# An Economic Analysis of Bank Risk-Taking: Stability, Deposit Contracts and Lending Relationships

by

ELENA CARLETTI

The London School of Economics and Political Science

Thesis submitted in partial fulfilment of the requirements of the degree of

**Doctor of Philisopy**

at the University of London

*A Philipp*

2

## Abstract

This thesis provides an economic analysis of bank risk-taking, addressing the relation between stability and competition, the efficiency of demandable debt as an incentive device for bankers, and the interaction between the structure of credit relationships, bank monitoring and loan rates.

Chapter 1 reviews the literature on stability, regulation and competition in banking. The survey is organised around two dimensions of instability: vulnerability to runs and panics, and excessive risk taking. It turns out that the existing literature largely ignores the impact of competition on stability and on the optimal regulatory design. The very few models addressing these issues do not provide conclusive results.

Chapter 2 goes deeper on the phenomenon of bank runs. A unified framework is presented within which the main literature is outlined and compared.

Chapter 3 develops a model that analyses both the benefits and the costs of market discipline as an incentive device for bankers. It is shown that demandable debt, by allowing for the possibility of runs, can induce bankers to monitor their projects. However, market discipline comes at a cost. Since depositors are not equally informed about bank future solvency, they may commit mistakes in their withdrawal decisions, forcing the closure of a solvent bank or permitting the continuation of an insolvent one.

Chapter 4 turns the attention to the lending side, developing a model of overlapping moral hazard problems between banks and firms. The aim is to study how the number of bank relationships affects banks' monitoring decisions and how these affect loan rates and firms' choice between single and multiple relationships. It is shown that multiple lenders monitor less than a single lender, but they don't necessarily require higher loan rates. The firm's choice between single and multiple relationships is not univocal, depending on the severity of bank moral hazard as compared to firm moral hazard.

## Acknowledgments

I am deeply indebted to my supervisor Margaret Bray. Our weekly meetings have been a 'fixed point' during the years of work. She followed my dissertation step by step, giving me invaluable advice. I greatly appreciated her continuous encouragement and support, and the patience with which she answered my endless questions. I am also grateful to my supervisor Charles Goodhart, especially for teaching me to go beyond the mathematical aspects of the models. Also, I would like to thank Vittoria Cerasi, Marco Pagano and Michele Polo for having initiated me to graduate studies and supported me throughout.

For discussions and insights at various stages, I would like to express my gratitude to Sudipto Bhattacharya, Vittoria Cerasi, Marco Da Rin, Vincenzo Denicolò, Serdar Dinc, Antoine Faure-Grimaud, Xavier Freixas, Paolo Garella, Ulrich Hege, Haizhou Huang, Charlotte Ostergaard, Xavier Vives. I am also appreciative of my fellow students, including Markus Brunnermeier, Giulio Fella and Philip Kalmus for various suggestions, and Eric Ben-Hamou, Simon Emrich and Alex Muermann, for their help on mathematical and computer questions. I could not have written chapter 4 without the numerous discussions on 'matematica' with Marco Polenghi and on game theory with Jorn Rothe. I am also very grateful to them for their friendship and their support from the very first steps of the dissertation.

I am grateful to several institutions for their support. I benefited from the stimulating research environment at the FMG, for which I would like to thank the director David Webb and also Bob Nobay, for the comments he gave me on various parts of the thesis and his support over the years. I am also indebted to the "Autorita' Garante della Concorrenza e del Mercato" and, especially, to Michele Grillo and Pier Luigi Parcu, for giving me the possibility to continue my studies. Financial support from Bocconi University and Ente "L. Einaudi" is gratefully acknowledged.

Many friends encouraged me continuously. I would like to mention, in particular, Adriana Barone, Alessio De Vincenzo, Valeria Gaudiello, Lisa Grazzini, Daniela Novi, Barbara Marchitto, with whom I shared most of the experience in London, Osmana Raie and Andrew Ellul, Monica Paiella and Anna Maria Reforgiato-Recupero, who read preliminary drafts of the chapters. I am also deeply indebted to my parents, who encouraged me in all circumstances.

My greatest debt, however, is to Philipp, to whom this thesis is dedicated.

