.10). ■

*Proof of Lemma 5.2*

Observe from (5.10) that for $\frac{\tau}{N} \leq Y - c_B$ one has $\frac{\partial E(\Pi)}{\partial N} < 0$. For $\frac{\tau}{N} > Y - c_G$, one computes

$$\frac{\partial E(\Pi)}{\partial N} = -\frac{\gamma[Y-c_G] + [1-\gamma][Y-c_B]}{N^2} + \frac{\tau}{2N^3}.$$

Note that Assumption (5.2) guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

For $Y - c_B \leq \frac{\tau}{N} \leq Y - c_G$ one computes

$$\frac{\partial E(\Pi)}{\partial N} = -\frac{\gamma\tau}{2N^3} - \frac{1-\gamma}{N^2}[\frac{[Y-c_B]}{N} - \frac{\tau}{2}].$$

Note that Assumption (5.2) guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$. ■

*Proof of Proposition 5.1*

For $\frac{\tau}{N} \leq Y - c$ it is known that in equilibrium $\frac{\tau}{2N^2} = F$. That is, as long as $F < \hat{F}$, where $\hat{F} = \frac{[Y-c]^2}{2\tau}$, one has $\frac{\tau}{N} \leq Y - c$. In this case $\Pi(N) = \frac{\tau}{2N^2}$. Note that $\frac{\partial \Pi(N)}{\partial c} = 0$.

For $F \geq \hat{F}$, one has $\frac{\tau}{N} > Y - c$ and $\frac{\partial \Pi(N)}{\partial c} = -\frac{2}{N} + 2\frac{[Y-c]}{\tau}$, which is negative due to Assumption 5.2. ■

*Proof of Proposition 5.2*

Differentiate (5.14) with respect to $\tau$ to obtain

$$\frac{\partial \Pi(N)}{\partial \tau} = -\frac{1}{4N^2} + \frac{[Y-c]^2}{2\tau^2} \text{ for } \tau > [Y-c]N, \tag{5.18}$$

$$\frac{\partial \Pi(N)}{\partial \tau} = \frac{1}{4N^2} \text{ for } \tau \leq [Y-c]N. \tag{5.19}$$

Note that (5.19) is always positive.

Note that (5.18) is positive for $\frac{\tau}{N} \leq \sqrt{2}[Y-c]$ but is negative for $\frac{\tau}{N} > \sqrt{2}[Y-c]$. In equilibrium it is known that

$$\frac{2[Y-c]}{N} - \frac{\tau}{2N^2} - \frac{[Y-c]^2}{\tau} = F. \tag{5.20}$$

Now compute the entry costs $\bar{F}$ for which $\frac{\partial \Pi(N)}{\partial \tau} = 0$. Use $\tau = \sqrt{2}N[Y-c]$ in (5.20) to obtain

$$\bar{F} = \frac{[Y-c]^2}{\tau}[2\sqrt{2}-2]. \tag{5.21}$$

Note that Lemma 5.2 shows that the equilibrium $N$ is decreasing in $F$. Thus, for $F < \bar{F}$ one has $\frac{\partial \Pi(N)}{\partial \tau} > 0$ and for $F \geq \bar{F}$ one has $\frac{\partial \Pi(N)}{\partial \tau} < 0$. Solve (5.21) for $\tau$ to obtain $\bar{\tau}$. Thus, $\frac{\partial \Pi(N)}{\partial \tau} > 0$ for $\tau < \bar{\tau}$ and $\frac{\partial \Pi(N)}{\partial \tau} < 0$ for $\tau \geq \bar{\tau}$. ■

*Proof of Proposition 5.3*

The proof consists of two parts. First, I show that the equilibrium lies in the region $Y - c_B < \frac{\tau}{N} < Y - c_G$. Observe that for high heterogeneity $\Delta c > \sqrt{2}[Y - c_B]$ a milder condition also holds (i.e., $\Delta c > Y - c_B$). Combine this with Assumption 5.1 to see that the region $\tau \leq [Y - c_B]N$ does not exist. More specifically, combine high heterogeneity with $\frac{\tau}{N} \leq Y - c_B$ to obtain $c_B - c_G > Y - c_B > \frac{\tau}{N}$. However, this contradicts with Assumption 5.1. Note that for high heterogeneity ($Y - c_G > 2[Y - c_B]$) there is no equilibrium for $\frac{\tau}{N} > Y - c_G$. That is, rewrite $\Delta c > Y - c_B$ to obtain $Y - c_G > 2[Y - c_B]$, which means that $\frac{\tau}{N} > Y - c_G > 2[Y - c_B]$, but this contradicts with Assumption 5.2.

Second, I show that the proof holds in the region $Y - c_B < \frac{\tau}{N} < Y - c_G$. I introduce $d = \frac{1}{N[Y-c_B]}$ and $f = \frac{F}{[Y-c_B]^2}$ and $g = \frac{\Delta c^2}{[Y-c_B]^2}$. Note that $1 < \tau d < 2$. That is, Assumption 5.2 guarantees that $\tau d < 2$ and the condition $Y - c_B < \frac{\tau}{N}$ guarantees that $1 < \tau d$. Observe also that, for high heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$), one has $g > 2$.

In what follows I prove that there exists $\tilde{F} = \tilde{f}[Y - c_B]^2$ such that for $f < \tilde{f}$ one has $\frac{\partial d}{\partial \tau} > 0$ and for $f \geq \bar{f}$ one has $\frac{\partial d}{\partial \tau} < 0$ as long as $g > [\sqrt{2} + 1]^2$ and $\gamma \in (\gamma_1, \gamma_2)$, where

$$\gamma_1 = 1 - \frac{1}{g}, \quad \gamma_2 = \frac{1}{2} + \frac{\sqrt{1 - 2/g}}{2}. \tag{5.22}$$

Rewrite (5.10) for the region $Y - c_B < \frac{\tau}{N} < Y - c_G$ to obtain the equilibrium condition

$$\gamma\frac{\tau d^2}{2} + [1-\gamma][d - \frac{\tau d^2}{2} + \frac{\gamma g - 1}{\tau}] = f. \quad (5.23)$$

Differentiate (5.23) w.r.t. $\tau$ to obtain

$$\frac{\partial d}{\partial \tau} = \frac{\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]}{\tau^2\{\gamma\tau d + [1-\gamma][2-\tau d]\}}. \quad (5.24)$$

Note that $\gamma\tau d + [1-\gamma][2-\tau d] > 0$ because $\tau d < 2$. Observe that $d$ is increasing in $f$. In order to have $\frac{\partial d}{\partial \tau} > 0$ for low $f$ and $\frac{\partial d}{\partial \tau} < 0$ for high $f$, one needs $\frac{\partial d}{\partial \tau}$ to be a negative function of $d$. One needs $\gamma > \frac{1}{2}$, but this is true because $\gamma > \gamma_1$.

Observe that $\frac{\partial d}{\partial \tau} > 0$ for the lowest $\tau d$ (i.e., $\tau d = 1$). That is,

$$\left[\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]\right]_{d\tau=1} = -\frac{1}{2} + \gamma[1-\gamma]g, \quad (5.25)$$

which is positive for at least $\gamma \in (\frac{1}{2} - \frac{\sqrt{1-2/g}}{2}, \frac{1}{2} + \frac{\sqrt{1-2/g}}{2})$, and therefore also for $\gamma \in (\gamma_1, \gamma_2)$, because $g \geq 2$.

Now observe that $\frac{\partial d}{\partial \tau} \leq 0$ for the highest $\tau d$ (i.e., $\tau d = 2$). Compute

$$\left[\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]\right]_{d\tau=2} = \gamma\{-1 + [1-\gamma]g\}, \quad (5.26)$$

which is negative if $\gamma > \gamma_1$. Thus, there exists $\bar{f}$ corresponding to a certain $\tau\tilde{d}$ such that for $f < \tilde{f}$ one has $d < \tilde{d}$ and $\frac{\partial d}{\partial \tau} > 0$ and for $f \geq \tilde{f}$ one has $d \geq \tilde{d}$ and $\frac{\partial d}{\partial \tau} \leq 0$. ■

*Proof of Lemma 5.3*

Insert $c_G = E(c) - [1-\gamma]\Delta c$ and $c_B = E(c) + \gamma\Delta c$ into (5.10). For $Y - c_B < \frac{\tau}{N} \leq Y - c_G$, one obtains

$$E(\Pi) = \gamma\frac{\tau}{2N^2} + [1-\gamma]\{2\frac{Y - E(c) - \gamma\Delta c}{N} - \frac{\tau}{2N^2} + \frac{\gamma\Delta c^2 - [Y - E(c) - \gamma\Delta c]^2}{\tau}\} - F. \quad (5.27)$$

Now differentiate profits in (5.10) (use also (5.27)) with respect to $\Delta c$ to obtain

$$\left.\frac{\partial E(\Pi)}{\partial \Delta c}\right|_{E(c)=const.} = \begin{cases} \frac{2\gamma[1-\gamma]\Delta c}{\tau} & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ 0 & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2\gamma[1-\gamma][\frac{[Y-c_G]}{\tau} - \frac{1}{N}] & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \quad (5.28)$$

Note that $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} > 0$ for $F$ sufficiently low such that $N > \frac{\tau}{[Y-c_G]}$. For sufficiently high $F$ one has $N < \frac{\tau}{[Y-c_G]}$ in which case $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} = 0$. Note also that for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one has $\frac{\partial E(\Pi)}{\partial \Delta c} > 0$. Observe also that (5.28) is the greatest for $\gamma = \frac{1}{2}$. ■

*Proof of Lemma 5.4*

Differentiate (5.10) with respect to $E(c)$ keeping $\Delta c$ constant to obtain

$$\frac{\partial E(\Pi)}{\partial E(c)}\bigg|_{\Delta c=const.} = \begin{cases} 0 & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ -\frac{2}{N} + \frac{2[Y-E(c)]}{\tau} & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2[1-\gamma][-\frac{1}{N} + \frac{[Y-c_B]}{\tau}] & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \tag{5.29}$$

Note that $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} = 0$ for $F$ sufficiently low such that $N > \frac{\tau}{[Y-c_B]}$. For sufficiently high $F$, we have $N < \frac{\tau}{[Y-c_B]}$ in which case $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} < 0$. To see this, note that $\frac{2}{N} - \frac{2[Y-E(c)]}{\tau} < 0$ because $\frac{\tau}{N} > Y - c_G$. Also observe $2[1-\gamma][\frac{1}{N} - \frac{[Y-c_B]}{\tau}] < 0$ because $Y - c_B < \frac{\tau}{N}$.

Note that $2[1-\gamma][\frac{1}{N} - \frac{Y-c_B}{\tau}]$ is maximized for $\gamma = 0$. ■

*Proof of Proposition 5.4*

We check for the sign of $DF$, where we define

$$DF = \frac{\partial E(\Pi)}{\partial \Delta c}\bigg|_{E(c)=const.} + \alpha\, \frac{\partial E(\Pi)}{\partial E(c)}\bigg|_{\Delta c=const.}. \tag{5.30}$$

Use (5.10) to compute (5.30) for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ to obtain

$$DF = 2\gamma[1-\gamma][\frac{Y-c_G}{\tau} - \frac{\gamma}{N}] - 2\alpha[1-\gamma][\frac{1}{N} - \frac{Y-c_B}{\tau}]. \tag{5.31}$$

Use (5.10), (5.28) and (5.31) to obtain

$$DF = \begin{cases} \frac{2\gamma[1-\gamma]\Delta c}{\tau} & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ -\alpha[\frac{2}{N} + 2\frac{Y-E(c)}{\tau}] & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2[1-\gamma]\{\gamma\frac{[Y-c_G]}{\tau} - \alpha[\frac{1}{N} - \frac{[Y-c_B]}{\tau}]\} & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \tag{5.32}$$

Note that for sufficiently low $F$ this yields $\frac{\tau}{N} \leq Y - c_B$ one has (5.30) always positive. For medium $F$ such that $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ there exists

$$\bar{\gamma} = \frac{\alpha\{\frac{\tau}{N} - [Y - c_B]\}}{[Y - c_G]},$$

such that for $\gamma > \bar{\gamma}$ the value in (5.30) is positive and for $\gamma \leq \bar{\gamma}$ the value in (5.30) is negative. For very high $F$ this yields$\frac{\tau}{N} > Y - c_G$, the value in (5.30) is negative for all $\gamma$. ■

*Proof of Corollary 5.2*

In the equilibrium $E(\Pi(N^*)) = F$. Differentiate it with respect to $N^*$ to obtain

$$\frac{\partial E(\Pi)}{\partial c_G}\frac{\partial c_G}{\partial N} + \frac{\partial E(\Pi)}{\partial N} = 0.$$

Compute

$$\frac{\partial N}{\partial c_G} = -\frac{\frac{\partial E(\Pi)}{\partial c_G}}{\frac{\partial E(\Pi)}{\partial N}}.$$

Note that Lemma 5.2 guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

Now observe that for $\tau > [Y - c_G]N$ one has

$$\frac{\partial E(\Pi)}{\partial c_G} = -\gamma[\frac{1}{N} - \frac{[Y - E(c)]}{\tau}] < 0.$$

For $[Y - c_G]N > \tau > [Y - c_B]N$ one has

$$\frac{\partial E(\Pi)}{\partial c_G} = -\gamma[1 - \gamma]\frac{c_B - c_G}{\tau} < 0.$$

For $\tau < [Y - c_B]N$ one also has $\frac{\partial E(\Pi)}{\partial c_G} < 0$. Thus, $\frac{\partial N}{\partial c_G}$ is always negative. ■

*Proof of Corollary 5.3*

In the equilibrium $E(\Pi(N^*)) = F$. Differentiate it with respect to $N^*$ to obtain

$$\frac{\partial E(\Pi)}{\partial c_B}\frac{\partial c_B}{\partial N} + \frac{\partial E(\Pi)}{\partial N} = 0.$$

Compute

$$\frac{\partial N}{\partial c_B} = -\frac{\frac{\partial E(\Pi)}{\partial c_B}}{\frac{\partial E(\Pi)}{\partial N}}.$$

Note that Lemma 5.2 guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

Now observe that for $\frac{\tau}{N} \leq Y - c_B$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = [1 - \gamma][-\frac{1}{N} + \frac{Y - E(c)}{\tau}],$$

which is positive.

Note that for $\frac{\tau}{N} > Y - c_G$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = \frac{\partial E(\Pi)}{\partial E(c)}\frac{\partial [E(c)}{\partial c_B}.$$

Observe that $\frac{\partial E(\Pi)}{\partial E(c)} = 2[-\frac{1}{N} + \frac{[Y - E(c)]}{\tau}]$, which is negative due to $\frac{\tau}{N} > Y - c_G > Y - E(c)$.

Now observe that for $Y - c_G > \frac{\tau}{N} > Y - c_B$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = 2[1 - \gamma]\{-\frac{1}{N} + \frac{\gamma[c_B - c_G] + [Y - c_B]}{\tau}\}.$$

That is, there exists an equilibrium $N$ where $\frac{\partial E(\Pi)}{\partial c_B} = 0$ and is defined by

$$\tilde{N} = \frac{\tau}{\gamma[c_B - c_G] + [Y - c_B]} = \frac{\tau}{Y - E(c)}. \tag{5.33}$$

This shows that there exists $\hat{F}$ such that for $F < \hat{F}$ one has $N > \hat{N}$ and $\frac{\partial E(\Pi)}{\partial c_B} < 0$, whereas for $F \geq \hat{F}$ one has $N \leq \hat{N}$ and $\frac{\partial E(\Pi)}{\partial c_B} \geq 0$. In particular, insert (5.33) into (5.10) to obtain

$$\hat{F} = \frac{[Y - E(c)]^2}{2\tau} + \frac{\gamma[1-\gamma]\Delta c^2}{\tau}. \tag{5.34}$$

Differentiate $\hat{F}$ to $\gamma$ to obtain

$$\frac{\partial \hat{F}}{\partial \gamma} = \frac{Y - c_G}{\tau} + \frac{\Delta c^2}{\tau}\{1 - \gamma + [1-\gamma]^2 + 2\gamma^2. \tag{5.35}$$

Note that (5.35) is positive. ∎

*Proof of Proposition 5.5*

Differentiate (5.13) with respect to $\Delta c$ to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]N\Delta c}{\tau} + [\frac{\tau}{4N} - F + \frac{\gamma[1-\gamma]\Delta c^2}{2\tau}]\frac{\partial N}{\partial \Delta c} \tag{5.36}$$

For low entry costs, such that $\tau \leq [Y - c_B]N$, differentiate the equilibrium condition in (5.11) to obtain

$$\frac{\partial N}{\partial \Delta c} = \frac{2\gamma[1-\gamma]\Delta c N^3}{\tau^2}. \tag{5.37}$$

Insert (5.37) into (5.36) and use (5.11) to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]N\Delta c}{\tau}[\frac{1}{2} - \frac{\gamma[1-\gamma]\Delta c^2 N^2}{\tau^2}]. \tag{5.38}$$

Note that Assumption 5.1 guarantees that (5.38) is always positive.

For $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one can see that

$$\frac{\partial N}{\partial \Delta c} = N\frac{2\gamma[1-\gamma][\frac{[Y-c_G]}{\tau} - \frac{1}{N}]N}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y-c_B] - \frac{\tau}{N}]} < N\frac{2\gamma[1-\gamma]\frac{\Delta c}{\tau}N}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y-c_B] - \frac{\tau}{N}]}. \tag{5.39}$$

Note also that

$$N[\frac{\tau}{2N^2} - F + \frac{\gamma[1-\gamma]\Delta c^2}{2\tau}] = \frac{\tau}{4N}[3 - 4\gamma] - [Y - c_B][1-\gamma][2 - \frac{[Y-c_B]N}{\tau}] - \frac{\gamma[1-\gamma]\Delta c^2 N}{\tau}. \tag{5.40}$$

Insert (5.39) and (5.40) into (5.36) to obtain

$$\frac{\partial TW}{\partial \Delta c} > \gamma[1-\gamma]N\frac{\Delta c}{\tau}\frac{\gamma\frac{\tau}{N} - \frac{\tau}{4N} + [1-\gamma]\frac{[Y-c_B]^2 N}{\tau} - \frac{\gamma[1-\gamma]\Delta c^2 N}{\tau}}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y-c_B] - \frac{\tau}{N}]}. \tag{5.41}$$

Note that Assumption 5.1 guarantees that (5.41) is always positive.

Observe that for $\frac{\tau}{N} > Y - c_G$ the equilibrium number of firms as determined with (5.11) does not change with $\Delta c$. Consequently, differentiate social welfare in (5.13) to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]\Delta c N}{\tau},$$

which is always positive. ■

# Chapter 6

# Bank Foreign Entry and the Role of Bank Capital


**Abstract**

This chapter asks how capital regulation affects foreign entry. I first show that tighter capital regulation may lead to more foreign entry, which is reminiscent of the results in Chapter 3. I show that especially bad domestic banks may have incentives to merge to prevent entry of foreign banks. Interestingly, increasing capital requirements augments the merger incentives of banks. I then allow for discriminatory policies against foreign bank entry. In particular, I allow for higher capital requirements imposed only on foreign banks.



Keywords: Foreign Entry, Capital Requirements, Expropriation
JEL CLASSIFICATION: G21, G28

## 6.1 Introduction

This chapter is a direct extension of Chapter 3. I will analyze how foreign market penetration comes about in banking. In particular I will analyze what makes banks competitive in foreign markets, and how their success in cross-border banking is affected by capital regulation, potential expropriation by domestic regulators and banks' own strategic choices.

I analyze these issues in an industrial organization framework of Chapter 3. I distinguish multiple banks and let banks in different countries differ in quality. These quality differences are linked to the banks' abilities in monitoring borrowers and affect the profitability and riskiness of their lending operations. The quality of the banks in a country determines the strength of its banking system. I let banks compete for borrowers and analyze how their choices of monitoring technology, and hence risk, are affected by entry of foreign banks, capital regulation, and the regulatory actions.

I will focus on three extensions of the analysis of bank competition as put forward in Chapter 3. In the first part, I allow for one-sided competition. What I mean by this is that one country opens up its banking system to banks from another country, but this other country keeps its own market closed. I also analyze a de novo bank, without current borrowers, that seeks to enter an established banking market. The de novo bank has no incumbency advantage, but all existing competitors have. This extension highlights the problems that a start-up bank faces.

In this part I show that opening up of a strong banking sector does not affect stability of banks and their investments in monitoring technology. However, opening up of a weak banking sector weakens domestic institutions. They may loose market share and invest less in monitoring technology. Increasing capital requirements further helps high quality foreign banks to enter. That is, higher capital requirements limit the distortionary benefits that deposit insurance brings for weak domestic banks and help strong foreign banks. Weak domestic banks may defend by merging. Merger increases their investment in monitoring technology and prevent an influx of high quality foreign banks.

I also show that higher capital requirements should positively affect merger incentives. That is, when having more capital at stake, banks aim for stability and value monitoring technology more. But precisely merger increases investment in monitoring. Hence, at higher capital requirements banks would merge more likely.

In the second part, I highlight the drawbacks of potential expropriation by domestic regulators and their effect on the behavior of entering banks. I analyze the situation in which the domestic regulator limits the size of entering banks and thereby gives an unfair competitive advantage to local players. Alternatively, the regulator can put entering banks at a disadvantage by imposing tighter capital requirements on them. In the key result I show that increasing capital requirements only for foreign banks may put them at competitive disadvantage. They may loose market share and become riskier. In this case, imposing higher capital requirements on banks increases their risk. This averse effect of capital requirements is especially strong if competition is high.

In the third part, I enrich the set-up by introducing a richer specification of the true added value of banks. In particular, I allow for a more dynamic setting in which banks' staying power becomes important. What I mean by this is that borrowers in dealing with banks might find it important that banks are around in the future as well. This encompasses the notion that the bank-borrower relationship includes mutual long term investments that make staying power important (otherwise these investments cannot be recouped).

This chapter is organized as follows. In Section 6.2, I analyze entry of a foreign banks. In Section 6.3, I allow for expropriation of foreign banks by the domestic regulation. Section 6.4 stresses the importance of staying power. Section 6.5 concludes.

## 6.2 Foreign Entry

In this section, I analyze two modes of entry. First, banks use cross-border lending to enter the foreign banking system. That is, banks lend to the borrowers abroad using their already established home based monitoring technology. I analyze this mode of entry in a simplified framework allowing for one sided competition in which only one country opens up borders for foreign banks but not the other country. Second, I analyze de novo entry in which the bank has to build monitoring technology in the new market from scratch.

I use the same model specification as given in Chapter 3, Section 3.2. I also assume that Assumption 3.1 in Chapter 3 holds. Assumption 3.1 in Chapter 3 allows us to focus on competition between banks of different types. That is, the competing bank will not try to grab borrowers from banks of equal type. However, the competing bank will try to win market share from banks of lower quality.

### 6.2.1 One sided competition

The analysis in Chapter 3 has focused on symmetric competition. All banks are at equal footing, and the expected gain in market share (stealing borrowers from other banks) equals in expected value sense the potential loss they face in their own market, see the discussion following Lemma 3.1 in Chapter 3. I are now going to focus on one-sided competition. I consider two countries, where the first opens its domestic banking market, but the second country keeps its market closed. In this setting, I analyze how the symmetric competition results from Chapter 3 are affected. I seek to answer the question whether countries should single-handedly free up their banking markets or whether this should only be done on a reciprocal basis.

I proceed as follows. The country that opens up its market I call the 'open' country (country $O$). The country that keeps its banking market closed but whose bank is allowed to enter country $O$ I call the "attacking" country $A$. This means that a bank from country $A$ can enter country $O$, but not vice versa. Since I want to analyze later whether domestic mergers are an effective response to the threat of competition from foreign banks, I assume that there are two domestic banks in country $O$, but just one in country $A$. To make matters interesting, the bank in country $A$ is good.[1] I let all banks be of equal size. I distinguish two cases. In Case 1, both banks in country $O$ are bad; in Case 2, the banks are good.

**Proposition 6.1.**

**Case 1 – The domestic banks in country $O$ are good:** *The banks in country O hold on to their market share and do not change their investments in monitoring technology, but their values decrease because of extra competition from opening up the market. For the bank in country A nothing changes.*

**Case 2 – The domestic banks in country $O$ are bad:** *The banks in country O lose market share and value; the good bank from country A now gains market share and value. Anticipating the loss of market shares, the banks in country O reduce their investments in monitoring technology while the bank in country A increases its investment.*

The results in this proposition are quite straightforward. When the domestic banks in the country that opens up are good (Case 1) they can hold on to their market, and also their investments in monitoring technology remain intact. If the banks are bad (Case 2), they will lose out to the foreign competitor and market share is lost. Anticipating this, they will reduce their levels of investment in monitoring.

I analyze next what impact capital requirements have on the results in Proposition 6.1. I focus on the effect that capital requirements have on the profitability of entering country $O$ when that country's banks are bad (Case 2 in Proposition 6.1).

**Corollary 6.1.** *The attractiveness of entering country O when the banks in that country are bad is increasing in the level of the capital requirements.*

This corollary contrasts to the results in Proposition 3.4 in Chapter 3. There I showed that higher capital requirements encourage de novo entry only when $q$ is high enough and $\gamma$ takes on interior values. This corollary shows that an existing bank finds it *always* more profitable to enter a new market when capital requirements are higher. Corollary 6.1 provides the interesting empirical implication that in countries with relatively weak banks increasing capital requirements facilitates entry of foreign banks. Whether higher capital requirements also encourage de novo entry depends crucially on the openness of the banking markets; that is, on the parameter $q$ (see Proposition 3.4 in Chapter 3). Only when this parameter is sufficiently high, can more de novo entry be expected. Thus raising capital requirements can have different effects on *de novo* entry versus entry coming from existing banks.[2]

[1] If the bank in country $A$ is bad, it could not succeed in grabbing market share in country $O$.

[2] The careful reader may counter that entry in Proposition 3.4 in Chapter 3 was analyzed from the perspective of a de novo bank that did not yet know its type, while in Corollary 6.1 the existing bank knows that it is good. Observe however that if the de novo bank in Proposition 3.4 in Chapter 3 knows that it is good for sure, it would always be more likely to enter in response to higher capital requirements when $q$ is high (see also Proposition 3.3, Chapter 3). In Corollary 6.1 I do not need this restriction. A potentially more important consideration is that I implicitly assume that the fixed-cost based monitoring technology of any bank is equally useful across borders. If this technology is country specific, I are in a situation of late entry in which the entering bank needs to build up a new (second) monitoring capacity. This increases the incumbency advantage of the existing incumbent banks and complicates entry. This situation is analyzed in Section 6.2.2.

I show next that domestic banks may choose to merge to protect their market against foreign competition. The following corollary establishes that – given the fixed costs in the monitoring technology – merging indeed helps protect market share.

**Corollary 6.2.** *A merger between (bad) domestic banks helps defend them against the threat of foreign entry if the incumbency advantage exceeds some minimum level; that is,*

$$S > [\underline{\nu}_G - \underline{\nu}_B]X - \frac{X^2}{c\rho N}, \tag{6.1}$$

*and if competition is not too high (i.e., $q < \hat{q}(S)$), where $\frac{\partial \hat{q}}{\partial S} > 0$.*

Corollary 6.2 reflects the scale advantage that comes from merging. That is, the merged bank is prepared to make a bigger (ex ante) investment in monitoring technology which elevates its added value in lending. This helps the merged bad bank mitigate its quality disadvantage. The restriction on $q$ follows because the lower $q$, the more difficult for an entrant to grab market share. Similarly, the condition on $S$ (see (6.1)) guarantees a minimum incumbency advantage to deter the entrant.[3]

Next, I ask the question whether opening up borders encourages domestic mergers. And if so, are merging incentives elevated more for good than for bad domestic banks? I can prove the following.[4]

**Corollary 6.3.** *For any small positive entry cost to the foreign entrant the threat of entry (weakly) increases the value of merging for bad domestic banks, but has no effect on the merger incentives of high quality domestic banks.*

This corollary shows that the threat of entry encourages weak domestic institutions to merge. Such merger discourages entry because of scale economies. Without the threat of entry domestic banks enjoy a relatively high valuation as stand-alone entities. Particularly for weak institutions this value is at risk when entry comes about. Merging then may help.

I analyze next what impact capital requirements have on incentives of banks to merge. I can show the following.

**Corollary 6.4.** *Increasing capital requirements augments banks' merger incentives.*

This corollary stems from the following effect. A merger gives the scale advantage to the merged bank and elevates its investment in monitoring technology. Higher capital requirements make such increase in monitoring technology more valuable. Hence, higher capital requirements induce banks to merge more eagerly.

[3]The condition (6.1) poses a potentially stricter lower bound on $S$ than Assumption 3.1 in Chapter 3 does.

[4]To simplify the analysis I assume that the foreign bank faces a small positive entry cost. Note that without entry cost even if the foreign entrant would not succeed in obtaining market share upon entry, it would affect the valuations of good domestic banks (see Proposition 6.1). The latter effect is not present when there is an entry cost.

Several policy implications readily follow from my analysis. What my results in this section show is that opening up a weak domestic banking sector to foreign competitors weakens the domestic institutions (i.e., they lose market share), reduce investments in monitoring and hence become riskier. Increasing capital requirements makes entry even more likely, and using similar arguments as in Corollary 3.3, could undermine monitoring incentives further. The liberalization then does not help improve the quality of the domestic banking sector. While allowing domestic institutions to merge helps them protect market share and possibly favorably affects their monitoring incentives, it simultaneously could prevent the influx of higher quality banks. This would suggest that opening such domestic market should allow for takeovers of weak domestic institutions by foreign entrants.

Lastly, increasing capital requirements might induce more foreign entry and more mergers. Interestingly, that could imply that tightening of capital requirements (Flannery and Rangan (2004) show that banks have increased their capital partially due to higher regulatory requirements) may be additional cause for the merger wave in the banking system.[5]

### 6.2.2 Asymmetric competition with late entrants

So far, all banks were (initially) allocated the same number of borrowers $1/N$. Now I extend the model to incorporate the possibility that banks may enter late and have no initial (incumbent) borrowers; their established competitors, however, do. This would correspond to foreign entry, in which the foreign bank must invest in a new country from scratch.

Banks are again either good or bad. Assume that all established banks consider the number of banks $N$ to be fixed. Thus, late entry is not anticipated and, hence, the existing banks have chosen the levels of monitoring as given in Proposition 3.1 in Chapter 3. Starting from this equilibrium, a de novo bank may consider late entry. However, this bank misses an incumbency advantage. It can only obtain borrowers by luring them away from existing banks. As before, a late entering bank does not know its own type, yet knows the cross-sectional distribution of being a good or bad type respectively $\{\gamma, [1-\gamma]\}$.

I can now analyze which factors affect the profitability of late entry.

**Proposition 6.2.** *Late entry occurs if $q$ is sufficiently high, the existing banking market is of intermediate quality (i.e., $\gamma \in [\underline{\gamma}(q), \bar{\gamma}(q)]$), and the incumbency advantage is sufficiently small; that is,*

$$S < [\underline{\nu}_G - \underline{\nu}_B]X - \frac{X^2}{c\rho N}. \tag{6.2}$$

Observe that restriction (6.2) puts a stricter upper bound on $S$ than Assumption 3.1 does.[6] This is intuitive. With late entry, the new entrant is at a distinct competitive disadvantage because it has no incumbent borrowers. Consequently, a substantial scale

[5] Berger, Demsetz, and Strahan (1999) review causes for consolidation of banking industry. The literature mostly focuses on cost efficiency and market power rationale for bank consolidation.

[6] In the proof of Proposition 6.2 in Chapter 3, I show that there is a non-empty set of parameter values for which late entry can occur. Interestingly, note that the conditions in Corollary 6.2 and Proposition 6.2 are the mirror image of each other, and hence can never be jointly satisfied. Because Corollary 6.2 refers to

advantage needs to be overcome. Hence, the incumbency advantage $S$ should be small. The other conditions in the proposition mimic those in the earlier results. That is, the banking market needs to be sufficiently open ($q$ high) such that the entering bank can obtain access to borrowers. The restrictions on $\gamma$ guarantee that the entering bank has a sufficiently favorable image about its own quality. This explains the lower bound $\underline{\gamma}(q)$; only then can it expect to take a market share. The upper bound $\bar{\gamma}(q)$ guarantees that (in expectation) it can expect to encounter some weaker banks to take market share from.

In the spirit of Proposition 3.4 in Chapter 3, I can also show that higher capital requirements make late entry more profitable whenever the conditions in Proposition 6.2 in Chapter 3 are satisfied. Thus

**Corollary 6.5.** *In a region in which late entry is profitable (i.e., condition* (6.2) *holds, $q$ is sufficiently high, and $\gamma \in [\underline{\gamma}(q), \bar{\gamma}(q)]$), higher capital requirements always enhance the profitability of late entry, and hence induce more late entry.*

Corollary 6.5 is similar to Proposition 3.4 in Chapter 3, yet less restrictive. That is, strengthening capital regulation always helps encourage late entry whenever late entry is profitable.

In Section 6.2.1 I analyze cross-border competition, in which foreign banks could use their own technology when entering a new market. In this section I have assumed de novo entry in which a foreign bank cannot use the monitoring technology from its home market when operating abroad. In particular, the foreign bank must invest in monitoring in a new market from scratch and cannot rely on the scale advantage from its home market. One could argue that the magnitude of the differences between home and foreign lending markets defines the mode of entry. Banks can more successfully apply the already-installed lending technology to similar cultural, regulatory, and legal environment; see Degryse and Ongena (2004). In this case, the analysis in Section 6.2.1 would apply. Yet, if the difference between lending markets is substantial, banks must invest in the new market from scratch and the analysis of this section applies.

### 6.2.3 Different regulation across countries

So far I have assumed that capital requirements are equal across countries. Now I allow the regulators to choose the optimal levels of capital for their own countries. I can show the following result.

**Lemma 6.1.** *The welfare optimizing regulator in a weak banking system chooses lower capital requirements than the regulator in a strong banking system.*

The intuition for Lemma 6.1 is the following. Capital regulation has a limited effect on bad banks (see Proposition 3.2 in Chapter 3). This is why the regulator of a weak banking

foreign entry of existing banks and Proposition 6.2 to de novo entry, they are not fully comparable. However, these results do show that, in circumstances in which merging is effective in deterring foreign entry, late entry is also not feasible (even without merging).

system is reluctant to set overly high capital requirements. High capital requirements result in additional cost to banks but do not change their risk much. This contrasts with the situation in a strong banking system. Capital regulation greately affects good banks; hence the regulator chooses high capital requirements to mitigate their risk. Summarizing, Lemma 6.1 shows that the regulators would optimize social welfare by selecting different capital requirements in banking systems of different strengths.

Now I analyze whether opening up banking systems harmonizes regulation of those banking systems. I compare optimal regulation in two different cases. First, banks enter through cross-border lending; that is, they extend loans abroad using their home-based monitoring technology (see Section 6.2.1). Second, banks enter through de novo entry; that is, they build a new, country-specific monitoring technology (see Section 6.2.2).

**Proposition 6.3.** *In case of cross-border lending, capital regulation becomes less harmonized across countries if banking systems open up their borders. In the case of de novo entry, capital regulation becomes more harmonized if banking systems open up.*

For cross-border lending, the following situation applies. If a weak country opens up its borders, domestic low-quality banks must compete with strong foreign competitors. Losing market share makes them even weaker and less responsive to capital regulation. Consequently, the regulator of a weak banking system is inclined to loosen capital regulation even further. The opposite holds for the regulator of a strong banking system. Opening banking systems up allows strong banks to take additional market share from weak banks. They grow larger and become more responsive to capital regulation. Correspondingly, the regulator of a strong banking system further increases the level of capital regulation.

This is different with de novo entry if entering foreign banks are regulated by the domestic regulator. The regulator of a weak banking system now regulates not only weak domestic banks but also strong de novo entrants. Due to higher average quality, banks become more responsive to capital regulation. Hence, the regulator of a weak domestic banking system increases capital requirements and capital regulation across countries becomes more harmonized.

Proposition 6.3 shows that opening up banking systems for cross-border competition may lead to less harmonization between the regulatory standards in countries with banks of different qualities. In contrast, opening up banking systems for de novo entry in which the foreign bank is regulated by the domestic regulator results in greater harmonization of capital regulation.

Proposition 6.3 triggers some thoughts regarding initiatives on harmonization of capital regulation; namely, the Basel I and Basel II accords. It shows that harmonization of capital regulation may not occur endogenously if banking systems are opened up. In addition, complete harmonization may not be socially optimal. This argument would promote a certain level of discretion of national regulators in the calibration of capital regulation. However, this is not to say that the level of harmonization of capital regulation is unimportant. Harmonization of capital regulation prevents potential favoritism of domestic banks by the domestic regulation. This is the focus of the next section.

## 6.3 Expropriation

Although Section 6.2.3 assumes that the regulator acts in the best interest of all participants in the lending markets (i.e., banks, borrowers, and depositors), this may not always be the case. Regulators may act in the interest of voters that may favor domestic over foreign banks.[7] The action of the regulator by which it favors certain (usually domestic) banks is called expropriation. I assume throughout this section that foreign and domestic banks are initially on an identical footing; that is, they have identical intrinsic monitoring technology, $\underline{\nu}_F = \underline{\nu}_D = 0$. What distinguishes them is that the regulator imposes different requirements on them.

Whereas Section 6.2.1 analyzes cross-border entry and Section 6.2.2 late entry, this section could represent the mode of entry in which a foreign bank acquires a domestic bank and, by doing this, obtains a market share in the entering country.

I analyze two cases. First, I allow the regulator to limit the size of the market share that the foreign bank can grab. Second, the regulator may demand higher capital requirements from foreign banks.

### 6.3.1 Limited market share

I first analyze the situation in which the regulator expropriates foreign banks by limiting their size. In particular, at $t = 0$ the market is divided between banks unevenly, such that domestic banks obtain a higher market share than foreign banks. In particular, at $t = 0$ each foreign bank obtains $m_F$ incumbent borrowers, and each domestic bank obtains $m_D$ incumbent borrowers, where $m_D > m_F$. The proportion of domestic banks is $\gamma$ and the proportion of foreign banks is $[1 - \gamma]$. I show that the lower market share might put foreign banks at a substantial competitive disadvantage.

There exist several regulatory tools that may allow the regulator to limit the size of foreign banks. First, the regulator may object to a takeover of a large domestic bank by a foreign bank. A foreign bank may then enter only through an acquisition of a small domestic bank or by building its branch network from scratch. Second, and related, when a foreign bank decides to enter, the domestic market might already be divided between domestic banks.

I can now prove the following proposition.

**Proposition 6.4.** *For intermediate $S$ (i.e., $\frac{q_D^2 X^2}{2c\rho} < S < \frac{q_D^2 X^2}{c\rho}$), there exists $\hat{m}_F$ such that for $m_F < \hat{m}_F$ foreign banks lose their market share in competition with domestic banks. I*

[7] Kroszner and Strahan (1999) show that deregulation in the U.S. was largely driven by the interest of large U.S. banks.

*have*

$$\nu_F^* = m_F[1-q\gamma]\frac{X}{c\rho}, \tag{6.3}$$

$$\nu_D^* = \{m_D + m_F q[1-\gamma]\}\frac{X}{c\rho}. \tag{6.4}$$

Proposition 6.4 shows that foreign banks are hampered twice because the regulator limits their size. First, this directly hampers their profits. Second, size limit puts foreign banks at a competitive disadvantage. That is, foreign banks invest less in monitoring technology. Consequently, they lose borrowers in competition with domestic banks. They become even smaller and riskier. In contrast, a size advantage allows domestic banks to gain additional market share in competition with foreign banks. This induces them to invest more in monitoring technology. Domestic banks become larger and safer.

The implications of Proposition 6.4 must be confronted with Proposition 6.1. If monitoring technology can be transfered across borders (Proposition 6.1), large foreign banks may easily take additional market share from domestic banks. The stability of domestic institutions might be at stake. However, if monitoring technology is country-specific and the size of foreign banks is limited (Proposition 6.4), the market share of foreign banks in the domestic country might be depleted by large domestic banks that use their scale advantage. In this case, small foreign banks may become risky and, if many foreign banks exist, they may even pose a threat to the stability of the domestic banking system.

### 6.3.2 Biased capital regulation

The other possibility for expropriation is the asymmetric application of prudential regulation. In particular, the regulator may use capital regulation to expropriate foreign banks. Even though capital regulation has been quantified up to a certain extent (see the Basel I and Basel II agreements), regulators still possess some discretionary power in setting the level of capital requirements for a specific bank. It is not my intention to argue that discretion must be eliminated. In particular, in Proposition 6.3 I even argue that certain discretion is valuable. Now I highlight its potential dark side.[8] More specifically, the regulator may abuse its power to expropriate foreign banks.

In particular, the regulator may impose tighter capital requirements on foreign banks $k_F$ and looser capital requirements on domestic banks $k_D$, such that $k_F > k_D$. I define $X_F \equiv Y - [1-k_F]r_D$ and $X_D \equiv Y - [1-k_D]$, where $Y$ is a return of the borrower and $r_D$ is a cost of debt. I can now show the following result.

**Proposition 6.5.** *For intermediate $S$ (i.e., $\frac{Y^2}{2c\rho N} < S < \frac{[Y-r_D]^2}{c\rho N}$), there exists $\hat{k}_F$ such that*

[8]Discretionary power might be advantageous because it is hard to quantify all the risk that banks are exposed to.

*for $k_F > \hat{k}_F$ foreign banks lose their market share in competition with domestic banks. Thus*

$$\nu_F^* = [1 - q\gamma]\frac{X_F}{c\rho N}, \tag{6.5}$$

$$\nu_D^* = \{1 + q[1-\gamma]\}\frac{X_D}{c\rho N}, \tag{6.6}$$

*where $\nu_F^* < \nu_D^*$.*

The intuition for Proposition 6.5 is the following. If higher capital requirements are imposed only on foreign banks, this hampers their competitive ability. Even though foreign banks are otherwise identical to domestic banks, higher capital requirements increase their funding costs (because capital is costlier than deposits). Consequently, foreign banks may lose borrowers in competition with domestic banks. Foreign banks may become smaller, invest less in monitoring, and become riskier.

The key result of Proposition 6.5 is that raising capital requirements for one group of banks (for foreign banks) but not for the other (domestic banks) puts the first group at a competitive disadvantage. If differences in imposed capital requirements are substantial ($k_F > \hat{k}_F$), foreign banks are hampered so much that they lose market share and consequently they invest less in monitoring and become riskier. Hence, foreign banks become riskier because of higher capital requirements. In contrast, domestic banks gain market share and invest more in monitoring technology. Consequently, domestic banks become safer despite lower capital requirements.

Next, I show that higher competition exacerbates the distortions created by the differences in capital requirements.

**Corollary 6.6.** *Increasing competition augments the region in which imposing higher capital requirements on foreign banks increases their risk (i.e., $\frac{\partial \hat{k}_F}{\partial q} < 0$).*

Corollary 6.6 shows that differences in capital requirements across banks have the biggest effect if competition is high. If competition is high (high $q$), domestic banks can use the advantage of lower capital requirements over the larger market share. Hence, even a small difference in capital requirements between foreign and domestic banks ($\hat{k}_F$ is smaller if $q$ is high) puts foreign banks at such a competitive disadvantage that they lose market share and become riskier.

Proposition 6.5 and Corollary 6.6 shed some light on a related problem. Banks may use several mechanisms of "regulatory arbitrage" to circumvent capital requirements (see Jones (2000)). By doing this, some banks may lower the *effective* capital requirements. Proposition 6.5 then shows such banks may gain an unfair competitive advantage. Consequently, they could take additional market share, increase their investment in monitoring technology, and become safer.

I have analyzed expropriation of foreign banks by the domestic regulator. Now I turn to a more dynamic setting in which the staying power of a bank becomes important.

## 6.4 Staying Power of a Bank

The above analysis has focused on entry into banking markets. Now I extend this by allowing banks to exit the market. This brings new dynamics into the model. This affects banks' investments in monitoring and also their entry decision.

In particular, at $t = 1$ banks are subjected to a shock $h$ with probability $p$. With probability $[1 - p]$ there is no shock. If the shock occurs, its size $h$ is uniformly distributed at the interval $[0, H]$. If the shock occurs, a bank may still avoid the loss $h$ by exiting the market. Hence, the bank weights the loss $h$ of a shock with the expected profit if it continues with its operations. If the shock exceeds the bank's expected profit, the bank exits from the market. In this case, the bank avoids the shock $h$, but also does not obtain any profits.

I assume for simplicity that the conditions for borrowers do not change if their bank exits the market. That is, the borrowers' success probability continues to correspond to the monitoring technology of the initial bank. The borrower is simply transfered to another bank and it receives financing at the same conditions as agreed with its incumbent bank. I also assume that there is no competition for borrowers and that the market share of foreign banks $m_F$ is lower than the market share of domestic banks $m_D$; that is, $m_F < m_D$ (similarly to in Section 6.3.1).

I can now prove the following lemma.

**Lemma 6.2.** *Smaller foreign banks exit more often than larger domestic banks.*

Lemma 6.2 stems from the fact that foreign banks expect lower profits than domestic banks because of their smaller market share. Consequently, a relatively small shock that has no influence on large domestic banks might force foreign banks to pull back from their investments abroad.

What Lemma 6.2 shows is that foreign banks are more affected by the exit possibility. In particular, small foreign banks face an additional disadvantage. Due to their small size, they are more prone to exit the banking system more often than large domestic banks. To gain staying power they must become sufficiently large.

The following proposition analyzes what impact the probability of exit has on the level of investment in monitoring.

**Proposition 6.6.** *A higher probability of a shock $p$ raises the investment in monitoring technology if a shock is small ($H > \bar{H}$) but lowers the investment in monitoring technology if a shock is large ($H \leq \bar{H}$).*

This proposition stems from the following effect. If a small shock is expected, banks increase investments in monitoring technology and bank stability improves. That is, a small shock increases the value of monitoring; namely, banks with higher monitoring technology better defend against potential shock. In contrast, if a large shock occurs, banks are forced to exit regardless of their monitoring technology. If the bank exits, its investment in monitoring is useless. Hence, a higher probability of a large shock lowers bank monitoring.

Proposition 6.6 shows that the response of the bank to an increased probability of a shock crucially depends on the size of the shock. Leaning on Proposition 6.6, the regulators need not pay excessive attention to potential small shocks that may hamper bank assets. Even though a small shock might lower a bank's profitability, it increases investment in monitoring and bank stability. However, regulators should be more concerned if an expected shock is large. A large expected shock has two effects. First, if it occurs, it forces banks to exit. Second, banks that anticipate such a shock invest less in monitoring technology ex-ante. Hence, they become less stable even in the case of no shock.

I can now show the following result.

**Proposition 6.7.** *Increasing capital requirements augments the investment in monitoring technology if a shock is large ($H > \hat{H}$) and cost of monitoring is low ($c < \bar{c}$), but lowers the investment in monitoring if a shock is small ($H \leq \hat{H}$) and the cost of monitoring is high ($c \geq \bar{c}$).*

Increasing capital requirements has two opposite effects on monitoring incentives. The familiar effect is that higher capital requirements force banks to correctly price their risks and limit their reliance on deposit insurance. This effect increases the level of investment in monitoring technology. Yet, in the presence of the bank's exit option, an opposite effect of capital requirements is at work. Higher capital requirements hamper banks' profits, which in turn elevate the probability of exit. Banks expecting a higher probability of exit invest less in monitoring.

Two parameters determine whether the positive or negative effect of capital requirements on monitoring incentives prevails. First, if the expected shock to the bank's value is small ($H \leq \hat{H}$), the bank's decision to exit depends substantially on its expected profits. In this case, increasing capital requirements lowers banks' profits, augments exit, and also hampers investment in monitoring. Yet, if the expected shock to a bank's value is high ($H > \hat{H}$), a shock typically forces a bank to exit. Even though higher capital requirements hamper expected profits, this does not mainly influence the bank's exit decision. Hence, the negative effect of higher capital requirements on monitoring technology is limited and the positive effect prevails.

Second, if the cost of monitoring is low ($c < \bar{c}$), the positive effect of capital requirements on monitoring becomes more pronounced. Namely, banks respond more to higher capital requirements by increasing their monitoring levels. In contrast, if the cost of monitoring is high ($c \geq \bar{c}$), the bank's expected profit is low and the negative effect of capital requirements is at work. In particular, increasing capital requirements significantly lowers the value of the bank, which leads to more exit. A bank that anticipates a higher probability of exit invests less in monitoring. Consequently, increasing capital requirements results in less monitoring.

## 6.5 Conclusions

This chapter provides some key insights on the interplay between regulation and foreign entry. In particular, I show that opening up a weak banking sector leads to entry of strong foreign banks. Weak domestic banks may lose their market share, invest less in monitoring, and become riskier. Higher capital requirements induce foreign entry even further. Domestic banks may react to this. That is, they want to merge to gain scale advantages and prevent an influx of strong foreign banks.

I show also that higher capital requirements increase merger incentives. This is because higher capital requirements make investment in monitoring more valuable, which is exactly what merger brings.

I also allow for expropriation. In particular, the domestic regulator may act in favor of domestic banks and impose stricter regulation on foreign banks. I show that increasing capital requirements only for foreign banks may put them at a competitive disadvantage. They may lose market share, invest less in monitoring technology, and become riskier. However, domestic banks gain in that they face weaker competition. This helps them in preserving monitoring, and helps offset the potential adverse impact that cross border banking may have on a weak domestic banking system. As argued in Chapter 3 (see Section 3.7), just opening a domestic market to foreign competition might have severe adverse effects; acquisition of local players by foreign entrants might be preferred.

## 6.6 Appendix

*Proof of Proposition 6.1*

Note that all banks are (initially) of equal size. This implies that country $O$ (with two banks) is twice as large as country $A$. In total there are 3 banks, each with $1/N = 1/3$ of total borrowers. I normalize the total borrowers (over the two countries) to one to provide symmetry with the earlier analysis.

**Case** 1: Proposition 3.1 establishes that banks do not gain market share from banks of equal type. This immediately implies that good banks in country $O$ hold on to their market share. Banks do not change their levels of monitoring. To see this, note that, because there are only good banks in the market $\gamma = 1$, which implies that the optimal level of monitoring $\nu_G^*$ is not a function of $q$, see (3.9). One-sidedly opening up borders, however, increases the competition parameter in country $O$. This reduces the value of banks in country $O$; that is, observe that (3.17) is a decreasing function of competition parameter $q$, for $\gamma = 1$; that is,

$$\left.\frac{\partial V_G^*}{\partial q}\right|_{\gamma=1} = -\frac{1}{N}[-k + \frac{\nu_B X}{\rho}] + \frac{S}{\rho N} - \frac{1}{\rho N}[\underline{\nu}_G - \underline{\nu}_B]X - \frac{X^2}{2c[\rho N]^2}\{2[1-q]q + 2q\},$$

which is always negative (see that Assumption 3.1 guarantees that $S < [\underline{\nu}_G - \underline{\nu}_B]X$).