# Contents

# List of Tables

**Chapter 2**



**Chapter 3**

# List of Figures

# Introduction

The role of financial intermediation is one of the fundamental issues in theoretical economics and finance. The efficiency of the process through which savings are channelled into productive activities is key for growth and general welfare. In some economies, often in an earlier stage of their development, banks perform most of this financial intermediation through loan contracts. In other economies, in particular those with favourable legal systems, capital markets allow more firms to tap investors directly through the issuance of debt securities. However, whatever the relative orientation of the financial systems, banks always play a special role in financial intermediation. They can be regarded as a 'vehicle' to solve problems of asymmetric information between lenders (ultimately savers) and borrowers (ultimately firms investing for the purpose of producing goods). They specialise in assessing the relative viability and profitability of the different projects put forward by entrepreneurs. Hence, they will be particularly involved in the type of firm projects in which the informational disadvantages of 'savers' are relatively high.

Banks are not only special in that they 'produce' information about investment projects but also in the way they raise their funds. In fact, they rely to a significant extent on (many small) short-term demandable deposits, which they pool and then invest in long-term loans to production firms. This maturity mismatch between assets and liabilities makes banks play the additional role of providers of liquidity to depositors but, also, exposes them to the possibility of runs (and systemic crises). This vulnerability to runs (and to systemic crises) represents the source of bank instability originating on the liability side of the balance sheet. A second source of instability relates to bank risk-taking on the asset side. Because of their substantial financing from many small, relatively uninformed depositors and an often-existing public safety net in response to the previously mentioned vulnerability, banks can be particularly prone to taking on 'excessive' risk in the choice of which projects to finance.

Bank stability, both on the liability and on the asset side, is influenced by a variety of (partly interlinked) factors, which can be decomposed into those related to the endogenous functioning of private markets and those more related to public policies. Within the former category bank stability is influenced, on the one hand, by macroeconomic fluctuations and, on the other hand, by microeconomic market structures and the competitive environment. For example, it matters whether the financial system is more bank-based or more market-based, how large is the number of banks and their firm lending relationships, how interbank lending relationships are structured, how important is the 'special' deposit financing as compared to equity or bond financing, and how important is the degree of information asymmetries between both creditors and banks, as well as between borrowers and banks.

Regarding public policies, one can again distinguish between the macroeconomic and the microeconomic side, but also between 'ex ante' and 'ex post' policies. 'Ex post' policies relate to crisis management, once a financial instability has materialised, whereas 'ex ante' policies relate to the maintenance of a stable environment in the first place. On the macroeconomic side, bank stability is particularly affected by monetary policies, but also by fiscal policies (both 'ex ante' and 'ex post'). On the microeconomic side, bank stability is influenced 'ex ante' by financial regulation, prudential supervisory practices and payment system oversight, special competition policies for the banking sector (e.g. special regulations for market entry, mergers and acquisitions or collusive agreements) and 'ex post' by deposit insurance arrangements (however, in some countries they are private or semi-private) as well as lender of last resort interventions to individual banks ('micro LOLR').

The present thesis discusses several key aspects of bank stability among those enumerated above, following primarily a theoretical approach. The main emphasis is on the microeconomic side. It is organised in four chapters, whose content and main conclusions I summarise below.

Chapter 1 provides an extensive review of the theoretical literature on bank stability, regulation and competition. The survey is divided into two parts, each of them organised around the two dimensions of instability previously described: vulnerability to runs and systemic crises (bank fragility on the liability side) and excessive risk taking (bank moral hazard problem on the asset side).

The first part of the chapter focuses on the problem of bank stability and the measures that aim to preserve it. It starts with reviewing selected contributions on the problem of bank fragility (both individual runs and systemic crises) and the need for public intervention, with particular emphasis on deposit insurance schemes and lender of last resort facilities, i.e. ex post micro policies. Then, it analyses the moral hazard problem induced by ill-designed public regulation and safety net arrangements and discusses the proposals advanced to ameliorate them. These are divided into two categories: The first, defined as regulation-oriented, includes proposals for the reform of deposit insurance pricing and bank closure policy. The second category, defined as market-oriented, relates to the proposal of scaling back deposit insurance in order to reintroduce market discipline in the form of uninsured depositors' monitoring.