**Case** 2: Proposition 3.1 establishes that bad banks lose their market share to good banks. When the borders are closed, there are only bad banks in country $O$ (i.e., $\gamma = 0$), and banks in country $O$ invest $\nu_{BC}^* = X/c\rho N + \underline{\nu}_B$ (see (3.8)) in monitoring technology. After opening up borders, a bad bank in country $O$ competes with equal probability with a good or bad bank; this means that $\gamma = 1/2$. Bad banks in country $O$ now invest $\nu_{BO}^* = [1 - q/2]X/c\rho N + \underline{\nu}_B$. Thus, summarizing,

$$\nu_{BC}^* = X/c\rho N + \underline{\nu}_B, \quad \nu_{BO}^* = [1 - q/2]X/c\rho N + \underline{\nu}_B. \tag{6.7}$$

Observe that $\nu_{BO}^* < \nu_{BC}^*$. From (3.16) it follows that opening borders decreases the country $O$ bank values $V_B^*$ because $V_B^*$ is decreasing in both $q$ and $\gamma$; thus $V_{BO}^* < V_{BC}^*$. To compute the monitoring level of the good bank in country $A$ before the borders of country $O$ are opened, insert $\gamma = 1$ and $q = 0$ into (3.9) to obtain $\nu_{GC}^* = X/c\rho N + \underline{\nu}_G$. After borders are opened, the good bank has access to the borrowers from the bad banks in country $O$. Now $q > 0$ and $\gamma = 0$, and from (3.9) I have $\nu_{GA}^* = [1 + q]X/c\rho N + \underline{\nu}_G$. Summarizing,

$$\nu_{GC}^* = X/c\rho N + \underline{\nu}_G, \quad \nu_{GA}^* = [1 + q]X/c\rho N + \underline{\nu}_G. \tag{6.8}$$

■

*Proof of Corollary 6.1*

Use adapted versions of (3.7) to compute the values of the good bank in country $A$ before ($V_{GC}^*$) and after ($V_{GO}^*$), when it gains access to country $O$. The adapted versions of (3.7) that take into account one-sided competition are

$$V^*_{GC} = \frac{-k + \nu^*_{GC} X/\rho}{N} - c\frac{[\nu^*_{GC} - \underline{\nu}_G]^2}{2} \tag{6.9}$$

and

$$V^*_{GA} = \frac{-k + \nu^*_{GA} X/\rho}{N} + \frac{q}{\rho N}\{-S + [\nu^*_{GA} - \nu^*_{BO}]X\} - c\frac{[\nu^*_{GA} - \underline{\nu}_G]^2}{2}. \tag{6.10}$$

Insert $\nu^*_{GA}$, $\nu^*_{GC}$ and $\nu^*_{BO}$ from (6.7) and (6.8) into (6.9) and (6.10) to obtain

$$V^*_{GC} = \frac{-k + \underline{\nu}_G X/\rho}{N} + \frac{X^2}{2c\rho^2 N^2}, \tag{6.11}$$

and

$$V^*_{GA} = \frac{-k + \underline{\nu}_G X/\rho}{N} + \frac{[1+q]X^2}{c\rho^2 N^2} + \frac{q}{\rho N}\{-S + [\underline{\nu}_G - \underline{\nu}_B]X\} + \frac{3q^2 X^2}{2c\rho^2 N^2} - \frac{[1+q]^2 X^2}{2c\rho^2 N^2}. \tag{6.12}$$

Now compute the difference between (6.11) and (6.12) to see what the value of entering country $O$ is to the bank in country $A$. This yields

$$V^*_{GA} - V^*_{GC} = \frac{q}{\rho N}\{-S + [\underline{\nu}_G - \underline{\nu}_B]X\} + \frac{q^2 X^2}{c\rho^2 N^2}.$$

which is always increasing in $X$ and therefore also increasing in $k$. This completes the proof. ■

*Proof of Corollary 6.2*

Assume that the domestic banks behave in a closed domestic market as monopolists (i.e., $q = 0$). Merging then does not change the level of competition between domestic banks. Opening up the border increases $q$ to the level $q_O > 0$. Note that allowing for competition between domestic banks has no qualitative impact; it simply elevates all values of $q$ without changing the order. The investment in monitoring technology of a bad ($B$) merged ($M$) bank with open ($O$) borders is $\nu_{BOM}$. The merged bank can defend its borrowers from an entering good bank if

$$[\nu^*_{GA} - \nu^*_{BOM}]X < S, \tag{6.13}$$

where $\nu^*_{BOM}$ follows from maximizing,

$$V_{BOM} = \frac{2[1-q]}{N}[-k + \frac{\nu_{BOM} X}{\rho}] + 2\frac{q}{\rho N}\{S + [\nu^*_{BOM} - \nu^*_{GC}]X\} - c\frac{[\nu_{BOM} - \underline{\nu}_B]^2}{2}, \tag{6.14}$$

and $\nu^*_{GA}$ and $\nu^*_{GC}$ are given in (6.8). Maximizing (6.14) w.r.t. $\nu_{BOM}$ yields

$$\nu^*_{BOM} = \frac{2X}{c\rho N} + \underline{\nu}_B. \tag{6.15}$$

Hence, the merged bank can defend its borrowers from an entering good bank if (insert (6.8) and (6.15) into (6.13))

$$[\underline{\nu}_G - \underline{\nu}_B]X - [1 - q_O]X^2/c\rho N < S. \tag{6.16}$$

The condition in (6.16) is satisfied for a low enough $q_O$ because of the condition in (6.2). The left side of (6.16) is continuously increasing in $q_O$. Note from Assumption 3.1 that the condition in (6.16) is not satisfied for $q_O = 1$. Thus, there exists a $\hat{q}(S)$ such that the condition in (6.16) is satisfied for all $q_O$ for which $q_O < \hat{q}(S)$. Note also that $\frac{\partial \hat{q}(S)}{\partial S} > 0$. This concludes the proof. ■

*Proof of Corollary 6.3*

For simplicity, it is assumed as in the proof of Corollary 6.2 that the domestic banks behave in a closed domestic market as monopolists (i.e., $q = 0$). Opening up the border increases $q$ to the level $q_O > 0$. The value of each domestic bad bank when the border is opened and there is no merger is (see (3.7))

$$V_{BO} = \frac{1 - q_O}{N}[-k + \frac{\nu_{BO}X}{\rho}] - c\frac{[\nu_{BO} - \underline{\nu}_{BO}]^2}{2}. \tag{6.17}$$

Note that a bad bank loses its borrower to the good entering bank (this happens with probability $q_O$). Each bad domestic bank maximizes (6.17) by selecting the monitoring level $\nu_{BO} = \nu^*_{BO} = \frac{[1-q_O]X^2}{c\rho^2 N^2} + \underline{\nu}_B$ to obtain the value

$$V^*_{BO} = \frac{1 - q_O}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + \frac{[1 - q_O]^2}{2}\frac{X^2}{c\rho^2 N^2}. \tag{6.18}$$

Observe that the value of a merged bad bank facing $q = 0$ is

$$V_{BOM} = \frac{2}{N}[-k + \frac{\nu_{BOM}X}{\rho}] - c\frac{[\nu_{BOM} - \underline{\nu}_B]^2}{2}. \tag{6.19}$$

The situation is as follows. Assume (6.1) holds, and then a potential entrant will abstain from entering. To save on the entry cost, it will not even pose any competitive threat. Hence $q = 0$. The merged bank now maximizes (6.19) by investing $\nu_{BOM} = \nu^*_{BOM} = \frac{2X^2}{c\rho^2 N^2} + \underline{\nu}_B$. Inserting this into (6.19) yields

$$\frac{V^*_{BOM}}{2} = \frac{1}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + \frac{X^2}{c\rho^2 N^2}. \tag{6.20}$$

Compute the benefits of merging from (6.20) and (6.18) to obtain

$$MB_{BO} = \frac{V^*_{BOM}}{2} - V^*_{BO} = \frac{q_O}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + [1 - \frac{[1 - q_O]^2}{2}]\frac{X^2}{c\rho^2 N^2}. \tag{6.21}$$

If the borders are closed, the value of a bad bank is

$$V_{BC} = \frac{1}{N}[-k + \frac{\nu_{BC}X}{\rho}] - c\frac{[\nu_{BC} - \underline{\nu}_{BC}]^2}{2}. \tag{6.22}$$

The optimal monitoring is $\nu^*_{BC} = \frac{X}{c\rho N} + \underline{\nu}_B$. Insert this into (6.22) to get

$$V^*_{BC} = \frac{1}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + \frac{1}{2}\frac{X^2}{c\rho^2 N^2}. \tag{6.23}$$

The value of a merged bad bank is the same as given in (6.20). The benefits of merging are (use (6.23) and (6.20))

$$MB_{BC} = \frac{V^*_{BCM}}{2} - V^*_{BC} = \frac{1}{2}\frac{X^2}{c\rho^2 N^2}. \tag{6.24}$$

Compute the difference between (6.24) and (6.21) to obtain

$$MB_{BO} - MB_{BC} = \frac{V^*_{BCM}}{2} - V^*_{BC} = \frac{q}{N}[-k + \frac{\underline{\nu}_B X}{\rho}] + \frac{1 - [1 - q_O]^2}{2}\frac{X^2}{c\rho^2 N^2}. \tag{6.25}$$

Note that (6.25) is always positive. Thus, for bad banks merging is more beneficial when borders are opened.

In the case of good domestic banks, opening up borders has no impact. The entry cost together with anticipating zero market share prevents entry even without a merger, and hence there is no valuation impact. This concludes the proof. ■

*Proof of Corollary 6.4*

Differentiate (6.24) with respect to $k$ to obtain

$$\frac{\partial MB_{BC}}{\partial k} = \frac{r_D X}{c\rho^2 N^2}, \tag{6.26}$$

which is positive. Note also that the derivative of (6.21) with respect to $k$ is positive. This concludes the proof. ■

*Proof of Proposition 6.2*

The entering bank only knows its type once it has entered. If it is bad, it cannot win any borrowers because of the incumbency disadvantage and its value is zero. If it turns out to be good, its only possibility is to attract borrowers from bad banks. Its value is (use (3.7))

$$V_{G,\text{late}} = \frac{q[1-\gamma]}{\rho N}\{-S + [\nu_{G,\text{late}} - \nu^*_B]X\} - c\frac{[\nu_{G,\text{late}} - \underline{\nu}_G]^2}{2}. \tag{6.27}$$

The first part in (6.27) represents the profits from the borrowers that the entering bank takes in expectation from bad banks (see (3.6)). The entering bank competes with a bad bank w.p. $q[1-\gamma]$. Maximizing (6.27) yields

$$\nu^*_{G,\text{late}} = q[1-\gamma]X/c\rho N + \underline{\nu}_G. \tag{6.28}$$

Now I show that the late entrant bank conditional on being good can overcome the incumbency advantage of the existing bad banks. For this one needs $[\nu^*_{G,\text{late}} - \nu^*_B]X > S$. Use (3.8) and (6.28) to obtain

$$[\underline{\nu}_G - \underline{\nu}_B]X > S + [1-q]X^2/c\rho N. \tag{6.29}$$

Note that (6.2) assures that (6.29) is satisfied. Insert (3.8) and (6.28) into (6.27) to obtain

$$V^*_{G,\text{late}} = \frac{q[1-\gamma]}{\rho N}\{-S + [\underline{\nu}_G - \underline{\nu}_B]X\} - \frac{q[1-\gamma][2-q-q\gamma]X^2}{2c\rho^2 N^2}. \tag{6.30}$$

The expected value of the bank prior to entering is

$$V^*_{\text{late}} = \gamma \times 0 + [1-\gamma] \times V^*_{G,\text{late}}. \tag{6.31}$$

Use (6.27) to write (6.31) as

$$V^*_{\text{late}} = \frac{q\gamma[1-\gamma]}{\rho N}\left\{-S + [\underline{\nu}_G - \underline{\nu}_B]X - \frac{[2-q-q\gamma]X^2}{2c\rho N}\right\}. \tag{6.32}$$

Observe that (6.32) is zero for $q = 0$. The expression in (6.32) is continuous and increasing in $q$. Use Assumption 3.1 to see that the expression in curly brackets is strictly positive for $q = 1$. Thus late entry only occurs for a sufficiently high $q$ and an entry cost sufficiently small. In addition, it readily follows that (6.32) is maximized at an interior $\gamma$, and that for a sufficiently small entry cost and sufficiently high $q$, late entry is observed for $\gamma \in [\underline{\gamma}, \bar{\gamma}]$. ■

*Proof of Corollary 6.5*

Differentiate (6.32) w.r.t. $k$ to obtain

$$\frac{\partial V^*_{\text{late}}}{\partial k} = \frac{q\gamma[1-\gamma]r_D}{\rho N}\left\{\underline{\nu}_G - \underline{\nu}_B - \frac{[2-q-q\gamma]X}{c\rho N}\right\}. \tag{6.33}$$

If late entry occurs, $V^*_{\text{late}} > 0$. Observe from (6.32) that this implies that

$$[\underline{\nu}_G - \underline{\nu}_B]X - \frac{[2-q-q\gamma]X^2}{2c\rho N} > 0. \tag{6.34}$$

The expression (6.34) guarantees that, if late entry occurs, (6.33) is positive. Because late entry is more profitable, the value of late entry (6.32) surpasses the entry cost for a larger range of parameter values, thus, $\frac{\partial \underline{\gamma}}{\partial k} < 0$ and $\frac{\partial \bar{\gamma}}{\partial k} > 0$. ■

*Proof of Lemma 6.1*

This proof follows the proof of Proposition 3.6 in Chapter 3 and is omitted. ■

*Proof of Proposition 6.3*

This proof follows the proof of Proposition 3.6 in Chapter 3 and is omitted. ■

*Proof of Proposition 6.4*

I first conjecture that domestic banks take borrowers from foreign banks but neither domestic nor foreign banks can take borrowers from banks of the same type. Conditional on the fact that foreign banks obtain borrowers from foreign banks but not from domestic banks, their profits are

$$V_F = m_F[1-q][-k+\frac{\nu_F X}{\rho}] + m_F\frac{q}{\rho}[1-\gamma]\{S+[\nu_F-\nu_F^*]X\} - c\frac{\nu_F^2}{2}. \tag{6.35}$$

Maximization with respect to $\nu_F$ yields

$$\frac{\partial V_F}{\partial \nu_F} = \{m_F[1-q] + m_F q[1-\gamma]\}\frac{X}{\rho} - c\nu_F = 0. \tag{6.36}$$

Solve (6.36) to obtain (6.3).

Conditional on the fact that domestic banks obtain borrowers from foreign banks but not from domestic banks, their profits are

$$\begin{aligned} V_D = {} & m_D[1-q][-k+\frac{\nu_D X}{\rho}] + m_D\frac{q}{\rho}\gamma\{S+[\nu_D-\nu_D^*]X\} + m_D\frac{q}{\rho}[1-\gamma]\{S+[\nu_D-\nu_F^*]X\} \\ & + m_F\frac{q}{\rho}[1-\gamma]\{-S+[\nu_D-\nu_F^*]X\} - c\frac{\nu_D^2}{2}. \end{aligned} \tag{6.37}$$

Maximization with respect to $\nu_D$ yields

$$\frac{\partial V_D}{\partial \nu_D} = \{m_D[1-q] + m_D q\gamma + m_D q[1-\gamma] + m_F q[1-\gamma]\}\frac{X}{\rho} - c\nu_D = 0. \tag{6.38}$$

Solve (6.38) to obtain (6.4).

Now, I show that a conjecture that domestic banks cannot take borrowers in competition with another domestic bank is satisfied as long as $S > \frac{m_D^2 X^2}{2c\rho}$. To prove this, it is enough to prove that a domestic bank does not deviate to higher monitoring levels when surrounded only by domestic banks and where competition is high (i.e., $\gamma = 1$ and $q = 1$). Observe that the value of a deviating bank is

$$\hat{V}_D = \frac{m_D}{\rho}\{S+[\hat{\nu}_D-\nu_D^*]X\} + \frac{q_D}{\rho}\{-S+[\hat{\nu}_D-\nu_D^*]X\} - c\frac{\nu_D^2}{2}. \tag{6.39}$$

Maximize (6.39) with respect to $\hat{\nu}_D$ to obtain

$$\hat{\nu}_D^* = \frac{2m_D X}{c\rho}. \tag{6.40}$$

Compute the difference between value of a deviating and non-deviating domestic bank to obtain

$$\hat{V}_D(\hat{\nu}_D^*) - V_D(\nu_D^*) = \frac{m_D^2 X^2}{2c\rho^2} - S\frac{m_D}{\rho^2}. \tag{6.41}$$

If $S > \frac{m_D^2 X^2}{2c\rho}$, (6.41) is negative and a domestic bank never deviates to win a borrower in competition with another domestic bank.

Now I show that, for $S < \frac{m_D^2 X^2}{c\rho}$, there exists a $\hat{m}_F$ such that for $m_F < \hat{m}_F$ a foreign bank loses its borrower in competition with a domestic bank. I prove this by showing that deviation from (6.3) to higher monitoring levels to defend borrowers is not profitable. A foreign bank that defends its borrower in competition with domestic banks has the following profits:

$$\hat{V}_F = m_F[1-q][-k+\frac{\hat{\nu}_F X}{\rho}]+m_F[1-\gamma]\frac{q}{\rho}\{S+[\hat{\nu}_F-\nu_F^*]X\}+m_F\gamma\frac{q}{\rho}\{S+[\hat{\nu}_F-\nu_D^*]X\}-c\frac{\hat{\nu}_F^2}{2}. \tag{6.42}$$

The value of monitoring that maximizes (6.42) is given as

$$\hat{\nu}_F^* = \frac{m_F X}{c\rho}. \tag{6.43}$$

A deviating foreign bank obtains an additional return on borrowers for which it competes with domestic banks:

$$\gamma\frac{q}{\rho}\{S+[\hat{\nu}_F^*-\nu_D^*]X\}. \tag{6.44}$$

Note, however, that

$$S+[\hat{\nu}_F^*-\nu_D^*]X = S - \frac{m_D X^2}{c\rho} + \frac{[1-q]m_F X^2}{c\rho}. \tag{6.45}$$

Note that there exists a sufficiently small $\hat{m}_F$, such that for $m_F < \hat{m}_F$ the expression in (6.45) is negative. This concludes the proof. ■

*Proof of Proposition 6.4*

I first conjecture that domestic banks take borrowers from foreign banks but neither domestic nor foreign banks can take borrowers from banks of the same type. Later I confirm that these conjectures are true. I define $X_F = Y - [1-k_F]r_D$ and $X_D = Y - [1-k_D]r_D$. Conditional on the fact that foreign banks obtain borrowers from foreign banks but not from domestic banks, their profits are

$$V_F = \frac{1-q}{N}[-k+\frac{\nu_F X_F}{\rho}] + \frac{q}{\rho N}[1-\gamma]\{S+[\nu_F-\nu_F^*]X_F\} - c\frac{\nu_F^2}{2}. \tag{6.46}$$

Maximization with respect to $\nu_F$ yields

$$\frac{\partial V_F}{\partial \nu_F} = [1-q\gamma]\frac{X_F}{\rho N} - c\nu_F = 0. \tag{6.47}$$

Solve (6.47) to obtain (6.5).

Conditional on the fact that domestic banks obtain additional borrowers from foreign banks but not from domestic banks, their profits are

$$\begin{aligned} V_D &= [1-q][-k+\frac{\nu_D X_D}{\rho N}]+\frac{q}{\rho N}\gamma\{S+[\nu_D-\nu_D^*]X_D\}+\frac{q}{\rho N}[1-\gamma]\{S-k_D+k_F \\ &\quad +\nu_D X_D-\nu_F^* X_F\}+\frac{q}{\rho N}[1-\gamma]\{-S-k_F+k_D+\nu_D X_D-\nu_F^* X_F\}-c\frac{\nu_D^2}{2}. \end{aligned} \tag{6.48}$$

Maximization with respect to $\nu_D$ yields

$$\frac{\partial V_D}{\partial \nu_D}=\{1-q+q\gamma+q[1-\gamma]+q[1-\gamma]\}\frac{X_D}{\rho N}-c\nu_D=0. \tag{6.49}$$

Solve (6.49) to obtain (6.6). The following condition assures that $\nu_D$ is lower than 1:

$$\frac{2Y}{c\rho N}<1. \tag{6.50}$$

A conjecture that domestic banks cannot take borrowers in competition with another domestic bank is satisfied as long as $S>\frac{Y^2}{2c\rho}$ (see the proof of Proposition 6.4). Now I show that, for $S<\frac{[Y-r_D]^2}{c\rho}$, there exists a $\hat{k}_F$ such that for $k_F>\hat{k}_F$ a foreign bank loses its borrower in competition with a domestic bank. I will prove this by showing that deviation from (6.5) to higher monitoring levels to defend borrowers is not profitable. A foreign bank that defends its borrower in competition with domestic banks has the following profits:

$$\hat{V}_F=\frac{1-q}{N}[-k+\frac{\hat{\nu}_F X}{\rho}]+[1-\gamma]\frac{q}{\rho N}\{S+[\hat{\nu}_F-\nu_F^*]X_F\}+\gamma\frac{q}{\rho N}\{S-k_F+k_D\hat{\nu}_F X_F-\nu_D^* X_D\}-c\frac{\hat{\nu}_F^2}{2}. \tag{6.51}$$

The value of monitoring that maximizes (6.51) is given as

$$\hat{\nu}_F^*=\frac{m_F X_F}{c\rho}. \tag{6.52}$$

A deviating foreign bank obtains an additional return $DF$ on borrowers for which it competes with domestic banks:

$$DF=\gamma\frac{q}{\rho}\{S-k_F+k_D+\hat{\nu}_F^* X_F-\nu_D^* X_D\}. \tag{6.53}$$

Note that the condition in (6.50) guarantees that $\frac{\partial DF}{\partial k_F}=-1+\frac{2X_F r_D}{N\rho}<0$. Note that in extreme cases where $k_F=1$ and $k_D=0$

$$S-k_F+k_D+\hat{\nu}_F^* X_F-\nu_D^* X_D=S+\frac{Y^2}{c\rho N}-\frac{[1+q][Y-r_D]^2}{c\rho}<0. \tag{6.54}$$

Note that there exists a sufficiently small $\hat{m}_F$, such that for $m_F<\hat{m}_F$ the expression in

(6.45) is negative. This concludes the proof. ■

*Proof of Lemma 6.2*
At $t = 1$ the cost of investment in monitoring is already sunk. The expected profits conditional on the absence of a shock are

$$\Pi_\tau(t = 1) = m_\tau[-k + \frac{\nu_\tau X}{\rho}] \text{ for } \tau \in \{F, D\}. \tag{6.55}$$

However, with probability $[1 - m]$ a shock $f$ occurs. Note that for a foreign bank $\Pi_F(t = 1) = m_F[-k + \frac{\nu_F X}{\rho}]$ and for a domestic bank $\Pi_D(t = 1) = m_D[-k + \frac{\nu_D X}{\rho}]$. Both $m_D > m_F$ and $\nu_D > \nu_F$. Hence, $\Pi_D(t = 1) > \Pi_F(t = 1)$. Hence, probability that $h$ exceeds $\Pi_D$ is lower than probability that $h$ exceeds $\Pi_F$. This concludes the proof. ■

*Proof of Proposition 6.6*
Conditional on a presence of a shock, bank profits are

$$\int_0^{\Pi(t=1)} \frac{[\Pi(t = 1) - h]}{H} \mathrm{d}h = \frac{\Pi(t = 1)}{2}. \tag{6.56}$$

Combine (6.56) and (6.55) to obtain the expected profit of a bank before investing in the monitoring technology (at $t = 0$):

$$\Pi_F = [1 - p]m_F[-k + \frac{\nu_F X}{\rho}] + p\frac{m_F^2[-k + \frac{\nu_F X}{\rho}]^2}{2H} - \frac{c\nu_F^2}{2}. \tag{6.57}$$

Maximize (6.57) w.r.t. $\nu_F$ to obtain

$$\frac{\partial \Pi_F}{\partial \nu_F} = [1 - p]\frac{m_F X}{\rho} + p[-k + \frac{\nu_F X}{\rho}]\frac{m_F^2 X}{\rho H} - c\nu_F = 0 \tag{6.58}$$

Solve for $\nu_F$ to obtain

$$\nu_F^* = \frac{[1 - p - \frac{m_F pk}{H}]m_F X}{\rho[c - \frac{pm_F^2 X^2}{\rho^2 H}]}. \tag{6.59}$$

Rearrange (6.59) to obtain

$$\nu_F^* = \frac{\rho k}{X} + \frac{[1 - p - \frac{ck\rho^2}{m_F X^2}]}{c - \frac{pm_F^2 X^2}{\rho^2 H}} \frac{m_F X}{\rho}. \tag{6.60}$$

Observe that for high $H$ one has $\frac{\partial \nu_F}{\partial p} < 0$ and for low $H$ one has $\frac{\partial \nu_F}{\partial p} > 0$. ■

*Proof of Proposition 6.7*
Use (6.60) to observe that $\frac{\partial \nu_F}{\partial k} < 0$ for high $c$ and for low $H$. Also note that that $\frac{\partial \nu_F}{\partial k} > 0$ for low $c$ and for high $H$.

■

# Chapter 7

# Demand Deposits and Bank Monitoring

**Abstract**

This chapter provides a novel rationale for why banks combine lending and deposit taking. I show that demand deposits may commit banks to monitoring in an environment in which monitoring is most valuable for long-term projects. Demand deposits, contrary to straight short-term and long-term debt, help commit a bank not to overleverage itself and this commits the bank to monitoring. I show that banks prefer demand deposits if the cost of early liquidation of borrowers is intermediate and if bank monitoring is costly.

Keywords: Demand Deposits, Bank Monitoring, Competition
JEL CLASSIFICATION: G21

## 7.1 Introduction

One of the important characteristics of banks is that they provide (withdrawable) demand deposits to investors and at the same time lend to borrowers and monitor them. This chapter addresses the question of why banks combine lending and deposit taking. The key idea developed in this paper is that banks may have a problem committing not to overleverage themselves; existing investors of the bank might suffer from subsequent new lending to their bank. What I show is that the specificness of demand-deposit contracts can be understood as a commitment device that prevents banks from excessive refinancing. That is, the threat of withdrawals of demand deposits disciplines the bank choice of the level of debt.

I analyze these issues in a multi-principal agent model in which investors (principals) contract with a bank (the agent). The bank's monitoring affects the success probability of its borrowers (i.e., companies that need bank financing). However, monitoring is costly and uncontractable, and a bank might be unable to commit to it.

I show that in the setting in which monitoring is most valuable for long-term projects, a bank may commit to monitoring using demand-deposit contracts as a costly commitment mechanism. In particular, (expected) withdrawals on demand-deposit contracts could force banks to liquidate some of their borrowers early to repay debt and, in doing so, limit debt financing in later periods. Although partial liquidation is less profitable than continuation, it allows the bank to realize (small) profits that can be used as equity in the second period. This partial equity financing alleviates moral hazard.

This still does not explain why short-term borrowing could not accomplish the same. Observe that the bank has to commit to prematurely liquidating some borrowers. With a single investor, the bank achieves this by borrowing short-term. With short-term borrowing, the single investor can limit funding to the bank and, in doing so, force the bank to liquidate some borrowers. Consequently, if the bank is forced to only borrow from a single investor, the bank can successfully commit to monitoring as in Bizer and DeMarzo (1992).

However, borrowing from multiple investors complicates matters: investors can no longer control the amount of bank borrowing because the bank may raise additional funds from other investors. Hence, the bank can avoid liquidation of borrowers and the commitment to monitoring is more difficult. This is reminiscent of the commitment problem of borrowers vis-à-vis their bank in Bizer and DeMarzo (1992) in which the existing bank suffers if a competing bank additionally lends to its borrowers.

Interestingly, borrowing from multiple investors using demand deposits resolves the refinancing problem. In particular, the bank may now credibly offer low rates and promise to monitor its borrowers. Investors with demand deposits only accept low rates if the bank has low levels of debt. If the bank reneges on this, investors anticipate higher bank risk (i.e., no bank monitoring) and withdraw their demand deposits. The threat of withdrawals on demand deposits therefore serves as a commitment device. Interestingly, if a bank could renege and take on more debt, the subsequent withdrawals reduce bank debt and, via forced liquidation, induce banks to monitor again (unless of course the liquidations are too costly and losses are realized; this then would make a bank run come about).

The intuition for the use of demand deposits differs somewhat from the established theories. In my analysis, withdrawals are not a consequence of exogenous liquidity needs of investors as in Diamond and Dybvig (1983), but serve as a threat to good behavior (as in Calomiris and Kahn (1991) and Diamond and Rajan (2001)).[1] Note however that the actual withdrawals in my analysis could help induce good behavior (i.e., bank monitoring). That is, the role of demand deposits in my analysis is to balance the bank's debt to the level where the bank can still commit to monitoring. This expands the notion of Flannery (1994), who argues that short-term debt, and demand deposits in particular, contain the incentives of banks to shift risk to investors. In contrast to Flannery (1994), my analysis directly connects demand deposits with the bank's role on the asset side; that is, with monitoring borrowers. It shows that demand deposits prevent the bank from overleveraging itself and this commits it to monitoring. Bizer and DeMarzo (1992) indicate that putable and callable bonds could play a similar role.[2]

This analysis is related to the rationale that Jensen (1986) and Tirole (2006, p. 204) offer on the use of short-term and long-term debt. They argue that an entrepreneur would like to commit to not overinvesting in the interim period; that is, paying out cash that is not needed. Although the entrepreneur can simply make a promise not to excessively invest, such a promise is not credible. To resolve this, the entrepreneur borrows short-term. This *forces* him in the future to pay out cash flows to short-term debt holders. In this way, the entrepreneur commits himself to release the cash rather than excessively invest.

My analysis differs from Tirole (2006) in that he suggests that excessive investment might be contained with a combination of short-term and long-term claims. I point to the *refinancing* problem that may arise in the case of multiple investors. Using a demand deposit contract could mediate this problem. This also distinguishes my analysis from that of Jensen (1986) and Stulz (1990), who point to the potential cost of underinvestment when insufficient funds are available.

A few other studies identify the benefits of the coexistence of demand deposits and lending to borrowers. Berlin and Mester (1999) argue that access to core deposits enables banks to insulate borrowers against exogenous credit shocks. Core depositors respond less to a potential change in the interest rate. Consequently, banks more heavily funded with core deposits can smooth loan rates more. Kashyap, Rajan, and Stein (2002) claim that banks provide liquidity both on the asset side and on the liability side. If the demand for

[1]Whereas in my analysis demand deposits prevent the bank from exploiting some investors and refinancing from others, in Gorton and Pennacchi (1990), demand deposits serve to protect uninformed investors from being abused by informed investors. See also Hart and Tirole (1990), McAfee and Schwartz (1994), and Segal and Whinston (2003) for an analysis of contracting problems with multiple principles and multiple agents.

[2]Myers and Rajan (1998) highlight that (excessively) liquid bank assets create potential expropriation problems on the part of the bank. That is, those liquid assets can "disappear" easily. Sufficient illiquidity might be needed to commit the bank not to abscond with the assets. The bank run literature (Diamond and Dybvig (1983), Jacklin and Bhattacharya (1988), etc.) emphasizes the opposite: lack of liquidity is problematic. In my analysis I point to the need for sufficient liquidity. Thus, there is a careful balance that could possibly reconcile the different theories.

liquidity by depositors and borrowers is imperfectly correlated, banks may realize synergies in combining both functions. Song and Thakor (2004) argue that banks use core deposits to prevent withdrawals based on disagreement between investors and banks, which is the most acute for the most opaque relationship loans. They conclude that banks match the highest-value liabilities with the highest-value loans.

My analysis also yields the following empirical prediction. As the liquidation cost of bank borrowers increases, banks have greater difficulties in committing to monitoring. Banks have to liquidate more borrowers (to realize higher short-term profits) to commit themselves to higher monitoring. That is, the total amount of bank lending contracts.

This chapter is organized as follows. In Section 7.2, I present model specifications. Section 7.3 compares long-term and short-term borrowing and using demand deposits. Section 7.4 extends the analysis and lists empirical predictions. Section 7.5 concludes the chapter. All proofs are relegated to Appendix.

## 7.2 Model Specification

### 7.2.1 Preliminaries

There are three types of players in the model: borrowers (companies asking for loans), investors, and commercial banks. All players are risk neutral.

Banks specialize in lending and borrow from investors. Banks monitor their borrowers by establishing a long-term relationship with them. Monitoring of a bank prevents a borrower's opportunistic behavior and increases the long-term success probability of borrowers (and of a bank). However, monitoring is costly and to save this cost a bank may decide not to monitor some of its borrowers. Summarizing, moral hazard problems show up in future periods, and not in the short-term. A bank may also prematurely liquidate lending to borrowers to earn some (albeit small) short-term profits.

The funding of banks comes from investors. Banks may borrow short-term or long-term, or use demand deposits. In the case of demand deposits, an investor may withdraw his funds at $t = 1$. In this setting, investors face two problems. First, bank monitoring is unobservable. Banks may stop monitoring, knowing that investors bear most of the potential down-size loss. Second, the bank may borrow from multiple investors. The bank may exploit this and (re)finance from some investors in a way that may be unfavorable for others.

### 7.2.2 Model details

*Borrowers:* The basic characteristics of borrowers are as follows. At $t = 0$, borrowers undertake projects that demand \$1 of investment in total. Because borrowers have no initial funds, they must borrow from a bank. At $t = 2$, the project yields $R$ in the case of a success and \$0 in case of a failure. If the borrower becomes safe, he always succeeds. However, the borrower can become risky. Risky borrowers fail if a systematic shock in the economy occurs. That is, the success probability of a risky borrower is only $p < 1$.[3] Without bank monitoring, borrowers always become risky because then they gain additional high private benefits.

*Banks:* At $t = 0$, each bank initially receives \$1 of funds from investors. It lends these funds to borrowers. The bank can liquidate a proportion $x$ of its borrowers at $t = 1$. In this case, the return $R$ per \$1 of the borrower's project is lowered by the liquidation cost $c_L$, such that the bank receives $x[R - c_L]$. Subsequently, the bank may monitor borrowers that are not liquidated; each at a cost $c_M$. That is, the cost of monitoring $m$ borrowers is $mc_M$. If monitored, the borrower must become safe.[4] The bank is a monopolist; namely, it can take all the returns from the borrowers' projects. Bank monitoring is profitable; that is,

$$R - c_M > pR > 1 \text{ and } R - c_M > R - c_L > 1. \tag{7.1}$$

I assume that the bank must borrow all funds needed from investors in the form of debt. There is no deposit insurance.[5] The bank may borrow either short-term from $t = 0$ to $t = 1$ with a gross interest rate $r_{S1}$ or from $t = 1$ to $t = 2$ with a gross interest rate $r_{S2}$, or long-term from $t = 0$ to $t = 2$ with a gross interest rate $r_L$ or with demand deposits. In the case of demand deposits, the bank promises a gross interest rate $r_{D1}$ for early withdrawals (at $t = 1$). At $t = 1$, the bank sets a gross interest rate $r_{D2}$ for late withdrawals (at $t = 2$).

*Investors:* Investors are risk neutral and demand a competitive risk free return, which is normalized to 1.[6]

*Timeline:* The bank gathers funds at $t = 0$. That is, the bank offers to investors the type (short-term, long-term borrowing, or demand deposits) and promised rate of a contract. Investors then decide whether they supply funds. The bank lends the raised funds to borrowers. At $t = 1$, the bank may again borrow from investors in the short-term. The bank sets the second period rate on demand deposits. The bank also needs to repay potential short-term investors or potential early withdrawals of demand deposits. Investors with demand deposits decide whether to withdraw in sequential order. The bank may liquidate some borrowers. The bank may then choose to monitor the remaining borrowers. At $t = 2$, payoffs are realized. The bank repays the remaining investors from the funds collected from borrowers. See Figure 7.1.

[3] Borrowers' returns are correlated as in Holmstrom and Tirole (1997). In the case of diversified returns, banks could always commit to monitoring; see Diamond (1984).

[4] Note that monitoring affects the long-term success probability of the borrower. This is aligned with evidence in Lummer and McConnell (1989). They distinguish between new bank loans and renewals, and show that renewals have a positive announcement effect but new bank loans do not. This suggests that banks monitor their borrowers through long-term relationships; see also Boot (2000).

[5] For the rationale on the use of bank capital, see Chapter 3 and Allen, Carletti, and Marquez (2007).

[6] In Diamond and Dybvig (1983), banks only provide liquidity if depositors are risk averse and if they value sufficiently long-term returns. In this analysis I focus on risk-neutral depositors.

| $t = 0$: | $t = 1$: | $t = 2$: |
|---|---|---|
| ♠ The bank borrows short-term or long-term, or raises demand deposits from investors.<br><br>♠ The bank lends the funds to borrowers. | ♠ The bank can again borrow short-term from investors and sets the second period rate on demand deposits.<br><br>♠ The bank has to repay investors with short-term contracts and early withdrawals of demand deposits. The bank can liquidate some borrowers.<br><br>♠ The bank decides whether to monitor its borrowers. | ♠ Payoffs are realized.<br><br>♠ Investors are repaid, if possible. |

Figure 7.1: Timeline

## 7.3 Optimal Bank Borrowing

Now I analyze different types of borrowing by the bank. The bank may borrow long-term, or short-term from a single investor, or short-term from mutliple investors, or use demand deposits.

### 7.3.1 Long-term borrowing

Now I analyze the bank's incentives to monitor if it borrows long-term funds and there exists a single investor. First, I analyze the bank's profits if it does not monitor its borrowers. Second, I allow for bank monitoring and show that the bank either monitors all borrowers or none. Third, I analyze the conditions under which long-term borrowing induces bank monitoring.

The bank's expected profit conditional on not monitoring its borrowers is as follows. With long-term borrowing, the bank does not need to liquidate any borrowers at $t = 1$; hence, $x = 0$. Borrowers repay the bank $R$ with probability $p$. The bank promises a long-term gross interest rate $r_L$ to the investor. Hence, the expected profit of the bank conditional on not monitoring its borrowers $\Pi_{NM}(x, r_L)$ is

$$\Pi_{NM}(0, r_L) = p[R - r_L]. \tag{7.2}$$

Alternatively, the bank may choose to monitor a proportion $m$ of its borrowers. In this case, it incurs an additional cost of monitoring $mc_M$; however, monitored borrowers now succeed with probability 1. That is, with additional probability $1 - p$ the bank gains $mR$ from the borrowers but has to pay $r_L$ to the investor (if possible). Hence, the expected profit of the bank $\Pi(x, m)$ is

$$\Pi(0, m) = \Pi_{NM}(0, r_L) + [1 - p] \max(mR - r_L, 0) - mc_M. \tag{7.3}$$

Observe that (7.3) is convex in $m$. Hence, it has its maximum either at $m = 0$ or at $m = 1$.

For $m = 0$, (7.3) is given in (7.2). For $m = 1$, (7.3) becomes

$$\Pi_M(0, r_L) = R - c_M - r_L. \tag{7.4}$$

The bank only monitors its borrowers if it has an incentive to do so. That is, the bank's profit conditional on monitoring in (7.4) must be larger than the one without monitoring in (7.2). Rearranging yields the following condition.

$$[1 - p][R - r_L] \geq c_M \tag{7.5}$$

If the investor anticipates bank monitoring, he supplies funds at $r_L = 1$. Use this with (7.5) to see that the bank monitors as long as $c_M \leq \bar{c}_M$, where

$$\bar{c}_M \equiv [1 - p][R - 1]. \tag{7.6}$$

For $c_M > \bar{c}_M$, the bank can no longer commit to monitoring; hence, it only succeeds with probability $p$. The investor anticipates this and only supplies funds at $r_L = \frac{1}{p}$. This immediately yields the following proposition.

**Proposition 7.1.** *If the bank borrows long-term, it can commit to monitoring for a low cost of monitoring (i.e., for $c_M \leq \bar{c}_M$, where $\bar{c}_M = [1 - p][R - 1]$) but not for a high cost of monitoring (i.e., for $c_M > \bar{c}_M$).*

The intuition for Proposition 7.1 is the following. The investor cannot write a contract with a bank contingent on the bank's monitoring. This creates the moral hazard problem. In particular, the bank may stop monitoring to save the cost of monitoring, knowing that the investor carries most of the downsize loss. For a small monitoring cost, the moral hazard problem is small and the bank can commit to monitoring. The investor anticipates bank monitoring and supply funds at $r_L = 1$. However, for a high monitoring cost, the moral hazard problem becomes dominant and the bank can no longer commit to monitoring, although monitoring is the first best option (see the condition in (7.1)). The investor anticipates no bank monitoring and only supplies funds at high rate $r_L = \frac{1}{p}$.

### 7.3.2 Short-term borrowing: A single investor

Now I show that the bank may mediate the moral hazard problem through *short-term* borrowing. I continue to assume that there exists a single investor. I proceed as follows. I first compute the expected profit of the bank at $t = 2$ conditional on liquidation of a proportion of $x$ borrowers at $t = 1$. I show that the bank either monitors all borrowers or no one. Second, I analyze the conditions under which the bank can commit to monitoring.

In the first period, the bank can borrow short-term at the risk-free rate (i.e., $r_{s1} = 1$). That is, the investor anticipates that he can be repaid at $t = 1$ because the bank can always liquidate a sufficient amount of borrowers (the condition in (7.1) guarantees that liquidation of borrowers yields more than the investor requires; i.e., $R - c_L > 1$). If the bank liquidates $x$ borrowers, it earns $x[R - c_L]$ and repays this amount to the investor at $t = 1$. That is, in the second period the investor lends $1 - x[R - c_L]$ to the bank.[7]

If the bank does not monitor its borrowers, it realizes the following expected profit at $t = 2$. Proportion $1 - x$ of borrowers is not liquidated and continues with projects. These projects yield $R$ with probability $p$; thus, the bank expects a revenue of $p[1 - x]R$. The bank promises a gross interest rate $r_{S2}$ to the investor. If successful, the bank must repay $\{1 - x[R - c_L]\}r_{S2}$ to its investor at $t = 2$. Hence, the profit of the bank conditional on not monitoring its borrowers is

$$\Pi_{NM}(x, r_{S2}) = [1 - x]pR - p\{1 - x[R - c_L]\}r_{S2}. \tag{7.7}$$

Alternatively, the bank may choose to monitor a proportion $m$ of its borrowers. In this case, it incurs additional cost of monitoring $mc_M$; however, monitored borrowers now succeed with probability 1. That is, with additional probability $1 - p$ the bank gains $[1 - x]mR$ from the borrowers but has to pay $\{1 - x[R - c_L]\}r_{S2}$ to the investor (if possible). Hence, the expected profit of the bank is

$$\Pi(m) = \Pi_{NM}(x, r_{S2}) + [1 - p]\max([1 - x]mR - \{1 - x[R - c_L]\}r_{S2}, 0) - [1 - x]mc_M. \tag{7.8}$$

Observe that (7.8) is convex in $m$. Hence, it has its maximum either at $m = 0$ or at $m = 1$. For $m = 0$, (7.8) is given in (7.7). For $m = 1$, (7.8) becomes

$$\Pi_M(x, r_{S2}) = [1 - x][R - c_M] - \{1 - x[R - c_L]\}r_{S2}. \tag{7.9}$$

The bank only monitors its borrowers if it has an incentive to do so. The bank's profit conditional on monitoring in (7.9) must be larger than the one without monitoring in (7.7). Rearranging yields the following condition.

$$[1 - x][R - \frac{c_M}{1 - p}] - \{1 - x[R - c_L]\}r_{S2} \geq 0. \tag{7.10}$$

The incentive constraint of the bank in (7.10) deserves a further discussion. Observe that liquidation of borrowers at $t = 1$ helps the bank more successfully commit to monitoring (compare (7.10) and (7.5)). This is because liquidation of borrowers allows the bank to realize some (small) profits that can be used as equity in the second period. This partial equity financing alleviates moral hazard in the second period and the bank can successfully commit to monitoring.

Now I analyze under which conditions the investor is willing to invest in a bank for another period; that is, from $t = 1$ to $t = 2$. These conditions crucially depend on the amount of liquidated borrowers, which defines whether the bank can commit to monitoring.

[7] I assume that the bank cannot expropriate the investor by liquidating the borrowers and directly taking funds out of the bank without paying the investor.

First, the bank may merely stop monitoring and offer $r_{S2} = \frac{1}{p}$. In this case, the investor anticipates no monitoring. He accepts the bank's offer and continues to provide full financing to the bank; that is, no borrower is liquidated. With probability $p$, the bank succeeds, gains $R$, and repays $r_{S2}$ to the investor. Hence, the profit of the bank is

$$\Pi_{NM}(0, \frac{1}{p}) = pR - 1. \tag{7.11}$$

Second, the bank may offer a low rate $r_{S2} = 1$. In this case, the investor only partially refinances the bank (the condition in (7.10) has to hold). The bank must liquidate some borrowers and realize (small) profits that can be used as equity financing in the second period and can commit the bank to monitoring. In this case, the condition (7.10) becomes

$$R - 1 - \frac{c_M}{1-p}[1-x] - c_L x \geq 0. \tag{7.12}$$

The bank's profit is (insert $r_{S2} = 1$ into (7.9))

$$\Pi_M(x, 1) = R - 1 - c_M[1-x] - xc_L. \tag{7.13}$$

Hence, the bank that commits to monitoring solves the following maximization problem.

$$\begin{aligned} \Pi_M(x^*, 1) &= \max_x R - 1 - c_M[1-x] - xc_L, \qquad (7.14) \\ & \quad \text{s.t.} \\ & \quad R - 1 - \frac{c_M}{1-p}[1-x] - c_L x \geq 0. \qquad (7.15) \end{aligned}$$

That is, the bank maximizes its profit in (7.13) under the constraint in (7.15). The constraint in (7.15) is identical to (7.12) and guarantees that the bank has an incentive to monitor its borrowers.

The bank decides whether to commit to monitoring and offer $r_{S2} = 1$ to the investor and gain (7.14) or not to commit to monitoring and offer $r_{S2} = \frac{1}{p}$ to the investor and gain (7.11). Figure 7.2 graphically depicts the maximization problem of the bank. The bank maximizes (7.14) at the constraint in (7.15); that is, $\Pi_M(x, 1) > \Pi_{NM}(x, 1)$, and at the constraint that (7.14) is bigger than (7.11); that is, $\Pi_M(x, 1) > \Pi_{NM}(0, \frac{1}{p})$.

I can now show the following proposition.

**Proposition 7.2.** *With a single investor and short-term borrowing, one has*

- *For a low monitoring cost $c_M \leq \bar{c}_M$, the investor fully refinances the bank at $r_{S2} = 1$. The bank liquidates no borrowers but monitors them.*

- *For a high monitoring cost $\bar{c}_M \leq c_M$ and for a low liquidation cost $c_L \leq \bar{c}_L(c_M)$, the investor only partially refinances the bank at $r_{S2} = 1$. The bank liquidates $x^*$ of borrowers and monitors the others.*

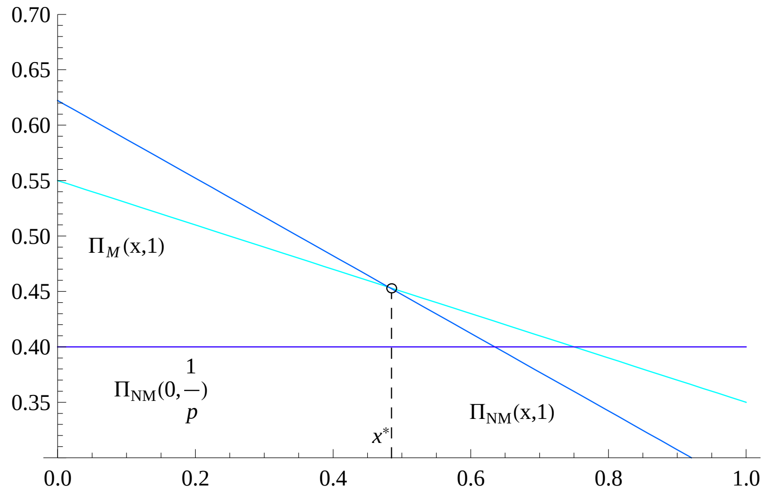


Figure 7.2: The bank's profits.

- *For a high monitoring cost $c_M > \bar{c}_M$ and for a high liquidation cost $c_L > \bar{c}_L(c_M)$, the bank does not monitor. The investor fully refinances the bank at $r_{S2} = \frac{1}{p}$.*

Proposition 7.2 shows that there exist three regions of bank financing. For a low monitoring cost $c_M < \bar{c}_M$, the bank can easily commit to monitoring. In this region, liquidation of borrowers is not necessary (i.e., it would only lower the total profit of the bank). However, for a high monitoring cost $c_M \geq \bar{c}_M$, commitment to monitoring is more difficult. Two regions can be distinguished. For a low liquidation cost $c_L < \bar{c}_L$, the investor at $t = 1$ only partially refinances the bank, forcing it to liquidate a part of its borrowers at $t = 1$. This commits the bank to monitoring. For a high liquidation cost $c_L > \bar{c}_L$, however, liquidation no longer helps commit the bank to monitoring. The bank neither monitors nor liquidates its borrowers.

The argument why liquidation of some borrowers commits banks to monitoring deserves further explanation. Monitoring helps the long-term ($t = 2$) success rate of the borrower (and hence of the bank), but not the short-term ($t = 1$) success rate. The bank may liquidate borrowers at $t = 1$ to gain some (albeit limited) profits. These profits can be used as equity (i.e., the investor can be partially repaid at $t = 1$) and partial equity financing alleviates moral hazard. Note that the investor limits funding to the bank and, in doing so, forces the bank to liquidate some of its borrowers. Hence, the bank can successfully commit to monitoring.[8]

The next corollary connects the monitoring cost with the proportion of liquidated borrowers.

**Corollary 7.1.** *The proportion of liquidated borrowers $x^*$ increases in a monitoring cost $c_M$ and in a liquidation cost $c_L$ (i.e., $\frac{\partial x^*}{\partial c_M} > 0$ and $\frac{\partial x^*}{\partial c_L} > 0$).*

[8] In this analysis, banks commit to monitoring by borrowing short-term. Allen, Carletti, and Marquez (2007) argue that in order to commit to monitoring banks may raise excessive levels of capital.

Corollary 7.1 builds on the intuition that banks with high monitoring costs must liquidate more borrowers. Only then do they realize sufficient profits that serve as partial equity financing and commit them to monitoring.

I also obtain the following empirical prediction. As the liquidation cost of bank borrowers increases, banks have to liquidate more borrowers (to realize higher short-term profits) to commit themselves to higher monitoring (i.e., $x^*$ increases in the liquidation cost; see Corollary 7.1). Consequently, the total amount of bank lending contracts.

### 7.3.3 Short-term borrowing: Multiple investors

Now I extend the previous analysis to allow for *multiple* investors. At the refinancing stage, at $t = 1$, the bank may again borrow short-term, in this case, from multiple investors. More specifically, investors sequentially decide whether they lend their funds to the bank at the rates offered.

With multiple investors the following problem occurs. Now the bank can neither commit to equal rates to all investors, nor can the bank commit to limit the amount of short-term borrowing. Proposition 7.2 shows that with a single investor the bank partially refinances but must still liquidate $x^*$ borrowers and this commits it to monitoring. A single investor would anticipate higher risk and demand higher return for all its investment if the bank wants to refinance more and liquidate fewer than $x^*$ borrowers. With multiple investors, however, the bank may deviate and refinance more and liquidate fewer than $x^*$ borrowers. Additional investors anticipate that the bank stops monitoring and demand a higher rate to be compensated for additional risk; however, the initial investors can no longer change their rates and may now be worse off.

The profit of the bank that refinances is as follows. It stops monitoring; hence, the expected borrowers' return is $pR$. Initial $\{1-x[R-c_L]\}$ investors anticipated that the bank would monitor. They demanded $r_{S2} = 1$. However, the bank deviates to no monitoring and the expected payment to the initial investors is $p\{1-x[R-c_L]\}$. The bank also refinances at additional $x[R-c_L]$ investors. They anticipate no monitoring and demand $r_{S2} = \frac{1}{p}$. Hence, the expected payment to additional investors is $x[R-c_L]$. Summarizing, the bank that refinances now earns

$$\Pi_R(x) = pR - p\{1 - x[R - c_L]\} - x[R - c_L]. \tag{7.16}$$

Rearranging (7.16) yields

$$\Pi_R(x) = p[R - 1] - pxc_L + pxR - xR + xc_L. \tag{7.17}$$

The bank monitors its borrowers if its profit conditional on monitoring in (7.9) is larger than that with refinancing and without monitoring in (7.17). Rearranging yields the following

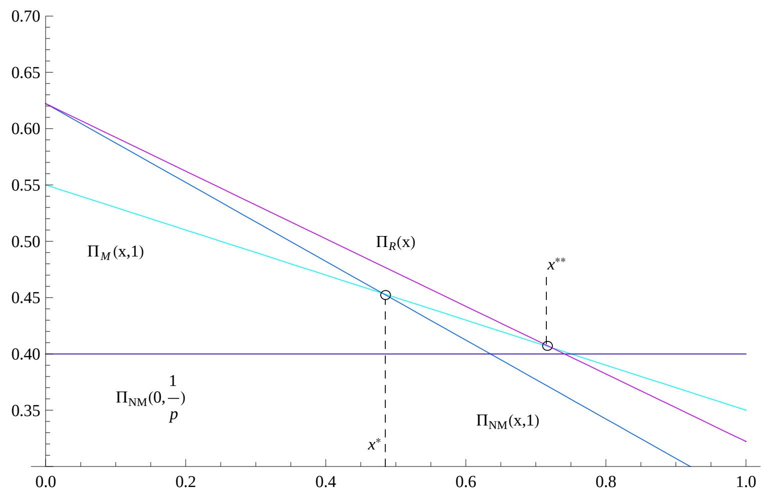


Figure 7.3: The bank's profits (if $c_L > R[1-p]$).

condition.

$$R - 1 - \frac{c_M}{1-p}[1-x] - c_L x - \frac{x[R - c_L - pR]}{1-p} \geq 0, \tag{7.18}$$

The profit of the bank if it commits to monitoring is

$$\Pi_M(x^*, 1) \quad = \quad \max_x R - 1 - c_M[1-x] - c_L x, \tag{7.19}$$

$$\text{s.t.}$$

$$R - 1 - \frac{c_M}{1-p}[1-x] - c_L x \geq 0, \tag{7.20}$$

$$R - 1 - \frac{c_M}{1-p}[1-x] - c_L x - \frac{x[R - c_L - pR]}{1-p} \geq 0. \tag{7.21}$$

That is, the bank maximizes its profit in (7.19) conditional on two constraints. First, the bank must have incentives to monitor at the promised liquidation level $x$; see (7.20). Second, the bank must have no incentives to refinance and then stop monitoring; see (7.21).