The second part of chapter 1 turns the attention to the impact of competition on stability (i.e. the micro market side), both from a positive and a normative angle. The main scope is to examine whether the presumed negative impact of competition on bank stability (in both senses) finds support in the theoretical literature. It turns out that the existing models largely disregard these issues, focusing on situations of perfect competition or total monopoly. The very few models that endogenise aspects of industrial organisation in banking are then reviewed. At the present stage, however, the results are still far from being conclusive.

Chapter 2 adds an in-depth literature analysis of the bank run phenomenon, adopting a uniform modelling framework within which the main theoretical contributions can be outlined and compared. Two types of theories are distinguished: models that focus on the role of banks as providers of flexibility to depositors in their timing of consumption and models that focus on the agency relation between the banker and his depositors. These two categories, denoted as 'first generation' literature and 'second generation' literature, respectively, provide different explanations for the occurrence of bank runs and draw different conclusions concerning their efficiency or inefficiency. Depending on the reasons behind depositors' decisions to withdraw prematurely, first generation models see runs as irrational events ('sunspot' approach) or as depositors' rational responses to the arrival of poor information on the bank's future solvency ('asymmetric information' approach). Information-based runs are efficient if based on perfect information, since they force the liquidation of valueless assets. Second generation models consider the threat of runs as discipline device for bankers' moral hazard problems. Runs are information-based and their efficiency depends on whether depositors base their withdrawal decisions on noisy or perfect information.

In chapter 3, I develop a model that analyses both the benefits and the costs of demandable debt, one of the micro factors influencing bank stability from the liability side. Expanding on the 'second generation' literature described in chapter 2, a moral hazard problem between a banker and his depositors is analysed. I investigate whether demandable debt, by allowing the possibility of runs, can constitute an efficient incentive device.

A banker needs to raise deposits to invest in an illiquid project. The project is risky and its success probability is higher when the banker (privately) decides to monitor than when he does not. Deposit contracts can be either of a more long-term nature or demandable at any time. With the former, depositors can demand a predetermined repayment only at maturity. With the latter, they have the right to withdraw at any point in time. Starting from a situation with standard debt in which the banker finds it optimal not to monitor the project, the analysis shows that demandable debt can induce him to change to monitoring. This is because this contract allows depositors to discipline the banker by triggering a run whenever they (rationally) expect the value of the bank's assets to be low.

However, market discipline can come at a cost. When depositors are heterogeneously informed on the value of the bank's assets, those who are uninformed may commit 'mistakes' in their withdrawal decisions. They may either leave their deposits at the bank when prospects are low, or withdraw them when prospects are high. The former error leads to the continuation of an insolvent bank; the latter error can induce informed depositors to ignore their information and join the withdrawal queue. When this is the case, an inefficient information-based run occurs, which forces a solvent bank into liquidation. The inefficiency of market discipline originates within a single bank and does not depend on the imposition of a sequential service constraint rule. Rather, it is due to both the informational externality of depositors' withdrawal decisions and the illiquidity of the bank's assets.

As described above, banks are valuable because of their superior ability to reduce the informational asymmetries that plague direct lending relative to other financial institutions. The extent to which banks are effective in playing this role, however, depends on their incentives to monitor. One main factor that influences banks' monitoring decisions is the structure of credit relationships. Banks can choose to monitor their borrowers with different intensity depending on whether they are the only or one of several lenders to the firm. These decisions affect, in turn, the cost of loans and the firms' choice between single and multiple credit relationships. Chapter 4 addresses these issues by developing a one-period model with overlapping moral hazard problems between banks and firms.

An entrepreneur seeks bank financing to undertake a project. He operates subject to a moral hazard problem. He may prefer not to exert effort and enjoy a private benefit instead of being diligent and increasing the success probability of the project. Banks can force the entrepreneur to behave well through monitoring. However, monitoring is not contractible and is costly, exhibiting diseconomies of scale. This creates a bank moral hazard problem: Banks choose the monitoring intensity as to maximise their profits, which depend on the size of monitoring costs, on the entrepreneur's behaviour and on the number of lenders. The firm can borrow either from one bank or from two banks. The equilibrium levels of monitoring and the loan rates depend on which structure of bank relationships the firm chooses.