Figure 7.3 graphically depicts the maximization problem of the bank. The bank decides to monitor its borrowers if the solution to (7.19) is higher than (7.11); that is, if $\Pi_M(x,1) > \Pi_{NM}(0, \frac{1}{p})$. In addition, the bank must have incentives to monitor at liquidation level $x$; that is, $\Pi_M(x,1) > \Pi_{NM}(x,1)$. Furthermore, the bank must have no incentive to refinance; that is, $\Pi_M(x,1) > \Pi_R(x)$.

I can now show the following result.

**Proposition 7.3.** *With multiple investors and short-term borrowing, one has:*

- *For a low monitoring cost $c_M \leq \bar{c}_M$, investors fully refinance the bank at $r_{S2} = 1$. The bank liquidates no borrowers, but monitors them.*

- *For a high monitoring cost $c_M > \bar{c}_M$ and for a low liquidation cost $c_L < R[1-p]$, the investors only partially refinance the bank at $r_{S2} = 1$. The bank liquidates $x^*$ of borrowers and monitors the others.*

- *For a high monitoring cost $c_M > \bar{c}_M$ and for an intermediate liquidation cost $R[1-p] \leq c_L \leq \hat{c}_L(c_M)$, where $R[1-p] \leq \hat{c}_L(c_M) < \bar{c}_L(c_M)$, investors only partially refinance the bank at $r_{S2} = 1$. The bank liquidates $x^{**}$ (where $x^{**} > x^*$) of borrowers and monitors the others.*

- *For a high monitoring cost $c_M > \hat{c}_M$ and for a high liquidation cost $c_L > \hat{c}_L(c_M)$, investors fully refinance the bank at $r_{S2} = \frac{1}{p}$. The bank does not monitor.*

The intuition for Proposition 7.3 is the following. With many investors, the bank's ability to commit to monitoring crucially depends on the liquidation costs of the bank's borrowers. For a low liquidation cost, the bank can commit to monitoring as easily as in Proposition 7.2. However, for an intermediate liquidation cost (i.e., for $R[1-p] \leq c_L \leq \hat{c}_L$), the bank is confronted with the additional refinancing option. In particular, the bank may avoid liquidation by refinancing more at other investors. Investors anticipate this refinancing problem and are only willing to fund the bank up to the lower level of indebtedness. That is, the bank is forced to liquidate more borrowers than in the case of a single investor (i.e., $x^{**} > x^*$ borrowers). Because liquidation is costly, the bank is worse off in this region compared to the situation with a single investor. For a high liquidation cost $c_L > \hat{c}_L$, the bank can no longer commit to monitoring in the case of multiple investors, although this may be possible with a single investor (i.e., in the region $\hat{c}_L < c_L < \bar{c}_L$; see Proposition 7.2).

### 7.3.4 Demand deposits

Now I show that demand deposits contain the refinancing problem. With demand deposits, the timing of the model is the following. At $t = 1$, the bank may first refinance, depositors then sequentially decide whether to withdraw or not, and the bank liquidates borrowers if necessary.

I can show the following result.

**Proposition 7.4.** *In the case of multiple investors with demand deposits, Proposition 7.2 still applies.*

Interestingly, offering demand deposits resolves the refinancing problem. That is, banks with demand deposits can no longer abuse investors and refinance at $t = 1$. In particular, if banks refinance, investors anticipate no monitoring and withdraw their demand deposits unless they are compensated for additional risk. The threat of withdrawals of demand deposits disciplines a bank's choice of the level of debt. Interestingly, even if the bank raises additional debt and renege on liquidation, the subsequent withdrawals reduce bank debt and, via forced liquidation, induce the bank to monitor again.

Demand deposits are important for a high monitoring cost $c_M > \bar{c}_M$ and for an intermediate liquidation cost $R[1-p] \leq c_L < \bar{c}_L(c_M)$. In this region, demand deposits contain the refinancing problem that occurs with straight short-term borrowing from multiple investors (compare Proposition 7.2 with Proposition 7.3).

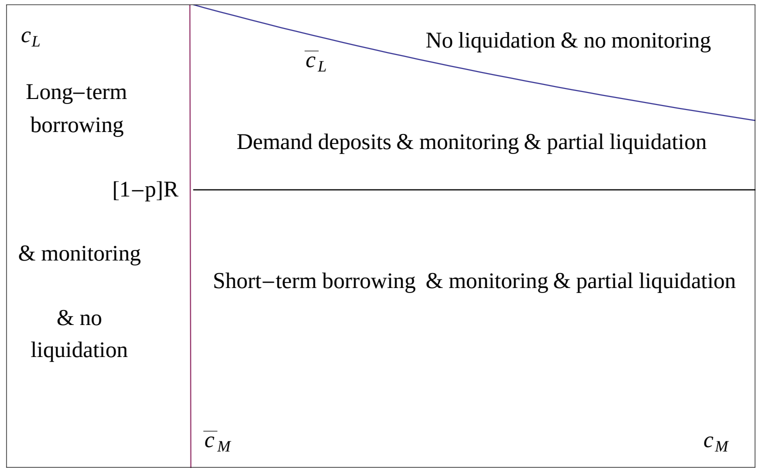


Figure 7.4: Different regions of bank financing.

Figure 7.4 combines the predictions of Proposition 7.1, Proposition 7.2, Proposition 7.3 and Proposition 7.4 to show that different regions of bank financing exist. Only in the first region, in which a monitoring cost is low $c_M < \bar{c}_M$, can banks borrow entirely long-term and commit to monitoring. For a higher cost of monitoring (i.e., for $c_M \geq \bar{c}_M$), three additional regions exist. For a low liquidation cost (i.e., if $c_L < R[1-p]$), the bank can borrow short-term. For an intermediate liquidation cost (i.e., if $R[1-p] \leq c_L \leq \bar{c}_L(c_M)$), the bank raises demand deposits. For a high liquidation cost $c_L > \bar{c}_L(c_M)$, the bank cannot commit to monitoring and liquidates no borrower.

The following corollary shows that the bank may combine demand deposits with long-term or short-term borrowing.

**Corollary 7.2.** *For $c_M > \bar{c}_M$ and for $R[1-p] < c_L < \hat{c}_L$, the bank can combine demand deposits with (not more than $1 - x^{**}[R - c_L]$) short-term or long-term borrowing, liquidates $x^*$ of borrowers, and commits to monitoring.*

The intuition for this corollary is the following. The bank can only borrow long-term and short-term to the extent that it can still successfully commit to monitoring. First, the bank has to commit to liquidate at least $x^*$ borrowers. Note that investors with demand deposits force the bank to liquidate at least $x^*$ borrowers (the bank borrows short-term or long-term $1 - x^{**}[R - c_L]$, which is less than $1 - x^*[R - c_L]$). That is, investors with demand deposits withdraw such that the bank is forced to liquidate $x^*$ of borrowers and this commits the bank to monitoring.

Moreover, with the combination of demand deposits and short-term (or long-term) borrowing as given in Corollary 7.2, the bank cannot refinance at $t = 1$ to prevent liquidation of its borrowers. To see this, note the following. Proposition 7.3 shows that refinancing is not possible if the bank liquidates $x^{**}$ borrowers; that is, the bank can borrow $1 - x^{**}[R - c_L]$ short-term or long-term and still has no incentives to refinance. However, at this stage (i.e., at $1 - x^{**}[R - c_L]$ short-term or long-term debt), the bank may additionally raise demand deposits. This is because investors with demand deposits can defend against bank refinancing (see Proposition 7.4); that is, they can withdraw if the bank tries to refinance.

## 7.4 Extensions and Empirical Predictions

This section first extends the analysis to the situation of different liquidation costs of bank borrowers. Second, it presents empirical predictions of the analysis.

### 7.4.1 Borrowers with different liquidation costs

Now I allow for different liquidation costs of borrowers. In particular, I assume that in addition to borrowers with liquidation cost $c_L$ there also exist otherwise identical borrowers but with liquidation cost $R$. Hence, liquidation of such borrowers yields zero profit (i.e., $R - c_L = 0$). Such borrowers may have no collateral that the bank can seize in the case of early liquidation; thus, the total return in the case of liquidation is zero.

The above analysis predicts that the bank cannot lend only to such borrowers. In particular, Proposition 7.2 shows that, for a high monitoring cost $c_M > \bar{c}_M(c_M)$, banks can only lend to borrowers with a sufficiently low liquidation cost $c_L < \bar{c}_L(c_M)$. Liquidation of borrowers with low liquidation costs allows the bank to earn (small) profits and this commits it to monitoring in the second period. This shows that borrowers with a high liquidation cost $c_L > \bar{c}_L(c_M)$ and a high monitoring cost $c_M > \bar{c}_M$ could not obtain financing directly from investors on the financial markets.

Interestingly, if the bank combines lending to borrowers with different liquidation costs, it can still commit to monitoring, as the following Proposition shows.

**Proposition 7.5.** *The bank may lend to $x^*$ of borrowers with a liquidation cost $c_L$ and to $1 - x^*$ of borrowers with a liquidation cost $R$.*

The intuition for Proposition 7.5 is the following. The bank needs to liquidate $x^*$ of borrowers to commit to monitoring (see Proposition 7.2). However, the bank does not liquidate other borrowers. Hence, the liquidation costs of other borrowers do not matter.

Proposition 7.5 connects this analysis to Jacklin and Bhattacharya (1988). They show that banks may invest partially in liquid and illiquid projects and, in doing so, optimally serve the liquidity needs of depositors. My analysis shows that investing in liquid projects helps the bank commit to monitoring. More specifically, depositors withdraw and force the bank to liquidate its liquid projects. This limits the bank's indebtedness and commits the bank to monitoring in the later period. In this way, the bank can also finance illiquid projects that are not viable on a stand-alone basis.

This brief extension shows that banks may combine lending to borrowers with different liquidation costs. Lending to several borrowers also gives the bank the ability to monitor borrowers for which commitment to monitoring is difficult.

### 7.4.2 Implications and empirical predictions

Now I list the empirical/stylized facts that are consistent with the predictions following from my analysis.

1. Proposition 7.3 predicts that banks may borrow long-term for a low monitoring cost. For a higher monitoring cost, banks may borrow short-term if they lend to borrowers with low liquidation costs. Banks that lend to borrowers with intermediate liquidation costs will predominately use demand deposits.

2. This analysis highlights some drawbacks of narrow banking proposals. Narrow banking proposals call for the separation of the two core activities of the bank such that demand deposits should be invested in liquid securities whereas illiquid loans should be financed with noncheckable long-term liabilities (see Bryan (1988)). Proposition 7.3 and Proposition 7.4 show that the separation of demand deposits and lending may diminish bank monitoring for a high monitoring cost $c_M > \bar{c}_M$ and for an intermediate liquidation cost $R[1-p] < c_L < \bar{c}_L$. This yields the prediction that narrow banking proposals may have the worst effect in developing economies, where the cost of monitoring is high. In contrast, the separation of lending and deposit taking may not have a negative effect in developed economies if the cost of monitoring is low.

3. Calomiris (1999) and Calomiris and Litan (2000) argue that market forces could help fine tune bank regulation. They propose that banks above a certain size threshold should be mandated to finance a certain portion of their assets by long-term, uninsured subordinated debt. Corollary 7.2 shows that the mandated long-term debt should not be too high, otherwise banks may become risky exactly because of this regulatory measure.

4. Corollary 7.1 shows that banks respond to a higher monitoring cost and to a higher liquidation cost of their borrowers by tightening their lending. This is related to Dell'Ariccia and Marquez (2006), who show that banks may loosen their credit standards if information asymmetries decrease.

## 7.5 Conclusions

This chapter presents a novel rationale for why banks combine demand deposits with lending to borrowers that need monitoring. It shows that a demand-deposit contract may serve as a commitment device. In particular, withdrawals on demand-deposit contracts could force the bank to liquidate a proportion of its borrowers. Liquidation yields (small) profits that can be used as equity in the second period. Partial equity financing alleviates moral hazard and commits the bank to monitoring.

I also show that demand deposits resolve a refinancing problem that appears with short-term borrowing. More specifically, if a bank borrows straight short-term debt from multiple investors, they can no longer control the level of bank debt: the bank may refinance from other investors rather than liquidate borrowers. Demand deposits resolve this refinancing problem. If the bank reneges on the promised level of liquidation, investors anticipate higher risk. They withdraw their demand deposits and, via forced liquidation, induce the bank to monitor again. This points to the novel role of demand deposits: demand deposits balance bank debt to the level where the bank can still commit to monitoring.

## 7.6 Appendix

*Proof of Proposition 7.1*
For $c_M \leq \bar{c}_M$, the condition in (7.5) is satisfied. In addition, use (7.2) and (7.4) to see that $\Pi_{NM}(0, \frac{1}{p}) < \Pi_M(0, 1)$. ■

*Proof of Proposition 7.2*
First, note that as long as $c_M \leq \bar{c}_M$, where $\bar{c}_M$ is defined in (7.6), the condition in (7.15) is satisfied for $x = 0$. Hence, the bank can commit to monitoring even without liquidating borrowers.

Now I analyze the situation in which $c_M > \bar{c}_M$. In this case, the investor only refinances the bank up to a certain degree to commit the bank to monitoring. However, liquidation is costly for the bank. To see this, differentiate (7.14) w.r.t. $x$ to obtain

$$\frac{\partial \Pi_M(x, 1)}{\partial x} = R - c_L - [R - c_M],$$

which is negative because of the condition in (7.1). Hence, $x$ has to be as low as possible; that is, the incentive constraint in (7.15) is satisfied with equality. Solving (7.15) for $x$, one obtains

$$x^* = \frac{c_M - [R-1][1-p]}{c_M - c_L[1-p]}. \tag{7.22}$$

If the bank liquidates $x^*$ borrowers, it can commit to monitoring. The question is whether the bank wants to commit to monitoring; that is, whether (7.14) is greater than (7.11). Insert (7.22) into (7.14) to obtain

$$\Pi_M^* = \frac{R - 1 - c_L}{c_M - c_L[1-p]}[1-p][R - c_M] - \{1 - \frac{c_M - [R-1][1-p]}{c_M - c_L[1-p]}[R - c_L]\}.$$

Rearranging yields

$$\Pi_M^* = pc_M \frac{R - 1 - c_L}{c_M - c_L[1-p]}. \tag{7.23}$$

This is lower than (7.11) as long as

$$pc_M \frac{R - 1 - c_L}{c_M - c_L[1-p]} < pR - 1. \tag{7.24}$$

This condition is identical to $c_L > \bar{c}_L(c_M)$, where

$$\bar{c}_L(c_M) \equiv \frac{c_M[1-p]}{pc_M - [1-p][pR-1]}. \tag{7.25}$$

Observe that, for $c_M > \bar{c}_M$, one has $pc_M > [1-p][pR-1]$. Hence, (7.25) is positive.

Observe also that, for $c_M > \bar{c}_M$ and $c_L > \bar{c}_L(c_M)$, the bank does not commit to monitoring because (7.11) is greater than (7.23). The investor anticipates that the bank does not monitor. Hence, he only provides funding if the bank offers at least $r_{S2} = \frac{1}{p}$. ■

*Proof of Corollary 7.1*
Rewrite (7.22) as

$$x^* = 1 - [1-p]\frac{R-1-c_L}{c_M - c_L[1-p]}. \tag{7.26}$$

For $c_L \leq \bar{c}_L(c_M)$ and $c_M > \bar{c}_M$, one has $c_L \leq \frac{c_M}{1-p[R-\frac{c_M}{1-p}]}$ and $\frac{c_M}{1-p[R-\frac{c_M}{1-p}]} < \frac{c_M}{1-p}$. Hence, $c_L < \frac{c_M}{1-p}$. Hence, (7.26) is positive. Consequently, $\frac{\partial x^*}{\partial c_M} > 0$. ■

*Proof of Proposition 7.3*
For $x = 0$, the conditions in (7.20) and (7.21) are simultaneously satisfied if $c_M \leq \bar{c}_M$, where $\bar{c}_M$ is as defined in (7.6). In this case, the bank can commit to monitoring even without liquidation. Hence, $r_{S2} = 1$.

Now I analyze the situation in which $c_M > \bar{c}_M$. If $c_L < R[1-p]$, the constraint in (7.20) is more binding than (7.21). Hence, (7.21) can be dismissed. In this case, Proposition 7.2 holds. That is, the bank liquidates $x^*$ borrowers and commits to monitoring, and investors accept $r_{S2} = 1$.

If $c_L \geq R[1-p]$, the constraint in (7.21) is more binding than (7.20). Hence, (7.20) can be dismissed. Because liquidation is costly for the bank, $x$ must be such that the incentive constraint in (7.21) is satisfied with equality. Solving for $x$ yields

$$x^{**} = \frac{[R-1][1-p] - c_M}{c_L[1-p] - c_M - R[1-p] + c_L}. \tag{7.27}$$

Compare with (7.22) to see that $x^{**} > x^*$. If the bank liquidates $x^{**}$ borrowers, it can commit to monitoring. The question is whether the bank wants to commit to monitoring; that is, whether (7.14) is greater than (7.11). This is true if

$$[1-x^{**}][p_H R - c_M] - [1 - x^{**}R_{liq}] > p_L R - 1. \tag{7.28}$$

Inserting (7.27) into (7.28) yields the condition

$$c_L < \hat{c}_L \text{ where } \hat{c}_L(c_M) = \frac{c_M[R-1] - R^2[1-p]}{c_M - R - 1 + pR}. \tag{7.29}$$

For $c_M > \bar{c}_M$ and for $c_L > \hat{c}_L(c_M)$, investors anticipate that the bank refinances and does not monitor. Hence, they only provide funding if the bank offers at least $r_{S2} = \frac{1}{p}$. ■

*Proof of Proposition 7.4*
If the bank refinances, investors with demand deposits withdraw if they are not compensated for higher risk. That is, refinancing as in (7.17) is no longer possible. Hence, the bank solves the optimization problem as put forth in (7.14) and Proposition 7.2 holds. ■

*Proof of Corollary 7.2*
Note that the only difference between Proposition 7.2 and Proposition 7.3 occurs in the region $\hat{c}_L(c_M) \leq c_L \leq \bar{c}_L(c_M)$. The situation with demand deposits is the following. If the

bank starts refinancing, demand depositors withdraw (if not offered $\frac{1}{p}$), to such extent that $x^*$ borrowers are liquidated. In this case, Proposition 7.2 shows that the bank can commit to monitoring. Hence, the bank can only refinance and stop liquidating $x^*$ borrowers if it offers $\frac{1}{p}$ to all depositors. However, Proposition 7.2 shows that this is not profitable. ■

*Proof of Proposition 7.5*
Note that Proposition 7.2 shows that the bank can commit to monitoring if it liquidates $x^*$ borrowers and continue lending to $1 - x^*$ borrowers. Hence, it suffices that the bank lends to $x^*$ borrowers with low liquidation costs $c_L$ and can lend to $1 - x^*$ borrowers with high liquidation costs $R$. At $t = 1$, the bank only liquidates the borrowers with low liquidation costs and continue lending to the borrowers with high liquidation costs. ■

# Chapter 8

# The Maturity of Monitored Finance: Covenants and Insufficient Liquidation

**Abstract**

The characteristics of an optimal lending contract in a model where a bank obtains proprietary information by monitoring its borrowers are analyzed. A key to the analysis is that, with a short-term contract, the incumbent bank's rent-seeking behavior will trigger competitive offers such that bad borrowers can also obtain financing and liquidation is insufficient. This happens if competition for borrowers is high and competition between banks is based on *simultaneous* offers. Using a long-term contract with covenants leads to liquidation of more bad borrowers. The chapter shows that lower-risk borrowers use short-term contracts while higher-risk borrowers finance themselves with long-term contracts with covenants. As competition for borrowers increases and/or is expected to increase in the future, long-term contracts with covenants become more attractive to borrowers. Moreover, the use of covenants and the contracts maturity are positively correlated with monitoring precision.

Keywords: Loan Maturity, Bank Monitoring, Covenants, Competition
JEL CLASSIFICATION: G21

## 8.1 Introduction

This chapter looks into how competition affects the effectiveness of different contractual arrangements between a bank and its borrowers.

The issue is analyzed in a model in which the bank's main role is to provide for optimal liquidation of (bad) projects. After granting credit, monitoring provides the credit-granting bank with proprietary signals about its borrowers' qualities. The (incumbent) bank rejects financing bad borrowers (the ones with negative signals) and tries to earn rents on good borrowers. Another (competing) bank may appear and compete for borrowers from the incumbent bank. However, it is at a disadvantage due to not having signals about the borrowers' qualities.

In my model, borrowers initially choose either short-term contracts or long-term contracts (with or without covenants) and they seek offers. Banks make loan offers for the chosen contracts. Subsequently, the incumbent bank obtains proprietary information about the borrowers' qualities. Borrowers that choose short-term contracts again search for competing banks. If borrowers do not obtain further financing, their projects are liquidated; otherwise they continue with their projects.

A key to the analysis is that with a long-term contract a bank cannot exploit good borrowers (meaning the ones for which it has received *good* signals). The only borrowers that seek funding from competing banks are bad borrowers. All other banks realize this and thus abstain from funding. Hence, liquidation is efficient. With a short-term contract the rent-seeking behavior of the incumbent bank will have triggered competitive offers such that bad borrowers can also obtain financing and liquidation is insufficient. Insufficient liquidation worsens the lending terms of a short-term contract. This happens when banks grant offers simultaneously.

The reason a short-term contract induces insufficient liquidation deserves further discussion. Due to rent-seeking behavior of the incumbent bank facilitated by a short-term contract, the competing bank might be willing to lend and try to attract good borrowers. However, the competing bank may also end up financing some of the rejected borrowers. More specifically, the incumbent bank obtains proprietary monitoring signals and it uses this informational advantage to take rents from good borrowers and reject bad borrowers. The competing bank is unable to distinguish between borrowers but it may still compete for them. This is because the competing bank may attract good borrowers that the incumbent bank attempts to exploit excessively. This compensates the competing bank for the possibility that it finances rejected (i.e., bad) borrowers, but may also prevent optimal liquidation.

The drawback of a short-term contract highlighted here, insufficient liquidation, is different from the ones in Rajan (1992), Sharpe (1990), and Von Thadden (1995). They employ similar types of models where private information gives bargaining power to the incumbent bank. In their studies, however, this creates inefficiencies on the borrowers' side. In particular, borrowers that anticipate lower profits either invest inefficiently (see Rajan (1992)), put too little effort into their projects (see Sharpe (1990)) or may invest too much in short-term projects and too little in long-term projects (see Von Thadden (1995)).

Berglof and Von Thadden (1994) and Bolton and Scharfstein (1996) argue that renegotiation makes liquidation less efficient. They show that a financial structure with multiple investors (with different maturity of their claims) may restore efficient liquidation. In this case, insufficient liquidation is not the consequence of renegotiation. Rather, increased competition prevents the full use of private information obtained through bank monitoring.

This analysis is related to studies that examine the impact of competition on the efficiency of bank monitoring. Petersen and Rajan (1995) argue that competition may hamper the intensity of bank-borrower relationships and lead to reduced availability of credit. Hauswald and Marquez (2006) combine the hold-up problem with spatial competition. They show that, as competition increases, banks invest less in monitoring and therefore make less efficient lending decisions. In contrast, Boot and Thakor (2000) argue that competition may make bank monitoring even more important. They show that relationship lending helps insulate banks from pure price competition. In their view, banks move to relationship lending as competition increases. My analysis shows that competition may prevent the full use of private information obtained through monitoring.[1]

This analysis still brings a positive view to the relation between competition and monitored finance, however. In particular, banks can largely circumvent the detrimental effect of competition on the efficiency of monitoring (and liquidation of bad borrowers) by carefully designing their contractual relationships with borrowers. The analysis shows that it may be optimal to offer a long-term contract with a termination clause; that is, a long-term contract coupled with covenants. This gives the bank the power to terminate the contract. However, if the bank does not terminate the contract, it has to uphold the previously arranged contractual terms. Such contracts prevent the incumbent bank with proprietary monitoring signals from engaging in rent-seeking behavior. Consequently, the only borrowers that try to find a competing bank are rejected borrowers. Anticipating this, a competing bank is not willing to provide financing and liquidation is efficient. Von Thadden (1995) similarly proposes that short-term contract deficiencies may be diminished with a carefully designed long-term contract. In his model, an example of such contract may be a debt contract coupled with a credit line arrangement.

Rajan and Winton (1995) analyze bank monitoring under the mix of bank debt and public debt. Public debt holders may free-ride on bank monitoring, which makes monitoring less valuable. Rajan and Winton (1995) show that including a covenant improves a bank's incentives to monitor the borrower. In my case, covenants are valuable because they give the bank the right to act upon receiving negative signals and terminate the bad borrowers. In this respect a short-term contract is less efficient because uninformed banks may continue to finance an already-rejected bad borrower.

[1] A study by Dell'Ariccia and Marquez (2006) is also related. They show that as information asymmetries decrease, banks may loosen their credit standards. I show that an increase in competition may lead to less efficient liquidation of borrowers although the bank credit standards remain unchanged.

While this study focuses on the relationship between a bank and its borrowers, this analysis is applicable to other contractual relationships. Aghion and Bolton (1987) show that customers are willing to sign a long-term contract that limits their future bargaining power. They argue that a long-term contract can be used to extract efficiency rents from future entrants and partially pass those rents on to customers. They show that long-term contracts would optimally control their duration by including costly termination clauses. My analysis points to a different reason why a long-term contract with a termination clause (i.e., a covenant) may be optimal: it provides for an efficient use of proprietary information obtained during the contractual relationship.

While this analysis is limited to debt contracts, Bolton and Scharfstein (1990) and Hart and Moore (1998) set their analyses more broadly. In the incomplete contract framework, they show that debt contracts are optimal for a wide range of conditions because they prevent managers from taking unobservable cash flows.[2] In Hart and Moore (1998), the bank may excessively liquidate its borrowers if they cannot commit to repay late. In this sense, long-term debt protects borrowers from inefficient liquidation of the bank (see also Hart and Moore (1994)). In my analysis, long-term contracts with covenants allow for efficient liquidation but protect borrowers from rent-seeking behavior by their bank.

I also analyze the strategic considerations of how borrowers optimally select the contracts type. A long-term contract with covenants is socially optimal because it provides for optimal liquidation of bad projects. Nevertheless, borrowers may prefer a short-term contract for strategic reasons. In particular, a short-term contract allows borrowers to search for a competing bank in the future, giving them additional bargaining power, while a long-term contract permanently binds them to the incumbent bank. This then translates into the trade-off between short-term contracts and long-term contracts with covenants explored here.

Flannery (1986) predicts a positive relation between the maturity of a debt contract and a borrower's risk. He argues that good borrowers use short-term debt contracts so that the bank can learn their borrower type during the first period. Bad borrowers use long-term contracts to spare the transaction costs of writing the contract.[3] This analysis also predicts that, on average, higher-risk borrowers opt for long-term contracts. However, the intuition for this is different from Flannery's (1986).

In my model, borrowers do not know their type. The choice between contracts depends on the publicly known proportions of good versus bad borrowers in the economy. With a high proportion of bad borrowers, long-term contracts with covenants are preferred while

[2]Bolton and Scharfstein (1990) show how debt financing affects product market competition through deterring liquidation, while this analysis focuses on the impact of competition in lending markets on the maturity of debt.

[3]Diamond (1991b), however, involves another effect. He argues that much higher-risk borrowers may be refused the option of long-term debt. In his view, highly risky and safe borrowers finance with short-term debt and borrowers of medium risk finance with long-term debt. Recently, Berger, Espinosa-Vega, Frame, and Miller (2005) have presented evidence for a positive relation between debt maturity and the borrower's risk as suggested by Flannery (1986).

short-term contracts might be preferred if this proportion is relatively low. This will depend on the competitiveness of the banking system at $t = 0$. This is because efficient liquidation is especially important if the proportion of bad projects is high. For lower-risk borrowers, liquidation efficiency loses its importance. Borrowers then choose a short-term contract to gain bargaining power at $t = 1$ if the banking system is relatively non-competitive at $t = 0$. Comparative statics show that high-quality banks (with high monitoring precision) more frequently use long-term contracts with covenants. In contrast, low-quality banks predominantly stick to short-term contracts.

My analysis yields some novel predictions regarding the relation between competition and the maturity of contracts. Interestingly, as competition for borrowers increases, borrowers opt more frequently for long-term contracts with covenants. This is for two reasons. First, borrowers anticipate that greater competition eliminates banks' rents, hence they opt for more efficient long-term contracts with covenants. Second, greater expected competition in the future exacerbates the problem of insufficient liquidations that short-term contracts bring.

The chapter is organized as follows. The model is developed in Section 8.2. Section 8.3 analyzes the different types of contracts. The borrower's choice of contract is analyzed in Section 8.4, Section 8.5 allows for costly covenants and contains the empirical predictions. Section 8.6 is the conclusion.

## 8.2 Model Specifications

It is assumed throughout that the risk free rate is zero. Both types of agents in the model – banks and borrowers – are risk neutral.

*Borrowers:* At $t = 0$, each borrower needs \$1 of initial investment to finance his project. The parameter $\gamma$ describes the proportion of borrowers with good projects. With probability $\gamma$ a borrower has a good project which yields a sure return $R$ at $t = 2$, where it is ex-ante efficient to grant a loan; i.e., $R > \frac{1}{\gamma}$. With probability $[1 - \gamma]$ a borrower has a bad project. A bad project always fails and yields zero return. Borrowers do not know their project types. Borrowers can choose to liquidate the projects at $t = 1$; each project then yields \$1 (regardless of the type of a borrower).

*Financing options:* Borrowers approach a bank for financing. All contracts are debt contracts.[4] Borrowers choose either a short-term contract (for one period, i.e., from $t = 0$ to $t = 1$, possibly followed by another short-term loan, i.e., from $t = 1$ to $t = 2$) or a long-term contract; that is, for two periods from $t = 0$ to $t = 2$. A long-term contract does not require intermediate payments.

*Bank monitoring:* The bank has monitoring technology in place that, just prior to $t = 1$ (just prior to the termination of a short-term contract), gives the bank costless signals $S \in \{B, G\}$ which noisily reveal the qualities of the borrowers. The quality of the bank

[4] Townsend (1979) finds that debt contracts are optimal using the costly state verification approach.

results in the precision of its signals $\Phi$. If a borrower has a bad project, the bank always obtains a negative signal $S = B$. If a borrower has a good project, the bank obtains a positive signal $S = G$ with probability $\Phi$ and a negative signal $S = B$ with probability $[1 - \Phi]$.

In the case of a one-period loan the bank can condition its renewal decisions at $t = 1$ on the signals received. Note that only the initial credit-granting bank (hereafter: the incumbent bank) receives the signals. Thus, it has an informational advantage at $t = 1$ compared to competing banks. With a two-period loan the incumbent bank cannot terminate the loan at $t = 1$, unless a covenant is included in the contract that allows it to do so. The covenant (if included) stipulates that the bank is allowed to call the loan at $t = 1$. This can be considered a MAC (material adverse change) clause, as often accounted for in bank loan contracts. It is assumed that inclusion of such a covenant gives the incumbent bank the right to call the loan.[5] However, if the loan is not called, the incumbent bank has to continue financing at the previously agreed conditions.

*Competition and information structure:* At $t = 0$, borrowers choose the contract type. With probability $q_0$, borrowers can choose between competing banks. With probability $[1 - q_0]$, they only have one bank to choose from.[6] When there is only one bank to choose from, borrowers have no choice but to accept a monopolistic offer, provided this yields positive returns. When banks compete, banks compete for borrowers as Bertrand competitors.

At $t = 1$, in case of a one-period contract and/or a two-period contract that has been terminated by the incumbent (initial) bank at $t = 1$, borrowers search for offers. They succeed in locating a competing bank with probability $q_1$. With probability $[1 - q_1]$, they do not find a competing bank and can only choose the offer (if any) from the incumbent bank. If a competing bank is found, both the incumbent and competing bank simultaneously decide whether they give offers to the borrowers, and if so, under what conditions. Borrowers then choose the offer that yields the highest profits. If borrowers obtain no offers for further financing, their projects are liquidated.

*Time line:* At $t = 0$, borrowers choose the contract type. Subsequently, they search for offers. The bank makes loan offers to borrowers for the chosen contracts. Just prior to $t = 1$, the incumbent (initial) bank obtains proprietary information about each borrower's quality. Borrowers again search for competing banks. The incumbent and competing bank (if one is found) may then *simultaneously* give second-period offers to borrowers. If borrowers do not obtain further financing, their projects are liquidated. If borrowers do obtain further financing, they continue with their projects. The outcomes of the projects are realized at

[5]This might be because it is costly for borrowers to dispute a bank's loan termination decision. Also, empirical evidence seems to confirm that covenants are set extremely tightly. Chava and Roberts (2006) show that approximately 15-20% of outstanding loans are in violation during a typical quarter and, conditional on violating a covenant, a loan is delinquent about 40% of the time.

[6]The competition parameter $q_0$ (and $q_1$) can be interpreted in terms of a spatial model as the level of transportation costs. In particular, a high level of transportation costs may make the banking system fragmented, such that each bank behaves as a monopolist in a local market and borrowers are left with only one offer (in our model $q_0$ is low). However, as transportation costs decrease, borrowers can reach the competing bank (in our model $q_0$ is high). See Thisse and Vives (1988).

$t = 0$:

♠ The borrower has \$1 of funding needs.

♠ The borrower chooses either a short-term contract or a long-term contract (with or without covenants).

♠ A competing bank materializes with probability $q_0$.

♠ The borrower undertakes the project.

$t = 1$:

♠ The incumbent bank receives a signal about the quality of its borrower.

♠ The competing bank materializes with probability $q_1$.

♠ In the case of a short-term contract, the incumbent bank and competing bank simultaneously bid for a borrower.

♠ In the case of a long-term contract with covenants, the incumbent (initial) bank can terminate the contract upon receiving a negative signal.

♠ If the borrower has further financing, he continues his project. Otherwise his project is liquidated.

$t = 2$:

♠ Payoffs are realized.

Figure 8.1: Timeline

$t = 2$. Figure 8.1 depicts the sequence of events.

## 8.3 Analysis

This section presents analyzes of the different types of contracts. Long-term contracts with and/or without covenants are analyzed first. Short-term contracts are analyzed second. The model is solved by backward induction.

### 8.3.1 Long-term contract

First the borrowers' expected profits if they choose a long-term contract without covenants are analyzed. Second, the profitability of a long-term contract with covenants is considered.

A long-term contract without covenants is defined as a contract *without* a termination clause; hence, it does not allow the incumbent bank to terminate the financing at $t = 1$, even if it obtains a negative signal.[7] Consequently, borrowers can always hold on to financing once obtained at $t = 0$. The situation at $t = 0$ is the following: the total expected profits of the projects are $\gamma R - 1$. The borrowers either have offers from one bank (w.p. $1 - q_0$) or from two (w.p. $q_0$). In the case of two banks competing, borrowers seize all the rents. With probability $1 - q_0$, there is only one bank. This bank can seize all the profits. Hence, upon choosing a long-term contract without covenants the expected values of the projects for the

[7] Potential renegotiation between the bank and borrowers at $t = 1$ is not considered here. In particular, if borrowers can seize the entire proceeds from renegotiation, the bank will not monitor them (if monitoring comes at a small cost for a bank). This then prevents renegotiation. For more on renegotiation, see Rajan (1992) and Berlin and Mester (1992).

borrowers are

$$V_{LT} = q_0[\gamma R - 1]. \tag{8.1}$$

Now I focus on a long-term contract with covenants, starting with an analysis of the situation at $t = 1$. Covenants give power to the incumbent (initial) bank to terminate the financing if the monitoring signal is negative.[8] The question remains whether the incumbent bank is willing to do so. Conditional on receiving a negative signal, the profitability of lending is as follows: the incumbent bank receives a negative signal if a borrower is bad. This happens with probability $1-\gamma$. However, the bank might also receive a negative signal if the borrower has a good project. The probability of this event is $\gamma[1 - \Phi]$. Hence, the success probability of the borrowers for which the bank receives negative signals is

$$p(S = B) = \frac{\gamma[1 - \Phi]}{\gamma[1 - \Phi] + 1 - \gamma}. \tag{8.2}$$

The highest return that the bank can demand from borrowers is $R$. Hence, lending to the borrowers with negative signals is not profitable if

$$p(S = B)R - 1 < 0. \tag{8.3}$$

It is assumed that borrowers are liquidated conditional on negative signals. This amounts to assuming that monitoring is sufficiently precise; that is, if the following assumption is satisfied (use (8.2) and (8.3)).

**Assumption 8.1.** *Monitoring precision is sufficiently high, i.e.* $\Phi > \beta$, *where*

$$\beta \equiv \frac{\gamma R - 1}{\gamma[R - 1]}. \tag{8.4}$$

Assumption 8.1 allows me to focus on the main role that banks play in this model, namely liquidation of bad projects. In particular, banks monitor their borrowers to gather proprietary information about the quality of borrowers and, consequently, act upon this information to deny credit whenever optimal. Such borrowers may be liquidated if they do not obtain financing at a competing bank. As shown here, with a long-term contract with covenants, borrowers that are denied credit by their incumbent (initial) bank will always be liquidated. Assumption 8.1 guarantees that monitoring precision is sufficiently high that liquidation of bad projects indeed occurs. Note that the condition in Assumption 8.1 can be rewritten as

$$\gamma < \gamma_0, \text{ where } \gamma_0 = \frac{1}{R - \Phi[R - 1]}. \tag{8.5}$$

The condition in (8.5) shows that bank monitoring provides for liquidation only if borrowers are relatively unsafe on average. In this case, the incumbent bank rejects borrowers upon receiving negative monitoring signals. Because borrowers have no initial wealth, the bank

[8] Borrowers can have the signals verified in court. This prevents banks from lying.

cannot insure itself higher repayment than the return of the project $R$. However, lending at the interest rate lower than $R$ is not profitable and borrowers are liquidated.

Banks are valuable because they provide for liquidation of bad projects. If borrowers are too safe on average; that is, if $\gamma \geq \gamma_0$, the incumbent bank does not reject them upon receiving negative signals. In this case, no borrower is liquidated and bank monitoring does not play a socially beneficial role. Assumption 8.1 precludes this case.

To summarize, the model contains:

**Lemma 8.1.** *With a long-term contract with covenants, the incumbent (initial) bank continues to finance borrowers (under the agreed conditions) upon receiving positive signals. Upon receiving negative signals the incumbent bank terminates the funding and borrowers are liquidated.*

*Proof of Lemma 8.1:* Upon receiving positive signals, it is profitable to finance borrowers. Even though the incumbent bank would like to increase the interest rate, a long-term contract does not allow for that. The incumbent bank can only terminate the contract but in this case it receives zero profits. Hence, it prefers to continue financing the borrowers at the agreed-upon terms. Upon receiving negative signals it is not profitable to finance the borrowers as Assumption 8.1 guarantees. Hence, the incumbent bank denies credit. A competing bank, if found, anticipates that the rejected borrowers are unprofitable and does not finance them either. ■

Lemma 8.1 points to the benefit of including covenants in a long-term contract. Banks can use valuable monitoring information together with long-term lending. More specifically, the bank monitors borrowers to receive signals about their credit qualities. As a consequence, the incumbent (initial) bank may weed out bad borrowers upon receiving negative signals.

The total profit with a long-term contract with covenants is computed as follows. Borrowers are liquidated if the bank obtains negative signals (see Lemma 8.1). In that case, liquidation proceeds exactly recoup the investment in the project and the total profit is zero. If the bank receives a positive signal, the project financing will be continued. This happens with probability $\gamma\Phi$. That is, with probability $\gamma$ the borrower has a good project and, if and only if this is the case, the bank obtains a positive signal with probability $\Phi$. The gain on a good project is $R-1$. Hence, the expected profits on a long-term contract with covenants are

$$\gamma\Phi[R-1]. \tag{8.6}$$

However, the borrowers only obtain this profit if they are exposed to competing banks at $t=0$. This happens with probability $q_0$. With probability $1-q_0$, there is only one bank, and this bank takes all rents. Hence, the borrowers' expected value of the project with long-term financing (with covenants) is

$$V_{COV} = q_0\gamma\Phi[R-1]. \tag{8.7}$$

Recall that the profits *without* covenants are given in (8.1). The next lemma is the following:

**Lemma 8.2.** *Borrowers prefer a long-term contract with covenants over a long-term contract without covenants.*

*Proof of Lemma 8.2:* Observe that (8.7) is greater than (8.1) for $\Phi > \beta$, and this is true if Assumption 8.1 holds. ■

Lemma 8.2 shows that it is not optimal to design a contract that precludes the incumbent bank from using (costless) monitoring signals. Thus the long-term contract without covenants cannot be optimal. Hence, hereafter the analysis is narrowed to include only long-term contracts with covenants.

### 8.3.2 Short-term contract

This analysis first shows that liquidation with a short-term contract is less efficient than it is with a long-term contract with covenants. Second, the borrowers' expected profits are considered.

#### Insufficient liquidation

The first analysis considers the situation at $t = 1$. Two cases are analyzed individually. In the first case, the focus is on a short-term contract conditional on the absence of a competing bank at $t = 1$; that is, $q_1 = 0$. The incumbent bank then offers second-period financing to the borrowers upon receiving positive signals. Upon receiving negative signals, it denies further financing to such borrowers. Without the competing bank, the rejected borrowers cannot obtain further financing. Therefore, they are forced into liquidation.

The following lemma summarizes the above.

**Lemma 8.3.** *If $q_1 = 0$, borrowers with negative signals are liquidated and borrowers with positive signals obtain further financing.*

Note that liquidation in the case of a short-term contract with no competition; that is, $q_1 = 0$, is equally as efficient as liquidation in the case of a long-term contract with covenants. In both situations the borrowers are liquidated upon receiving negative signals.

The second case considers a short-term contract if borrowers always find a competing bank at $t = 1$; that is, $q_1 = 1$. In this case, the incumbent bank has an informational advantage over the competing bank because it has signals about the borrowers' quality. No pure strategy equilibrium exists, however, the following mixed strategy equilibrium exists (see Broecker (1990), Von Thadden (2004), and Hauswald and Marquez (2006)).

**Lemma 8.4.** *If $q_1 = 1$, the incumbent (initial) bank always makes an offer to borrowers with positive signals and it denies financing to borrowers with bad signals. The competing bank (acting simultaneously) offers a loan with probability $\beta$ and it denies credit with probability $1 - \beta$. Both banks randomly mix over the offered interest rates (the mixing probabilities are* *given in the appendix). The competing bank breaks even. The expected profit of the incumbent bank at $t = 1$ is*

$$\Pi^1_{ST}(q_1 = 1) = \Phi[1 - \gamma]. \tag{8.8}$$

*Proof of Lemma 8.4:* See Appendix. ■

The intuition for Lemma 8.4 is the following. The incumbent bank wants to use its proprietary information to earn profits. The incumbent bank rejects borrowers upon receiving negative signals (because monitoring is precise enough as guaranteed by Assumption 8.1). Things are more complicated for borrowers with good signals. The incumbent bank always makes an offer to such borrowers. However, it cannot offer a unique interest rate because the competing bank would respond by slightly overbidding and seizing all the borrowers. The optimal solution for the incumbent bank is to randomize over the interest rate offered. The competing bank responds by randomizing the interest rates and over-accepting versus rejecting each borrower with the probability $\beta$ versus $1 - \beta$.

Lemma 8.4 points to the key insight of this analysis. A short-term contract can prevent an efficient usage of proprietary information that the incumbent bank has obtained with monitoring. The intuition for this is the following. A short-term contract allows the incumbent bank to use its information advantage to expropriate rents. This rent-seeking behavior of the incumbent bank induces inefficiencies; namely, it allows the competing bank to bid for the borrowers. The competing bank may then not only attract borrowers with good projects but also the bad ones. As a consequence, borrowers can still be financed even if the incumbent bank obtains negative signals. Hence, liquidation is less efficient.

Interestingly, the following proposition shows that the efficiency of liquidation at $t = 1$ depends on the type of contract (combine Lemma 8.1, Lemma 8.3, and Lemma 8.4).

**Proposition 8.1.** *Liquidation is (weakly) more efficient with a long-term contract with covenants than with a short-term contract.*

With a short-term contract, borrowers may be further financed even upon receiving negative signals, whereas with a long-term contract all such borrowers are liquidated. The following corollary shows that this inefficiency stems from the combination of short-term contracts and competition for borrowers.[9]

**Corollary 8.1.** *With no competition at $t = 1$, a short-term contract and a long-term contract with covenants are equally efficient. Increasing competition at $t = 1$ hampers the efficiency of liquidation (only) in the case of a short-term contract.*

*Proof of Corollary 8.1:* Combine Lemma 8.1, Lemma 8.3, and Lemma 8.4. ■

The intuition for Corollary 8.1 is the following. In the case of a short-term contract, borrowers are liquidated upon being rejected by their incumbent bank if there is no competing

[9] Schmeits (2005) also shows that short-term contracts are dominated by long-term contracts. She argues that long-term discretionary contracts allow for an intertemporal wealth redistribution (as well as cross-sectional wealth distribution) that cannot be mimicked by short-term contracts.

bank at $t = 1$. Consequently, the monitoring signals are fully exploited and liquidation is efficient. Yet, as competition increases and a competing bank materializes, the rejected borrowers can possibly finance at the competing bank. This is because the competing bank cannot distinguish good from bad rejected borrowers. This creates an inefficiency in the liquidation process. In contrast, if the borrowers are rejected with a long-term contract with covenants, competition does not affect their liquidation because the competing bank knows that only borrowers upon receiving negative signals end up looking for financing. Hence, it is not willing to finance such borrowers.

**Profits at $t = 1$**

Now I compute the expected profits of the borrowers and incumbent bank at $t = 1$. I denote them with $V^1_{ST}$ and $\Pi^1_{ST}$, respectively. I proceed as follows. With probability $1 - q_1$, borrowers do not find a competing bank. In this situation, borrowers obtain zero profits and the incumbent bank obtains $\gamma\Phi[R - 1]$; see (8.6). That is,

$$V^1_{ST}(q_1 = 0) = 0, \quad \Pi^1_{ST}(q_1 = 0) = \gamma\Phi[R - 1]. \tag{8.9}$$

With probability $q_1$, borrowers find a competing bank. In this case, the incumbent bank obtains $\Phi[1-\gamma]$, see (8.8). The borrowers' profits are computed as follows. With probability $\gamma$, the borrower is good. In this case his project yields $R$. Conditional on the borrower being good, he obtains financing at the incumbent bank with probability $\Phi$ (upon receiving a positive signal) and at the competing bank with probability $[1-\Phi]\beta$. Hence, the probability of being financed is $\Phi + [1 - \Phi]\beta$ and the expected return is

$$\gamma\{\Phi + [1 - \Phi]\beta\}R. \tag{8.10}$$

The costs of financing the project are as follows. The borrower cannot pursue the project if he is rejected simultaneously by the incumbent bank and competing bank. The incumbent bank rejects the borrower with probability $\gamma[1 - \Phi] + [1 - \gamma] = 1 - \gamma\Phi$ and the competing bank does so with probability $1-\beta$. If the project is pursued, \$1 of funds are needed. Hence, the total investment needed is

$$1 - [1 - \gamma\Phi][1 - \beta]. \tag{8.11}$$

Subtracting (8.11) from (8.10) yields the total expected profits of a short-term contract

$$\Phi[1 - \gamma] + \beta[\gamma R - 1]. \tag{8.12}$$

Borrowers, however, cannot take the total profit because the incumbent bank can take at least the rents as given in (8.8). In sum,

$$V^1_{ST}(q_1 = 1) = \beta[\gamma R - 1], \quad \Pi^1_{ST}(q_1 = 1) = \Phi[1 - \gamma]. \tag{8.13}$$

With probability $1-q_1$, borrowers do not find a competing bank and the computed profits in (8.9) are relevant. With probability $q_1$, borrowers find a competing bank and the profits in (8.13) apply. That is,

$$V_{ST}^1 = q_1\beta[\gamma R - 1], \tag{8.14}$$
$$\Pi_{ST}^1 = [1-q_1]\gamma\Phi[R-1] + q_1\Phi[1-\gamma]. \tag{8.15}$$

**Profits at $t = 0$**

Now I compute the expected profits of the borrowers at $t = 0$ (I denote this with $V_{ST}$). With probability $1-q_0$, borrowers are only served by one bank. Hence, this bank has all the bargaining power and the borrowers receive nothing except what they anticipated to make after the first period contract.[10] That is, they earn (8.14), i.e.,

$$V_{ST}(q_0 = 0) = q_1\beta[\gamma R - 1]. \tag{8.16}$$

With probability $q_0$, however, borrowers have competing banks to choose from at $t = 0$. In this case, the bank anticipates that it will earn the profit as given in (8.15). Hence, it is willing to subsidize the borrowers in the first period by exactly the amount of its expected second-period profits.[11] The borrowers expect the following profits (add (8.14) and (8.15)):

$$V_{ST}(q_0 = 1) = q_1[\beta - \Phi][\gamma R - 1] + \gamma\Phi[R-1]. \tag{8.17}$$

Combining (8.16) with (8.17), I determine the expected profits of the borrowers choosing a short-term contract,

$$V_{ST} = q_1[\beta - q_0\Phi][\gamma R - 1] + q_0\gamma\Phi[R-1]. \tag{8.18}$$

## 8.4 Contract Choice

In this section I analyze the borrowers' choices between contracts. I discuss the strategic considerations of the borrowers' choices of contract with respect to the expected level of competition at $t = 0$. I conclude with some comparative statics results.

First, I limit the analysis to the situation where there is competition for the borrowers at $t = 0$. That is, the borrowers choose the contract type anticipating that they will find (two) competing banks at $t = 0$. I can now show the following proposition.

[10] A short-term contract does not allow the bank to demand a long-term repayment from the borrowers. A contract with a long-term repayment would already be a long-term contract. I also step aside from potential renegotiation that may appear at $t = 1$ between the incumbent bank and borrowers. In particular, the incumbent bank cannot demand higher repayment at $t = 1$ than is the liquidation value of the project. Berlin and Mester (1992) analyze the role of covenants in light of efficient renegotiation.

[11] Borrowers at $t = 1$ again borrow \$1 and consume the rest at $t = 1$. The results do not change if I allow borrowers to invest the proceeds of the first period into the second period of the project, assuming that they are uninformed about the signals.