The analysis shows that multiple lenders always exert a lower level of discipline than a single lender. This is because they face duplication of effort and sharing of benefits in monitoring. However, as a consequence of diseconomies of scale in monitoring, multiple lenders do not necessarily require a higher loan rate.

The entrepreneur's choice between single and multiple relationships is not univocal. It depends rather on the relative severity of bank moral hazard as compared to firm moral hazard. In general, he prefers a single lender when bank moral hazard is weak and multiple lenders when bank moral hazard is strong. For intermediate values of bank moral hazard, the firm's choice depends on its own inclination for moral hazard. The greater the private return for the entrepreneur when he does not exert effort, the more likely the firm will find it optimal to borrow from two banks, although it may require a higher loan rate than with a single bank.

# Chapter 1

# Stability, Regulation and Competition in Banking

## 1.1 Introduction

Bank stability has always been a major public concern. Episodes of bank runs and widespread bank failures have plagued the history of many countries, motivating the introduction of bank regulation and its successive modifications.

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

The potential instability of the banking system is the fundamental rationale for the introduction and the development of regulation. In most countries the institution of the safety net, namely a form of deposit insurance and a lender of last resort, has been the response to bank fragility; the implementation of important new regulatory legislation in the last decades and the wide ongoing debate over regulatory reforms on the academic front are the reactions to the solvency crises experienced by banks in the 1980s and 1990s.

The debate on financial stability and regulatory reforms concentrates on instruments that prevent fragility and ameliorate incentive problems, largely ignoring the effects that different market structures and competition among banks can have on the safety of the banking sector. Curiously, the literature on banking stability and regulation focuses mostly on situations of either perfect competition or total monopoly.

Competition in banking has traditionally been seen with suspicion, because of a general feeling that it exacerbates both problems of fragility and of excessive risk taking. These presumed destabilising effects have justified the imposition of measures to prevent excessive competition. Many countries have regulated the structure of their banking sector for a long time with interest rate ceilings, entry barriers, ownership limits and asset restrictions. The wave of liberalisation and deregulation that occurred in the 1980s has not been entirely smooth. It has been described as a major cause of the crises that happened in the last two decades. As a result, prudential regulation has been strengthened and harmonised across countries and deposit insurance has been reformed and extended.

These developments have renewed the debate on the presumed trade-off between competition and stability. They have induced scholars to introduce "industrial organisation aspects" in the analysis of bank stability and in the design of the optimal regulatory policy. At the present stage, however, the results are still far from being conclusive.

The object of this chapter is to examine the academic literature on bank stability, both from a positive and from a normative angle. In the first part of the chapter, I review selected contributions addressing the problem of bank fragility and the need for public intervention. Then, I analyse the distortion that public arrangements, such as deposit insurance and lender of last resort, can create in terms of bank excessive risk taking (the so called moral hazard problem) and the proposals suggested to ameliorate it. In the second part of the chapter, I survey the recent literature on competition and stability, both regarding fragility and excessive risk taking, and regarding the implications of different market structures for the optimal regulatory policy. While I focus the survey on the theoretical aspects of the debate on stability, regulation and competition, I will complement the theory with some empirical findings.

The chapter is organised as follows. Section 2 focuses on the specificity and fragility of the banking system, stressing the difference between individual banks' vulnerability to run (section 2.1) and systemic crises (sections 2.2 ). Section 3 discusses the introduction of public interventions, in the form of deposit insurance schemes and lender of last resort function (the so-called safety net), which aim to prevent the fragility of the system. Section 4 describes the moral hazard problem in banking as a consequence of a de facto complete deposit insurance. Section 5 analyses theoretical contributions on the lively debate prompted by recent banking distress, reviewing proposals for the reform of deposit insurance pricing (section 5.1), on the importance of regulatory monitoring and bank closure policy (section 5.2) and on the role of market discipline (section 5.3). The subsequent sections concentrate on the link between competition and stability: Section 6 presents recent contributions on the effect of competition on bank fragility; section 5 describes those relating to the interrelations between market structure and excessive risk taking; section 8 focuses on the importance of market structure for regulatory reforms. Section 9 concludes.

## 1.2 The Specificity and Fragility of the Banking System

The theory of financial intermediation has provided disparate reasons for the uniqueness of banks and the need of regulation. Banking theory has aimed firstly to explain the mere existence of banks and, then, their vulnerability to individual runs and systemic crises, seen as inherent to the nature of banking itself.

In the 1970s, the reasons for the existence and the main functions of intermediaries were traced back in the reduction of transaction costs,[1] in the optimisation of portfolios and maturity transformation,[2] and in the improvement of collateral and contractual clauses.[3]

In the 1980s, greater importance was attributed to informational asymmetries in explaining the failure of competitive markets and the emergence of intermediaries. In a market characterised by asymmetric information, intermediaries can improve social welfare: They have economies of scale in producing information and sending signals on project quality, which is private information of entrepreneurs,[4] they have technological economies of scale in monitoring their borrowers[5] and provide insurance to depositors who are uncertain in their timing of con-

[1] See Benston and Smith (1976).
[2] See Baltensperger (1980), Deshmukh et al. (1983).
[3] See Smith and Warner (1979).
[4] See Leland and Pyle (1977).
[5] See Diamond (1984).

sumption.[6] These characteristics are the key elements of bank specificity: They explain banks' advantages relative to an autarkic situation in which individual investors act directly but, at the same time, they expose intermediaries to the risk of fragility.

In the following, I describe the problem of bank fragility, making a distinction between individual and systemic crises. Individual fragility concerns single intermediaries' vulnerability to runs. Systemic fragility relates to the failure of a considerable number of institutions or of the system as a whole. The latter can be the result of a negative aggregate shock and/or of the propagation of negative individual shocks.

### 1.2.1 Individual Banks' Vulnerability to Runs

The key elements for the occurrence of individual runs are the maturity transformation that banks operate by investing short-term deposits in long-term assets, and the informational asymmetries existing between banks and their clients. Banks offer deposit contracts, which allow lenders to withdraw a nominal and fixed amount on demand. Insofar as banks use these deposits for illiquid and/or risky loans, there is the possibility of a liquidity crisis. If the proportion of depositors that withdraw their deposit early exceeds that expected by the bank, the intermediary may not be able to fulfil the withdrawal requests even if it sells a part or all of its assets. When this is the case, a bank run has origin: Depositors' massive early withdrawal demands force the bank to sell all its assets and to close down.

The combination of maturity transformation, that is short and fixed liabilities against long and illiquid assets, and of a liquidity premium, that is the costly liquidation of long term assets before maturity, explains the role of intermediaries as liquidity providers and, at the same time, their fragility. Bank runs can be irrational or information-induced events, depending on the different reasons behind the unexpected early withdrawals by depositors and on the type of asymmetric information between banks and their clients.

Diamond and Dybvig (1983) show that runs, when they occur, are an unfortunate and undesirable side effect of a contract whose only purpose is to provide consumption flexibility to depositors. They depict bank runs as multiple-equilibrium phenomena, which are exclusively related to the illiquidity of bank assets. If, for any reason, depositors loose confidence in future

[6] See Bryant (1980) and Diamond and Dybvig (1983).

bank solvency, they start to withdraw early and the bank is emptied of funds. As a consequence, the risk-sharing mechanism, achievable when depositors leave their funds at the bank until asset maturity, breaks down and welfare is reduced. Given the deterministic structure of bank asset returns, the run emerges as a bad Nash equilibrium in which depositors start withdrawing prematurely because they just believe others will do so.

Thus, in the multiple-equilibrium approach, runs are a self-fulfilling rational coordination failure due to "sunspots". An alternative explanation for the occurrence of bank runs is that they are tied to changes in fundamental variables rather than to unpredictable variables. A negative shock on bank asset increases the probability that banks are unable to meet their commitments. If depositors anticipate the impending shock, they try to withdraw their funds and force the closure of the bank.[7] Jacklin and Bhattacharya (1988) provide a formal support to this alternative explanation.[8] Introducing a random structure of bank asset returns, they show that bank runs are information-induced, in that they are depositors' rational responses to the arrival of negative information regarding the state of bank investment.