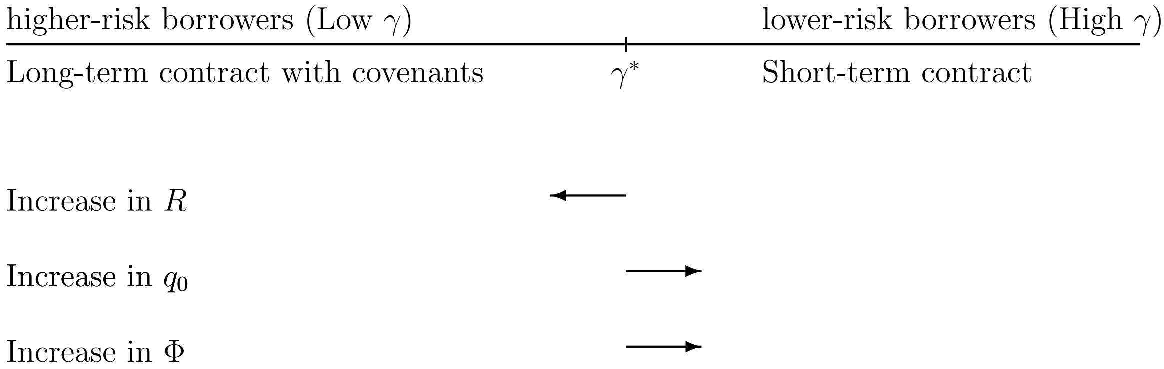


Figure 8.2: Ranges of optimal financing

**Proposition 8.2.** *If there is competition at $t = 0$, i.e. $q_0 = 1$, borrowers (weakly) prefer a long-term contract with covenants.*

*Proof of Proposition 8.2:* Note that (8.17) is greater (or equal if $q_1 = 0$) than (8.7). ■

The intuition is that a long-term contract with covenants allows for more efficient liquidation than a short-term contract (see Proposition 8.1). Namely, with a short-term contract the competing bank may also finance the rejected borrowers, which is inefficient. Hence, the total profit with a short-term contract is (weakly) lower than with a long-term contract with covenants. If competition at $t = 0$ is perfect, borrowers are able to take the total profit; hence, they choose a long-term contract with covenants.

**Corollary 8.2.** *If $q_0 = 1$, borrowers select the welfare optimal type of contract – a long-term contract with covenants.*

If competition at $t = 0$ is perfect, the borrowers anticipate taking all rents. Hence, they choose the type of contract that maximizes the total profit; that is, the welfare optimal contract. The welfare optimal type of contract is a long-term contract because it provides for efficient liquidation (see Proposition 8.1).

Now I generalize the analysis to the situation where competition at $t = 0$ is limited (described by parameter $q_0$). In this case, borrowers choose the type of a contract anticipating that they find themselves with competing banks only with probability $q_0$. With probability $1 - q_0$, they are left with only one bank.

I show the following proposition.

**Proposition 8.3.** *For $\gamma > \gamma^*$, where $\gamma^* \leq \gamma_0$, borrowers prefer a short-term contract. For $\gamma \leq \gamma^*$, borrowers prefer a long-term contract with covenants.*

*Proof of Proposition 8.3:* Subtract (8.7) from (8.18) to obtain

$$V_{ST} - V_{COV} = q_1[\beta - q_0\Phi][\gamma R - 1]. \tag{8.19}$$

Setting (8.19) to zero and solving for $\gamma$, I obtain the following solution (use (8.4))

$$\gamma^* = \frac{1}{R - q_0\Phi[R-1]}. \tag{8.20}$$

Use (8.4) and (8.19) to see that $V_{ST} > V_{COV}$ if $\gamma > \gamma^*$ and $V_{ST} \leq V_{COV}$ if $\gamma \leq \gamma^*$. ■

The intuition for Proposition 8.3 is the following. Borrowers may choose a short-term contract to partially protect themselves from being exploited by the incumbent (initial) bank. That is, with a short-term contract competition might also be realized at $t = 1$, and that allows borrowers to take second-period profits. For lower-risk borrowers (i.e., if $\gamma > \gamma^*$), liquidation efficiency is less important, and this consideration prevails and they choose a short-term contract. However, for higher-risk borrowers (i.e., if $\gamma \leq \gamma^*$), efficient liquidation of bad projects becomes important and borrowers choose a long-term contract with covenants.

Figure 8.2 graphically presents the result of Proposition 8.3. On average, higher-risk borrowers (where $\gamma$ is low; i.e., $\gamma \leq \gamma^*$) choose a long-term contract with covenants. On average, lower-risk borrowers (where $\gamma$ is high; i.e., $\gamma > \gamma^*$) choose a short-term contract. Figure 8.2 also presents the effects of an increase in competition $q_0$ and monitoring precision $\Phi$ on the use of long and short-term contracts. I analyze these comparative statics results now.

The return of borrowers' projects affect the choice of contract as follows.

**Corollary 8.3.** *As their return increases, borrowers more often use a short-term contract relative to a long-term contract with covenants; i.e.,* $\frac{\partial\gamma^*}{\partial R} < 0$.

*Proof of Corollary 8.3:* Note from (8.20) that $\gamma^*$ is decreasing in $R$. ■

The intuition for this Corollary is reminiscent of Proposition 8.3. Borrowers with higher return in the case of success $R$ want to protect themselves from being exploited by the incumbent (initial) bank. With a short-term contract, competition may also be realized at $t = 1$, and this may allow borrowers to take second-period profits. Hence, they choose a short-term contract more often.

The effect of competition on the choice of contract is as follows.

**Corollary 8.4.** *As competition for borrowers at* $t = 0$ *increases, borrowers more often use a long-term contract with covenants relative to a short-term contract, i.e.* $\frac{\partial\gamma^*}{\partial q_0} > 0$.

*Proof of Corollary 8.4:* Note from (8.20) that $\gamma^*$ is increasing in $q_0$. ■

As competition increases, borrowers need a short-term contract less to protect against monopolistic market power of the incumbent bank. Hence, they choose a long-term contract with covenants that optimizes the liquidation decision more often.

Next I analyze how monitoring precision influences the optimal financing mode of the borrowers.

**Corollary 8.5.** *As monitoring precision increases, borrowers more often choose a long-term contract with covenants relative to a short-term contract; i.e.,* $\frac{\partial\gamma^*}{\partial\Phi} > 0$.

*Proof of Corollary 8.5:* Use (8.20) to see that $\gamma^*$ is increasing in $\Phi$. ■

High monitoring precision exacerbates the insufficient liquidation of a short-term contract. More specifically, especially in the case of high-quality banks for which monitoring precision is high, borrowers should be liquidated upon receiving negative signals. Yet, with a short-term contract, such borrowers may still obtain financing at a competing bank. This is inefficient. With a long-term contract with covenants, borrowers are always liquidated upon receiving negative signals. Hence, higher precision of a monitoring signal (higher quality of the bank) makes long-term contracts with covenants more attractive (see Figure 8.2).

## 8.5 Model Extensions and Empirical Predictions

In this section I first analyze an extension in which covenants are costly. Second, I discern some empirical predictions from the analysis.

### 8.5.1 Costly covenants

I expand the previous model by assuming that including covenants in the contract is not costless for borrowers. In particular, I assume that including such covenants in the loan contract limits the flexibility of borrowers. More specifically, I assume that a borrower cannot undertake an additional action at $t = 1/2$ that provides a synergy benefit of $\alpha[\gamma R - 1]$ at $t = 1$. Note that the synergy benefit $\alpha[\gamma R - 1]$ is proportional to the expected return that a borrower obtains on the project when he is not liquidated. I let $\alpha$ be the proportionality parameter.

Now borrowers will prefer a short-term contract more often to keep their flexibility intact. The difference between the expected profit of borrowers with a short-term contract and long-term contract with covenants is now (compare with (8.19))

$$V_{ST} - V_{COV} = \{q_1[\beta - q_0\Phi] + \alpha\}[\gamma R - 1]. \tag{8.21}$$

Setting (8.21) to zero and solving for $\gamma$, it follows that borrowers are indifferent between a short-term contract and a long-term contract if the expression in (8.21) equals zero; that is, if

$$\gamma_2 = \frac{1}{R - [q_0\Phi - \frac{\alpha}{q_1}][R-1]}. \tag{8.22}$$

Borrowers prefer a short-term contract for $\gamma > \gamma_2$ and a long-term contract with covenants for $\gamma \leq \gamma_2$.

Because including covenants is costly, borrowers choose a long-term contract with covenants less often (use (8.20) and (8.22) to see that $\gamma_2 < \gamma^*$). The higher the costs of covenants are (i.e., the higher $\alpha$ is), the lower $\gamma$ has to be to choose for a long-term contract with covenants. Only if $\gamma$ is low enough, the additional value obtained due to the more efficient liquidation at $t = 1$ compensates borrowers for the lost flexibility.

I can now show the following result.

**Proposition 8.4.** *As the expected competition at $t = 1$ increases, borrowers more often prefer a long-term contract with covenants.*

*Proof of Proposition 8.4:* Use (8.22) to see that $\gamma_2$ is increasing in $q_1$. ■

The intuition for this result is the following. Increasing competition at $t = 1$ exacerbates the drawback of a short-term contract. In particular, it is more likely that borrowers obtain further financing upon receiving negative signals; that is, a competing bank materializes more often. That is, liquidation becomes less efficient with a short-term contract. In response, borrowers more often prefer a long-term contract with covenants to benefit from more efficient liquidation.

Proposition 8.4 is counterintuitive at first sight. One could expect that borrowers more often choose a short-term contract upon anticipating an increase in competition in the future. A short-term contract could allow borrowers to search for a competing bank in the future when competition is higher. Proposition 8.4 shows that another, stronger effect is at work. As competition increases in the future, the inefficiency of a short-term contract worsens; the liquidation decision becomes less efficient. Hence, borrowers prefer a long-term contract with covenants to bypass the lost efficiency of a short-term contract.

### 8.5.2 Empirical predictions and evidence

In this section, I list the empirical/stylized facts that are consistent with the predictions following from my analysis.

1. Lummer and McConnell (1989) classify announcements of loan agreements and revisions to agreements into announcement containing positive or negative information. They show that negative renewals have significantly negative abnormal returns. Cancellation of a bank loan results in the strongest negative response. Lemma 8.1 predicts that the termination of a long-term contract should result in liquidation of borrowers. This may be different with a short-term contract. In particular, Lemma 8.3 shows that upon termination of a short-term contract borrowers may obtain funds at the competing bank. This predicts that the termination of a long-term loan should result in a stronger negative announcement effect than the termination of a short-term contract. Proposition 8.3 predicts that this effect should be stronger if competition between banks is high.

2. Proposition 8.3 predicts that the relation between the maturity of loan contracts and borrowers' risk should be positive. Berger, Espinosa-Vega, Frame, and Miller (2005) provide evidence that higher-risk borrowers raise bank loans with longer maturities than lower-risk borrowers. Sorge and Zhang (2007) show that countries with higher dispersion of firms' credit qualities are characterized by more long-term debt. However, the empirical evidence is mixed. Barclay and Smith Jr. (1995) and Johnson (2003) point

to the non-monotone relation between debt maturity and borrowers' risk as predicted by Diamond (1991a). Ortiz-Molina and Penas (2006) use accounting data to assess a firm's risk. On the sample of small firms in the U.S., they show that riskier firms use loans with lower maturity. Hence, more research is needed on this point. My analysis stresses that focusing only on the maturity of loans may not be enough. Considering other contractual features such as loan covenants seems to be crucial.

3. The implication of Lemma 8.2 is that especially long-term loans should include covenants. Billett, King, and Mauer (2007) confirm the positive relation between the loan maturity and the inclusion of covenants.

4. Corollary 8.3 predicts that borrowers with high growth potential more often finance with a short-term debt. Barclay and Smith Jr. (1995) and Goyal, Lehn, and Racic (2002) confirm this relation.

5. Corollary 8.5 predicts that higher monitoring precision leads to higher maturity of lending contracts. First, this is aligned with the evidence that smaller firms – which are likely to be relatively opaque (i.e., monitoring precision is likely to be low) – finance with short-term debt (see Stohs and Mauer (1996) and Scherr and Hulburt (2001)). Second, Berger, Espinosa-Vega, Frame, and Miller (2005) directly confirm predictions from Corollary 8.5. They show that higher quality banks that have more advanced credit-scoring techniques and therefore higher monitoring precision grant loans with on average longer maturity.

6. Corollary 8.5 also predicts that higher quality banks (with high monitoring precision) should use covenants more frequently. This is related to the empirical evidence put forward by Smith Jr. and Warner (1979) and Dichev and Skinner (2002) that shows that covenants are set more tightly in private lending agreements (where monitoring precision is high) than in public lending agreements.

7. Corollary 8.5 establishes a relation between contract features and characteristics of banking systems. In high-quality banking systems, longer-term contracts and covenants would be used more often than in low-quality banking systems. This is confirmed by Caprio Jr. and Demirgüç-Kunt (1997), who show that in developing countries borrowers predominantly use short-term contracts. Additionally, Sorge and Zhang (2007) show that borrowers in countries with poor credit information and low quality of accounting standards use contracts with shorter maturities. A prediction on the comparative use of covenants in developing versus developed countries remains to be tested.

8. Corollary 8.4 and Proposition 8.4 predict some yet unveiled relation between competition and structural features of contracts. First, increasing competition should augment

the use of bank covenants. Second, increasing competition should augment the maturity of lending contracts. The closest empirical research to this end seems to be Caprio Jr. and Demirgüç-Kunt (1997). They show that in developing countries borrowers use shorter term contracts than their developed country counterparts. One possibility for this could be that in developing countries bank competition is limited and that this makes a short-term contract superior. As competition increases, in line with my analysis, short-term contracts are replaced by longer maturity contracts that include other contractual features such as covenants.

## 8.6 Conclusions

This chapter adds some key insights to understanding the interaction between the choice of the type of a contract and competition. I show that a short-term contract can lead to insufficient liquidation of (bad) borrowers. This is because a short-term contract can worsen the efficient use of information obtained with monitoring, particularly if competition for borrowers is high. Using a long-term contract with a termination clause could mitigate this inefficiency.

More specifically, a long-term contract with covenants prevents a bank from extracting rents from its borrowers. Only borrowers for which the bank has obtained negative signals are rejected. Other banks anticipate this and such borrowers are efficiently liquidated. In contrast, with a short-term contract, the incumbent (initial) bank exploits proprietary monitoring signals and seeks to extract rents from its borrowers. Now, the competing bank cannot distinguish whether borrowers are switching to gain better funding terms or because they were rejected due to the negative signals of their bank. In equilibrium, the competing bank also finances some rejected borrowers. This makes liquidation insufficient.

To summarize, efficient liquidation makes a long-term contract socially optimal. However, borrowers may opt for a short-term contract in order to gain bargaining power over the incumbent (initial) bank in the later period. This leads then to the key prediction of the analysis: As competition for borrowers increases, borrowers choose a longer term contract coupled with covenants. This is for two reasons. First, higher competition eliminates banks' rents; hence, borrowers opt for a more efficient long-term contract with covenants. Second, higher expected competition exacerbates the insufficient liquidation of a short-term contract.

## 8.7 Appendix

*Proof of Lemma 8.4*

The proof is similar to the proof of Proposition 1 in Hauswald and Marquez (2002); hence, the proof of existence of the mixed strategy equilibrium is omitted (see also Hauswald and Marquez (2006)). If the incumbent bank observes a positive signal, the project always succeeds; that is, $p(S = G) = 1$. If the incumbent bank observes a negative signal, the project succeeds with probability $p(S = B)$ as given in (8.2). In this case, the incumbent bank rejects the second-period financing. To see this, note that Assumption 8.1 guarantees that $p(S = B)R < 1$.

The probability that the incumbent bank is willing to make an offer is $Pr(G) = \gamma\Phi$. With probability $Pr(B) = 1 - \gamma\Phi$, the incumbent bank rejects the borrower. In a mixed strategy equilibrium, the incumbent and competing bank obtain equal profits for all interest-rate offers that they are willing to make. Thus,

$$[1 - F_C(r)][p(G)r - 1] = \Pi(S = G), \tag{8.23}$$

$$Pr(G)[1 - F_I(r|S = G)][p(G)r - 1] + Pr(B)[p(B)r - 1] = 0, \tag{8.24}$$

where I have defined the density distributions for the offers of the incumbent bank as $F_I(r)$ and the competing bank as $F_C(r)$. The lowest level of interest rate that the competing bank is willing to offer is $r = \frac{1}{\gamma}$. Hence, I have $F_C(\frac{1}{\gamma}) = 0$. Insert this into (8.23) to obtain the profitability of a bank upon obtaining a positive signal

$$\Pi(S = G) = 1/\gamma - 1. \tag{8.25}$$

However, the incumbent bank receives a positive signal with probability $\gamma\Phi$. The profit at $t = 1$ of the incumbent bank as expected at $t = 0$ is as given in (8.8). Rearrange (8.24) to obtain

$$\gamma\Phi[1 - F_I(r|S = G)][r - 1] + [1 - \gamma\Phi][\frac{\gamma[1 - \Phi]}{1 - \gamma\Phi}r - 1] = 0. \tag{8.26}$$

Some manipulations yield

$$F_I(r|S = G) = \frac{\gamma r - 1}{\Phi\gamma[r - 1]}. \tag{8.27}$$

From (8.23) one obtains

$$F_C(r) = \frac{\gamma r - 1}{\gamma[r - 1]}. \tag{8.28}$$

Note that the competing bank makes an offer with probability

$$F_C(R) = \frac{\gamma R - 1}{\gamma[R - 1]} = \beta, \tag{8.29}$$

which is identical to $\beta$ as defined in (8.4). With probability $1 - \beta$, the competing bank rejects the borrower. ■

# Bibliography


Agarwal, S., and R. Hauswald, 2007, "Distance and Information Asymmetries in Lending," Working paper.

Aghion, P., and P. Bolton, 1987, "Contracts as a Barrier to Entry," *American Economic Review*, 77(3), 388–401.

Aghion, P., and P. Howitt, 1992, "A Model of Growth through Creative Destruction," *Econometrica*, 60(2), 323–351.

Akerlof, G. A., 1970, "The Market for "Lemons:" Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Allen, F., 1990, "The Market for Information and the Origin of Financial Intermediation," *Journal of Financial Intermediation*, 1(1), 3–30.

Allen, F., and E. Carletti, 2005, "Credit Risk Transfer and Contagion," *Journal of Monetary Economics*, 53(1), 89–111.

Allen, F., E. Carletti, and R. Marquez, 2007, "Credit Market Competition and Capital Regulation," mimeo, University of Pennsylvania.

Allen, F., and D. Gale, 1997, "Financial Markets, Intermediaries, and Intertemporal Smoothing," *Journal of Political Economy*, 105(3), 523–546.

André, P., R. Mathieu, and P. Zhang, 2001, "A Note on: Capital Adequacy and the Information Content of Term Loans and Lines of Credit," *Journal of Banking & Finance*, 25, 431–444.

Barclay, M. J., and C. W. Smith Jr., 1995, "The Maturity Structure of Corporate Debt," *Journal of Finance*, 50(2), 609–631.

Bartelsman, E. J., and M. Doms, 2000, "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

Barth, J. R., G. Caprio Jr., and R. Levine, 2004, "Bank Regulation and Supervision: What Works Best?," *Journal of Financial Intermediation*, 13(2), 205–248.

Barth, J. R., S. Trimbath, and G. Yago, 2004, *The Savings and Loan Crisis: Lessons from a Regulatory Failure*. Milken Institute, Santa Monica, California.

Beck, T., A. Demirgüç-Kunt, and R. Levine, 2003, "Bank Concentration and Crises," NBER Working Paper 9921.

——— , 2005, "Bank Concentration and Fragility: Impact and Mechanics," NBER Working Paper 11500.

Berger, A. N., S. D. Bonime, L. G. Goldberg, and L. J. White, 2004, "The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on Entry in the Banking Industry," *Journal of Business*, 77(4), 797–834.

Berger, A. N., and C. Bouwman, 2007, "Bank Liquidity Creation," EFA 2007 Ljubljana Meetings.

Berger, A. N., R. S. Demsetz, and P. E. Strahan, 1999, "The Consolidation of the Financial Services Industry: Causes, Consequences and Implications for the Future," *Journal of Banking & Finance*, 23, 135–194.

Berger, A. N., M. A. Espinosa-Vega, W. S. Frame, and N. H. Miller, 2005, "Debt Maturity, Risk, and Asymmetric Information," *Journal of Finance*, 60(6), 2895–2923.

Berger, A. N., and T. Hannan, 1998, "The Efficiency Cost of Market Power in the Banking Industry: A Test of the "Quiet Life" and Related Hypotheses," *The Review of Economics and Statistics*, 80(3), 454–465.

Berger, A. N., and L. J. Mester, 1996, "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?," *Journal of Banking & Finance*, 21, 895–947.

Berger, A. N., N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. Stein, 2005, "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks," *Journal of Financial Economics*, 76(2), 237–269.

Berger, A. N., and E. Ofek, 1995, "Diversification's Effect on Firm Value," *Journal of Financial Economics*, 37(1), 39–65.

Berger, A. N., R. J. Rosen, and G. F. Udell, 2007, "Does Market Size Structure Affect Competition? The Case of Small Business Lending," *Journal of Banking & Finance*, 31(1), 11–33.

Berger, A. N., and G. F. Udell, 2006, "A More Complete Conceptual Framework for Financing of Small and Medium Enterprises," *Journal of Banking & Finance*, 30(11), 2945–2966.

Berglof, E., and E.-L. Von Thadden, 1994, "Short-Term versus Long-Term Interests: Capital Structure with Multiple Investors," *The Quarterly Journal of Economics*, 109(4), 1055–1084.

Berlin, M., and L. J. Mester, 1992, “Debt Covenants and Renegotiation,” *Journal of Financial Intermediation*, 2(2), 95–133.

——— , 1999, “Deposits and Relationship Lending,” *Review of Financial Studies*, 12(3), 579–607.

Bernanke, B. S., 1983, “Nonmonetary Effects of the Financial Crisis in Propagation of the Great Depression,” *American Economic Review*, 73(3), 257–276.

Besanko, D., and G. Kanatas, 1996, “The Regulation of Bank Capital: Do Capital Standards Promote Bank Safety?,” *Journal of Financial Intermediation*, 5(2), 160–183.

Besanko, D., and A. V. Thakor, 1993, “Relationship Banking, Deposit Insurance and Bank Portfolio Choice,” in *Capital Markets and Financial Intermediation*, ed. by C. Mayer, and X. Vives. Cambridge University Press, Cambridge, pp. 292–319.

Best, R., and H. Zhang, 1993, “Alternative Information Sources and the Information Content of Bank Loans,” *Journal of Finance*, 48(4), 1507–1522.

Bhaskar, V., and T. To, 2004, “Is Perfect Price Discrimination Really Efficient? An Analysis of Free Entry Equilibria,” *RAND Journal of Economics*, 35(4), 762–776.

Bhattacharya, S., A. W. A. Boot, and A. V. Thakor, 1998, “The Economics of Bank Regulation,” *Journal of Money, Credit, and Banking*, 30(4), 745–770.

Bhattacharya, S., and D. Gale, 1987, “Preference Shocks, liquidity and Central Bank Policy,” in *New Approaches to Monetary Economics*. Cambridge University Press, Cambridge.

Bhattacharya, S., and A. V. Thakor, 1993, “Contemporary Banking Theory,” *Journal of Financial Intermediation*, 3(1), 2–50.

Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, “A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades,” *Journal of Political Economy*, 100(5), 992–1026.

——— , 1998, “Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades,” *Journal of Economic Perspectives*, 12(3), 151–170.

Billett, M. T., M. J. Flannery, and J. A. Garfinkel, 1995, “The Effect of Lender Identity on a Borrowing Firm’s Equity Return,” *Journal of Finance*, 50(2), 699–718.

Billett, M. T., T.-H. D. King, and D. C. Mauer, 2007, “Growth Opportunities and the Choice of Leverage, Debt Maturity, and Covenants,” *Journal of Finance*, 62(2), 697–730.

Bizer, D. S., and P. M. DeMarzo, 1992, “Sequential Banking,” *Journal of Political Economy*, 100(11), 41–61.

Blum, J., 1999, "Do Capital Adequacy Requirements Reduce Risks in Banking?," *Journal of Banking & Finance*, 23(5), 755–771.

Blum, J., and M. Hellwig, 1995, "The Macroeconomic Implications of Capital Adequacy Requirements for Banks," *European Economic Review*, 39(3-4), 739–749.

Bolton, P., and M. Dewatripont, 2005, *Credit, Intermediation, and the Macroeconomy.* MIT Press, Cambridge, Massachusetts.

Bolton, P., and D. S. Scharfstein, 1990, "A Theory of Predation Based on Agency Problems in Financial Contracting," *American Economic Review*, 80(1), 93–106.

——— , 1996, "Optimal Debt Structure and the Number of Creditors," *Journal of Political Economy*, 104(1), 1–25.

Boot, A. W. A., 2000, "Relationship Banking: What do we Know," *Journal of Financial Intermediation*, 9(1), 7–25.

——— , 2003, "Consolidation and Strategic Positioning in Banking with Implications for Europe," in *Brookings-Wharton Papers on Financial Services: 2003*, ed. by R. E. Litan, and R. Herring. Brookings Institution Press, Washington.

——— , 2006, "Supervisory Arrangements, LoLR and Crisis Management in a Single European Banking Market," *Economic Review*, 2, 15–33, Central Bank of Sweden (Riksbank).

Boot, A. W. A., and S. I. Greenbaum, 1993, "Bank Regulation, Reputation and Rents: Theory and Policy Implications," in *Capital Markets and Financial Intermediation.* Cambridge University Press, Cambridge, New York, and Melbourne, pp. 262–285.

Boot, A. W. A., T. T. Milbourn, and S. Deželan, 2001, "Regulation and the Evolution of the Financial Services Industry," in *Challenges for Central Banking*, ed. by A. M. Santomero, S. Viotti, and A. Vredin. Kluwer, Massachussetts, pp. 39–58.

Boot, A. W. A., and A. V. Thakor, 1993, "Self-Interested Bank Regulation," *American Economic Review*, 83(2), 206–212.

——— , 1997, "Financial System Architecture," *The Review of Financial Studies*, 10(3), 693–733.

——— , 2000, "Can Relationship Banking Survive Competition?," *Journal of Finance*, 55(2), 679–713.

Bordo, M., B. Eichengreen, D. Klingebiel, and M. S. Martinez-Peria, 2001, "Is the crisis problem growing more severe?," *Economic Policy*, 16(32), 51–82.

Borio, C. E. V., 2003, "Towards a Macroprudential Framework for Financial Supervision and Regulation?," Discussion Paper 128, Bank for International Settlements.

Boscaljon, B., and C.-C. Ho, 2005, "Information Content of Bank Loan Announcements to Asian Corporations during Periods of Economic Uncertainty," *Journal of Banking & Finance*, 29(2), 369–389.

Boyd, J. H., and G. De Nicolo, 2005, "Bank Risk Taking and Competition Revisited," *Journal of Finance*, 60(3), 1329–1343.

Boyd, J. H., G. De Nicolo, and B. D. Smith, 2004, "Crises in Competitive versus Monopolistic Banking Systems," *Journal of Money, Credit, and Banking*, 36(3), 487–506.