The relationship between the bank and its depositors is now characterised by a bilateral asymmetric information: The bank knows depositors' aggregate consumption preferences but cannot distinguish the type of each depositor withdrawing early; further, depositors do not know the value of bank asset returns. At the interim period, a consumption shock is realised and some depositors turn out to be impatient; at the same time, some other depositors receive a signal related to the future bank asset value. If this signal is adverse enough, informed depositors join the queue of depositors withdrawing for consumption needs. Given the illiquidity of bank long-term assets, a rational, information-based run originates.

In a context characterised by shocks to bank asset returns, to the fraction of depositors that receive information on bank asset in the interim period and to the proportion of depositors wishing to withdraw early, Chari and Jagannathan (1988) show that bank runs can be both information-induced and panic phenomena. In other words, runs can happen either when some depositors have received negative information about bank future solvency or when nobody has received any signal. Panic runs are modelled as a signal-extraction problem: Uninformed

[7]Empirical evidence shows that most of the observed runs were tied to changes in fundamental variables rather than to unpredictable variables. See, for example, Gorton (1988).

[8]The pioneer but informal model of information-induced bank run is Bryant (1980).

depositors know that other depositors may be informed on future bank asset returns and they try to infer the state of bank solvency from the size of the withdrawal queue at the bank. Since the proportion of depositors withdrawing for consumption reasons is not observable, uninformed depositors may not be able to distinguish if a long withdrawal queue is due to some informed depositors receiving a negative information or only to a large proportion of agents desiring to consume early. A panic run generates when uninformed depositors confuse a high liquidity shock with fear of insolvency, that is when they withdraw although no one is informed about the bank future solvency.

To sum up, individual bank runs result from either/both a coordination failure among depositors or/and an expectation of poor performance of the bank. In terms of efficiency, the sunspot runs are clearly inefficient as they drive a solvent institution into failure; the information-based and the panic runs may be efficient. The former are efficient if depositors have correct information on the state of bank solvency but inefficient if depositors withdraw mistakenly on a solvent institution; the latter are efficient or inefficient depending on whether the bank forced into liquidation would turn to be solvent or insolvent if the run did not take place.

### 1.2.2 Systemic Crises: Aggregate Shock or Propagation of Individual Distresses

A major concern in banking is the occurrence of a systemic crisis, that is the occurrence of massive simultaneous bank failures. The risk of systemic crises is considered as the main rationale for policy interventions. Given the importance of the debate, among both academics and policymakers, centred on the terms systemic crises and systemic risk, I find it useful to provide first some conceptual definitions and then review the literature accordingly.

As recently argued by De Bandt and Hartmann (1999), the term systemic crisis can be given a narrow and a broad interpretation. The former refers to the situation in which the failure of one bank, or even only the release of bad news about its state of solvency, leads in a sequential fashion to the failure of numerous other banks or of the system as a whole. The latter includes also the simultaneous failure of many banks, or the crash of the system as a whole, as a result of a generalised adverse shock. In other terms, a systemic crisis in the narrow sense occurs when an idiosyncratic shock to a single institution propagates through the system causing a chain of subsequent failures. Differently, a systemic crisis in the broad sense may also be the direct consequence of an aggregate shock, which hits many institutions simultaneously. In both cases, a necessary condition for the occurrence of a systemic crisis is that the institutions affected by the shock, either in a first round (in case of an aggregate shock) or in a second round (in case of the propagation of an idiosyncratic shock), fail as consequence of the shock itself. Therefore, it cannot be identified as systemic crisis, neither in the narrow nor in the broad sense, the circumstance in which the occurrence of a shock, either idiosyncratic or aggregate, has adverse effects on the institutions affected but does not provoke the failure of any of them.[9]

Given the above definition of systemic crisis, the term systemic risk can be denoted as the risk, or the probability, that a systemic crisis, either in the narrow or in the broad sense, occurs.