Broecker, T., 1990, "Credit-Worthiness Tests and Interbank Competition," *Econometrica*, 58(2), 429–452.

Bryan, L. L., 1988, *Breaking Up the Bank: Rethinking an Industry under Siege*. Dow Jones-Irwin, Homewood, IL.

Bryant, J., 1980, "A Model of Reserves, Bank Runs, and Deposit Insurance," *Journal of Banking & Finance*, 4(4), 335–344.

Calem, P., and R. Rob, 1999, "The Impact of Capital-Based Regulation on Bank Risk-Taking," *Journal of Financial Intermediation*, 8(4), 317–352.

Calomiris, C. W., 1999, "Building an Incentive-Compatible Safety Net," *Journal of Banking & Finance*, 27(10), 1499–1519.

Calomiris, C. W., and C. M. Kahn, 1991, "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review*, 81(3), 497–513.

Calomiris, C. W., and R. E. Litan, 2000, "Financial Regulation in a Global Marketplace," in *Papers on Financial Services 2000*, ed. by R. E. Litan, and A. M. Santomero. Brookings Institution Press, Washington, pp. 283–340.

Calomiris, C. W., and J. R. Mason, 2004, "Credit Card Securitization and Regulatory Arbitrage," *Journal of Financial Services Research*, 26(1), 5–27.

Caprio Jr., G., and A. Demirgüç-Kunt, 1997, "The role of long term finance: Theory and evidence," Policy Research Working Paper Series 1746, The World Bank.

Carlson, M., and K. J. Mitchener, 2006, "Bank Branching, Bank Competition, and Financial Stability," *Journal of Money, Credit, and Banking*, 38(5), 1293–1328.

Chan, Y.-S., S. I. Greenbaum, and A. V. Thakor, 1986, "Information Reusability, Competition and Bank Asset Quality," *Journal of Banking & Finance*, 10(2), 243–253.

——— , 1992, "Is Fairly-Priced Deposit Insurance Possible?," *Journal of Finance*, 47(1), 227–245.

Chari, V. V., and R. Jagannathan, 1988, "Banking Panics, Information, and Rational Expectations Equilibrium," *Journal of Finance*, 18(3), 749–761.

Chava, S., and M. R. Roberts, 2006, "Is Financial Contracting Costly? An Empirical Analysis of Debt Covenants and Corporate Investment," Working paper.

Chen, Y., 1999, "Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities," *Journal of Political Economy*, 107(5), 946–968.

Claessens, S., and L. Laeven, 2004, "What Drives Bank Competition? Some International Evidence," *Journal of Money, Credit, and Banking*, 36(3), 563–583.

Clarke, M. Z., 2004, "Geographic Deregulation of Banking and Economic Growth," *Journal of Money, Credit, and Banking*, 36(5), 929–942.

Deep, A., and G. Schaefer, 2004, "Are Banks Liquidity Transformers?," Harvard University Working Paper.

Degryse, H., and S. Ongena, 2004, "The Impact of Technology and Regulation on the Geographical Scope of Banking," *Oxford Review of Economic Policy*, 20(4), 571–590.

——— , 2005, "Distance, Lending Relationships, and Competition," *Journal of Finance*, 60(1), 231–266.

Dell'Ariccia, G., and R. Marquez, 2004, "Information and Bank Credit Allocation," *Journal of Financial Economics*, 72(1), 185–214.

——— , 2006, "Lending Booms and Lending Standards," *Journal of Finance*, 61(5), 2511–2546.

DeMarzo, P. M., and D. Duffie, 1999, "A Liquidity-Based Model of Security Design," *Econometrica*, 67(1), 65–100.

Demirgüç-Kunt, A., and E. Detragiache, 2001, "Financial Liberalization and Financial Fragility," in *Financial Liberalization: How Far, How Fast?*, ed. by G. Caprio, P. Honohan, and E. J. Stiglitz. Cambridge University Press, Massachusetts, pp. 96–122.

——— , 2002, "Does Deposit Insurance Increase Banking System Stability? An Empirical Investigation," *Journal of Monetary Economics*, 49(7), 1373–1406.

Demirgüç-Kunt, A., B. Karacaovali, and L. Laeven, 2005, "Deposit Insurance around the World: A Comprehensive Database," World Bank Working Paper 3628.

Demsetz, H., 1973, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law & Economics*, 16(1), 1–9.

Dermine, J., 2006, "European Banking Integration: Don't Put the Cart before the Horse," *Financial Markets, Institutions & Instruments*, 15(2), 57–106.

Diamond, D. W., 1984, "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51(3), 393–414.

——— , 1991a, "Debt Maturity Structure and Liquidity Risk," *Quarterly Journal of Economics*, 106(3), 709–737.

——— , 1991b, "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy*, 99(4), 689–721.

Diamond, D. W., and P. H. Dybvig, 1983, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91(3), 401–419.

Diamond, D. W., and R. G. Rajan, 2000, "A Theory of Bank Capital," *Journal of Finance*, 55(6), 2431–2464.

——— , 2001, "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, 109(2), 287–327.

Diamond, P. A., 1971, "A Model of Price Adjustment," *Journal of Economic Theory*, 3(2), 156–168.

Dichev, I. D., and D. J. Skinner, 2002, "Large-Sample Evidence on the Debt Covenant Hypothesis," *Journal of Accounting Research*, 40(4), 1091–1123.

Dixit, A. K., and J. E. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3), 297–308.

Dowd, K., 2000, "Bank Capital Adequacy versus Deposit Insurance," *Journal of Financial Services Research*, 17, 7–15.

Dunne, T., M. J. Roberts, and L. Samuelson, 1989, "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, 104(4), 671–698.

Englund, P., 1999, "The Swedish Banking Crisis: Roots and Consequences," *Oxford Review of Economic Policy*, 15(3), 80–97.

European Commission, 2006a, "Communication from the Commission to the European Parliament and the Council Concerning the Review of Directive 94/19/EC on Deposit Guarantee Schemes," http://ec.europa.eu/internal_market/bank/docs/guarantee/comm9419_en.pdf.

——— , 2006b, "Proposal for a Directive of the European Parliament and of the Council amending Council Directive 92/49/EEC and Directives 2002/83/EC, 2004/39/EC, 2005/68/EC and 2006/48/EC as Regards Procedural Rules and Evaluation Criteria for the Prudential Assessment of Acquisitions and Increase of Shareholding in the Financial Sector," http://ec.europa.eu/internal_market/finances/cross-sector/index_en.htm#obstacles.

Evanoff, D. D., and E. Ors, 2002, "Local Market Consolidation and Bank Productive Efficiency," Bank of Chicago Working Paper.

Fields, L. P., D. R. Fraser, T. L. Berry, and S. Byers, 2006, "Do Bank Loan Relationships Still Matter?," *Journal of Money, Credit, and Banking*, 38(5), 1195–1209.

Financial Service Authority, 2003, "Reasonable Expectations: Regulation in a Non-Zero Failure World," September.

Flannery, M. J., 1986, "Asymmetric Information and Risky Debt Maturity Choice," *Journal of Finance*, 41(1), 19–37.

——— , 1994, "Debt Maturity and the Deadweight Cost of Leverage: Optimal Financing Banking Firms," *American Economic Review*, 84(1), 320–331.

Flannery, M. J., and K. P. Rangan, 2004, "Market Forces at Work in the Banking Industry: Evidence from the Capital Buildup of the 1990s," mimeo, University of Florida.

Focarelli, D., and F. Panetta, 2003, "Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits," *American Economic Review*, 93(4), 1152–1172.

Freixas, X., S. Hurkens, A. D. Morrison, and N. Vulkan, 2004, "Interbank Competition with Costly Screening," Oxford Financial Research Centre Working Paper.

Freixas, X., and J.-C. Rochet, 1998, "Fair Pricing of Deposit Insurance. Is It Possible? Yes. Is it desirable? No," *Research in Economics*, 52(3), 217–232.

——— , 1999, *Microeconomics of banking*. Massachusetts Institute of Technology, Cambridge, MA, Fourth printing, 1999.

Fudenberg, D., and J. Tirole, 1991, *Game Theory*. MIT Press, Cambridge.

Gale, D., and M. Hellwig, 1985, "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, 52(4), 647–663.

Gehrig, T., and P.-J. Jost, 1995, "Quacks, Lemons, and Self Regulation: A Welfare Analysis," *Journal of Regulatory Economics*, 7(3), 309–325.

Goldstein, I., and A. Pauzner, 2005, "Demand-Deposit Contracts and the Probability of Bank Runs," *Journal of Finance*, 60(3), 1293–1327.

Gordy, M. B., and B. Howells, 2006, "Procyclicality in Basel II: Can We Treat the Disease without Killing the Patient?," *Journal of Financial Intermediation*, 15(3), 395–417.

Gorton, G., 1988, "Banking Panics and Business Cycles," *Oxford Economic Papers*, 40(4), 751–781.

Gorton, G. B., and G. G. Pennacchi, 1990, "Financial Intermediaries and Liquidity Creation," *Journal of Finance*, 45(1), 49–71.

Gorton, G. B., and A. Winton, 2003, "Financial Intermediation," in *Handbook of the Economics of Finance*. North-Holland, Amsterdam, pp. 431–552.

Goyal, V. K., K. Lehn, and S. Racic, 2002, "Growth Opportunities and Corporate Debt Policy: The Case of the U.S. Defense Industry," *Journal of Financial Economics*, 64(1), 35–59.

Greenbaum, S. I., and A. V. Thakor, 1987, "Bank Funding Modes: Securitization versus Deposits," *Journal of Banking & Finance*, 11(3), 379–401.

Gropp, R., and J. Vesala, 2004, "Deposit Insurance, Moral Hazard and Market Monitoring," *Review of Finance*, 8(4), 571–602.

Gu, C., 2007, "Herding and Bank Runs," University of Missouri Working Papers.

Hart, O. D., and J. Moore, 1994, "A Theory of Debt Based on the Inalienability of Human Capital," *Quarterly Journal of Economics*, 109(4), 841–879.

——— , 1998, "Default And Renegotiation: A Dynamic Model Of Debt," *Quarterly Journal of Economics*, 113(1), 1–41.

Hart, O. D., and J. Tirole, 1990, "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity*, 1990(special issue), 205–276.

Hauswald, R., and R. Marquez, 2002, "Competition and Strategic Information Acquisition in Credit Markets," Working paper.

——— , 2006, "Competition and Strategic Information Acquisition in Credit Markets," *Review of Financial Studies*, 19(3), 967–1000.

Heinemann, F., and G. Illing, 2002, "Speculative Attacks: Unique Sunspot Equilibrium and. Transparency," *Journal of International Economics*, 58(2), 429–450.

Hellmann, T. F., K. C. Murdock, and J. E. Stiglitz, 2000, "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?," *American Economic Review*, 90(1), 147–165.

Herring, R., 2002, "International Financial Conglomerates: Implications for Bank Insolvency Regimes," Wharton School Working Paper.

Hoggarth, G., R. Reis, and V. Saporta, 2002, "Costs of banking system instability: Some empirical evidence," *Journal of Banking* & *Finance*, 26(5), 825–855.

Holmstrom, B., and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112(3), 663–691.

———, 1998, "Private and Public Supply of Liquidity," *Journal of Political Economy*, 106(1), 1–40.

Honohan, P., and J. E. Stiglitz, 2001, "Robust Financial Restraint," in *Financial Liberalization: How Far, How Fast?*, ed. by G. Caprio, P. Honohan, and E. J. Stiglitz. Cambridge University Press, Massachusetts, pp. 26–47.

Hotelling, H., 1929, "Stability in Competition," *Economic Journal*, 39(153), 41–57.

House of Commons - Treasury Committee, 2008, *The Run on The Rock*. TSO (The Stationery Office), Norwich, HC 56-I, Fifth Report of Session 2007-08 - Volume I: Report.

Hubbard, R. G., K. N. Kuttner, and D. N. Palia, 2002, "Are There Bank Effects in Borrowers' Costs of Funds? Evidence from a Matched Sample of Borrowers and Banks," *Journal of Business*, 75(4), 559–581.

Iyer, R., and M. Puri, 2007, "Who Runs? The Importance of Relationships in Bank Runs," Working paper.

Jacklin, C. J., 1987, "Demand Deposits, Trading Restrictions, and Risk Sharing," in *Contractual Arrangements for Intertemporal Trade*. University of Minnesota Press, Minneapolis, MN, pp. 26–47.

Jacklin, C. J., and S. Bhattacharya, 1988, "Distinguishing Panics and Information-Based Bank Runs: Welfare and Policy Implications," *Journal of Political Economy*, 96(3), 568–592.

James, C., 1987, "Some evidence on the uniqueness of bank loans," *Journal of Financial Economics*, 19(2), 217–235.

Jayaratne, J., and P. E. Strahan, 1996, "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation," *Quarterly Journal of Economics*, 111(3), 639–670.

Jayaratne, J., and P. E. Strahan, 1998, "Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking," *Journal of Law & Economics*, 41(1), 239–273.

Jensen, M. C., 1986, "Agency Costs of Free Cash Flow, Corporate Finance and Takeovers," *American Economic Review*, 76(2), 323–339.

Johnson, S. A., 2003, "Debt Maturity and the Effects of Growth Opportunities and Liquidity Risk on Leverage," *Review of Financial Studies*, 16(1), 209–236.

Jones, D., 2000, "Emerging Problems with the Basel Capital Accord: Regulatory Capital Arbitrage and Related Issues," *Journal of Banking & Finance*, 24(1-2), 35–58.

Kashyap, A. K., R. G. Rajan, and J. C. Stein, 2002, "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking," *Journal of Finance*, 57(1), 33–73.

Kashyap, A. K., and J. C. Stein, 2004, "Cyclical implications of the Basel II capital standards," *Economic Perspectives*, 28(Q I), 18–31.

Keeley, M. C., 1990, "Deposit Insurance, Risk, and Market Power in Banking," *American Economic Review*, 80(5), 1183–1200.

Kelly, M., and C. Ó Gráda, 2000, "Market Contagion: Evidence from the Panics of 1854 and 1857," *American Economic Review*, 90(5), 1110–1124.

Kim, D., and A. M. Santomero, 1988, "Risk in Banking and Capital Regulation," *Journal of Finance*, 43(5), 1219–1233.

Kopecky, K. J., and D. VanHoose, 2006, "Capital Regulation, Heterogeneous Monitoring Costs, and Aggregate Loan Quality," *Journal of Banking & Finance*, 30(8), 2235–2255.

Košak, M., and P. Zajc, 2006, "The East-West Efficiency Gap in European Banking," in *Competition and Profitability in European Financial Services. Strategic, Systemic and Policy Issues*, ed. by M. Balling, F. Lierman, and A. Mullineaux. Routledge, UK, pp. 63–79.

Krasa, S., and A. P. Villamil, 1992, "Monitoring the Monitor: An Incentive Structure for a Financial Intermediary," *Journal of Economic Theory*, 57(1), 197–221.

Kroszner, R. S., 1999, "On the Political Economy of Banking and Financial Regulatory Reform in Emerging Markets," *Research in Financial Services*, 10, 33–51.

Kroszner, R. S., and P. E. Strahan, 1999, "What Drives Deregulation? Economics and Politics of the Relaxation of Bank Branching Restrictions," *Journal of Political Economy*, 114(4), 1437–1467.

Lederer, P. J., and A. P. Hurter, 1986, "Competition for Firms: Discriminatory Pricing and Location," *Econometrica*, 54(3), 623–640.

Leland, H., and D. Pyle, 1977, "Informational Asymmetries, Financial Structure, and Financial Intermediation," *Journal of Finance*, 32(2), 371–387.

Lummer, S., and J. McConnell, 1989, "Further Evidence on the Bank Lending Process and the Capital Market Responses to Bank Loan Agreements," *Journal of Financial Economics*, 25(1), 99–122.

Matutes, C., and X. Vives, 2000, "Imperfect Competition, Risk Taking, and Regulation in Banking," *European Economic Review*, 44(1), 1–34.

McAfee, R. P., and M. Schwartz, 1994, "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity," *American Economic Review*, 84(1), 210–230.

Melitz, M. J., 2003, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Merton, R. C., 1977, "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory," *Journal of Banking & Finance*, 1(1), 3–11.

Millon, M. H., and A. V. Thakor, 1985, "Moral Hazard and Information Sharing: A Model of Financial Information Gathering Agencies," *Journal of Finance*, 40(5), 1403–1422.

Mishkin, F. S., 1999, "Financial consolidation: Dangers and opportunities," *Journal of Banking & Finance*, 23(2-4), 675–691.

Morgan, D. P., 2002, "Rating Banks: Risk and Uncertainty in an Opaque Industry," *American Economic Review*, 92(4), 874–888.

Morrison, A. D., and L. White, 2005, "Crises and Capital Requirements in Banking," *American Economic Review*, 95(5), 1548–1572.

Myers, S. C., and N. S. Majluf, 1984, "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do not Have," *Journal of Financial Economics*, 13(1), 51–76.

Myers, S. C., and R. G. Rajan, 1998, "The Paradox Of Liquidity," *The Quarterly Journal of Economics*, 113(3), 733–771.

Nash, J., 1950, "The Bargaining Problem," *Econometrica*, 18(2), 155–162.

Nickell, S. J., 1996, "Competition and Corporate Performance," *Journal of Political Economy*, 104(4), 724–746.

Ó Gráda, C., and E. N. White, 2002, "Who Panics During Panics? Evidence from a Nineteenth Century Savings Bank," NBER Working Papers 8856.

Ongena, S., and D. C. Smith, 2000, "Bank Relationships: A Review," in *Performance of Financial Institutions*, ed. by S. A. Zenios, and P. Harker. Cambridge University Press, Cambridge.

Ortiz-Molina, H., and M. F. Penas, 2006, "Lending to Small Businesses: The Role of Loan Maturity in Addressing Information Problems," Forthcoming in *Small Business Economics.*

Penas, M. F., and H. Unal, 2004, “Gains in Bank Mergers: Evidence from the Bond Markets,” *Journal of Financial Economics*, 74(1), 149–179.

Pennacchi, G. G., 2005, “Risk-based Capital Standards, Deposit Insurance, and Procyclicality,” *Journal of Financial Intermediation*, 14(4), 432–465.

Petersen, M. A., and R. G. Rajan, 1995, “The Effect of Credit Market Competition on Lending Relationships,” *Quarterly Journal of Economics*, 110(2), 407–443.

——— , 2002, “Does Distance Still Matter? The Information Revolution in Small Business Lending,” *Journal of Finance*, 57(6), 2533–2570.

Peura, S., and J. Keppo, 2006, “Optimal Bank Capital With Costly Recapitalization,” *Journal of Business*, 79(4), 2163–2202.

Prat, A., and A. Rustichini, 2003, “Games Played Through Agents,” *Econometrica*, 71(4), 989–1026.

Raith, M., 2003, “Competition, Risk, and Managerial Incentives,” *American Economic Review*, 93(4), 1425–1436.

Rajan, R. G., 1992, “Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt,” *Journal of Finance*, 47(4), 1367–1400.

Rajan, R. G., and A. Winton, 1995, “Covenants and Collateral as Incentives to Monitor,” *Journal of Finance*, 50(4), 1113–1146.

Ramakrishnan, R., and A. V. Thakor, 1984, “Information Reliability and a Theory of Financial Intermediation,” *Review of Economic Studies*, 51(3), 415–432.

Repullo, R., 2004, “Capital Requirements, Market Power and Risk-Taking in Banking,” *Journal of Financial Intermediation*, 13(2), 156–182.

Rochet, J.-C., 1992, “Capital Requirements and the Behaviour of Commercial Banks,” *European Economic Review*, 36(5), 1137–1170.

Rothschild, M., and J. E. Stiglitz, 1976, “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *The Quarterly Journal of Economics*, 90(4), 630–649.

Salop, S. C., 1979, “Monopolistic Competition with Outside Goods,” *Bell Journal of Economics*, 10(1), 141–156.

Santos, J. a. A. C., 2001, “Bank Capital Regulation in Contemporary Banking Theory: A Review of the Literature,” *Financial Markets, Institutions & Instruments*, 10(2), 41–84.

Saunders, A., 2000, *Financial Institutions Management, 3rd edition.* McGraw-Hill, New York.

Scherr, F. C., and H. M. Hulburt, 2001, "The Debt Maturity Structure of Small Firms," *Financial Management*, 30(1), 85–111.

Schmeits, A. B., 2005, "Discretionary Contracts and Bank-Firm Relationships," OLIN Working Paper.

Segal, I., 1999, "Contracting With Externalities," *The Quarterly Journal of Economics*, 114(2), 337–388.

Segal, I., and M. D. Whinston, 2003, "Robust Predictions for Bilateral Contracting with Externalities," *Econometrica*, 71(3), 757–791.

Sengupta, R., 2007, "Foreign entry and bank competition," *Journal of Financial Economics*, 84(2), 502–528.

Shaked, A., and J. Sutton, 1982, "Relaxing Price Competition Through Product Differentiation," *The Review of Economic Studies*, 49(1), 3–13.

Sharpe, S. A., 1990, "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships," *Journal of Finance*, 45(4), 1069–1087.

Slovin, M. B., S. A. Johnson, and J. L. Glascock, 1992, "Firm Size and the Information Content of Bank Loan Announcements," *Journal of Banking & Finance*, 16(6), 1057–1071.

Smith Jr., C. W., and J. B. Warner, 1979, "On financial contracting: An analysis of bond covenants," *Journal of Financial Economics*, 7(2), 117–161.

Song, F., and A. V. Thakor, 2004, "Relationship Banking, Fragility, and the Asset-Liability Matching Problem," Forthcoming in *Review of Financial Studies*.

Sorge, M., and C. Zhang, 2007, "Credit Information Quality and Corporate Debt Maturity: Theory and Evidence," Working paper.

Spence, A. M., 1973, "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–374.

Stein, J. C., 2002, "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *Journal of Finance*, 57(5), 1891–1921.

Stiglitz, J. E., 1987, "Competition and the Number of Firms in a Market: Are Duopolies More Competitive than Atomistic Markets?," *Journal of Political Economy*, 95(5), 1041–1061.

Stiroh, K. J., and P. E. Strahan, 2003, "Competitive Dynamics of Deregulation: Evidence from U.S. Banking," *Journal of Money, Credit, and Banking*, 35(5), 801–828.

Stohs, M. H., and D. C. Mauer, 1996, "The Determinants of Corporate Debt Maturity Structure," *Journal of Business*, 69(3), 279–312.

Stulz, R. M., 1990, "Managerial Discretion and Optimal Financing Policies," *Journal of Financial Economics*, 26(1), 3–27.

Syverson, C., 2004, "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(61), 1181–1222.

Thakor, A. V., 1990, "Investment "Myopia" and the Internal Organization of Capital Allocation Decisions," *Journal of Law, Economics and Organization*, 6(1), 129–154.

Thisse, J.-F., and X. Vives, 1988, "On The Strategic Choice of Spatial Price Policy," *American Economic Review*, 78(1), 122–137.

Tirole, J., 2006, *The Theory of Corporate Finance*. Princeton University Press, Princeton and Oxford.

Tompson, 2004, "Banking Reform in Russia," OECD Economics Department Working Papers, No 410, OECD Publishing.

Townsend, R., 1979, "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21(2), 265–293.

VanHoose, D., 2006, "Bank Behavior under Capital Regulation: What Does the Academic Literature Tell Us?," Indiana State University Working Paper.

Vega-Redondo, F., 2003, *Economics and the Theory of Games*. Cambridge University Press, Cambridge.

Vives, X., 2001a, "Competition in the Changing World of Banking," *Oxford Review of Economic Policy*, 17(4), 535–547.

——— , 2001b, "Restructuring Financial Regulation in the European Monetary Union," *Journal of Financial Services Research*, 19(1), 57–82.

von Neumann, J., and O. Morgenstern, 1944, *Theory of Games and Economic Behavior. 1953 edition*. Princeton University Press, Princeton, NJ.

Von Thadden, E.-L., 1995, "Long-Term Contracts, Short-Term Investment and Monitoring," *Review of Economic Studies*, 62(4), 557–575.

——— , 1997, "The Term-Structure of Investment and the Banks' Insurance Function," *European Economic Review*, 41(7), 1355–1374.

——— , 1999, "Liquidity Creation through Banks and Markets: Multiple insurance and Limited Market Access," *European Economic Review*, 43(4-6), 991–1006.

——— , 2004, "Asymmetric Information, Bank Lending and Implicit Contracts: The Winner's Curse," *European Economic Review*, 43(4-6), 991–1006.

Winton, A., 1997, "Competition among Financial Intermediaries When Diversification Matters," *Journal of Financial Intermediation*, 6(4), 307–346.

Yorulmazer, T., 2003, "Herd Behavior, Bank Runs and Information Disclosure," New York University Working Paper.

# Author Index

# Samenvatting

Dit proefschrift bestaat uit theoretische bijdragen over financiële intermediatie en regulering. Het belangrijkste thema is de optimale regulering van banken, en kapitaalregulering in het bijzonder. De belangrijkste vraag die wordt geanalyseerd is welk effect regulering heeft op concurrentie in het bankwezen, en, vice versa, wat het effect van concurrentie is op de effectiviteit van kapitaalregulering. Deze vraag wordt geanalyseerd in een industriële organisatie setting waarin zowel inter-bank concurrentie als nieuwe toetreding een rol speelt. De meest opvallende uitkomst van de analyse is dat hogere kapitaaleisen toetreding kunnen bevorderen. In essentie wat er gebeurt, is dat hogere kapitaaleisen ontmoedigend werken voor lagere kwaliteit banken en via dit opschoningeffect ("cleansing") concurrentie tussen valide spelers attractiever maakt. Andere kwesties die worden geanalyseerd in het proefschrift hebben betrekking op de vraag waarom banken het verstrekken van leningen en het accepteren van deposito's combineren, en hoe concurrentie tussen bestaande banken de karakteristieken van een optimale bankleningovereenkomst beïnvloeden. Het proefschrift omvat ook een overzichtshoofdstuk over de fundamentele bestaansredenen voor banken, inclusief overwegingen die aangeven waarom regulering in deze bedrijfstak zo belangrijk is. Dit laatste komt ook tot uitdrukking in een afzonderlijk hoofdstuk over optimaal reguleringsbeleid.