In the following, I concentrate on systemic crises in the narrow sense and review the related literature. The extent to which the mechanism for the propagation of an idiosyncratic shock through the system can be explained has important policy implications. The existence of a systemic risk in the narrow sense is indeed considered as the most relevant motivation for central banks' intervention.[10]

### The Contagious Run and the Domino Effect

As already observed, the concept of systemic crisis in the narrow sense refers to the disruptive situation caused by the propagation of failures from one bank to the other. This strong spillover effect, defined as contagion, can take place through two channels: the information channel and the credit channel.[11] The former refers to the mechanism through which a run on a single bank causes runs on other banks (contagious run). The latter refers to the propagation of the difficulties faced by a bank to others linked to the failing one through the payment system and/or interbank markets (domino effect).

A contagious run is a term used to describe the spread of the effects of a run from one

[9] For example, it is not a systemic crisis the situation in which the release of negative information about one institution or its failure leads to negative abnormal returns on the stock of other banks without causing their failure.

[10] Note however that in practise it can be difficult to distinguish the precise source of a systemic crisis as macroeconomic shocks and propagation of failures can be interwined.

[11] See, for example, Schoenmaker (1996a).

or more banks to others. In particular, it is used to describe the situation in which a run at a particular intermediary makes depositors at many other institutions withdraw massively. The banking literature has identified two types of contagious run depending on the reasons behind the propagation of depositors' loss of confidence from one institution to the other: pure (industry specific) contagion and noisy (firm specific) contagion. Pure contagion arises when negative information about one bank, such as fraud or low returns on specific risky assets, adversely affects all other banks, including those that have nothing in common with the first one. Noisy or firm specific contagion occurs when the failure of one bank reveals a bad, even if noisy, signal about other banks with common characteristics. In other words, if a run on a bank occurs, depositors at banks with similar asset and liability structure to the failing one fear that their banks are also vulnerable to the same economic shock and demand their deposits back. The more similar the banks are in size, location and markets served, the more likely it is that a greater number of banks will be affected and the greater the intensity of the contagion will be.

The two types of contagious run resemble the two different explanations for the emergence of an individual bank run. The mechanism underlying the change in depositors' beliefs about future solvency across banks is similar to the one underlying the loss of confidence by depositors within a single bank. A contagious run is the negative externality generated by the occurrence of an individual run, which can be irrational or information-induced. Indeed, industry-specific contagion is sometimes referred to as non-informational contagion while bank-specific contagion is usually referred to as informational, rational contagion.[12]

The second propagation mechanism of individual difficulties refers to the credit channel. The propagation of bankruptcy from one financial institution to another through the interbank market and/or payment systems is seen nowadays as the major source of systemic risk. To the extent that interbank relations are neither collaterized nor insured against, an institution's distress may trigger a chain of subsequent failures. In particular, institutions linked to the failing one may incur a liquidity or an insolvency problem depending on the intensity of the linkage and on the shock correlation in the system.

Payment systems represent the most important interrelations among banks. Their internal

[12] In this sense, the industry-specific contagion is like the "sunspot phenomenon" described in Diamond and Dybvig (1983) and the bank-specific contagion is information-induced like the run in Jacklin and Bhattacharya (1988).

arrangement determines how individual shocks propagate and therefore the severity of the risk of contagion. Depending on the timing and the methodology of settlement, payment systems can be classified in net settlement systems (only net balances are settled and at a certain point in time), pure gross systems (payments between members are settled without netting and a certain point in time), real-time gross systems (payments between members are settled without netting and occur immediately for every transaction) and correspondent banking (payments are settled bilaterally between a correspondent bank and members of a group of small or foreign banks). Net systems incorporate a lower risk of contagion than pure gross systems but a higher risk than real-time gross systems.

The two vehicles of contagion, the information channel and the credit channel, can work in conjunction as well as independently. In principle, a domino effect can take place even without a contagious run and a contagious run does not necessarily require the existence of interrelations among banks. In most cases, however, a systemic crisis (in the narrow sense) is the result of the propagation of an individual failure through both channels.

Surprisingly, the risk of contagion has received attention in academic research only very recently. The former models of system-wide runs, which consider the banking sector as a whole cannot explain the risk of contagion but only generalised crises.[13] The analysis of the propagation mechanism requires multiple banking systems.

In the following, I review the recent models of contagion. I distinguish between contributions which model contagion through the information channel only (Chen (1999)), those that model contagion only through the credit channel (Rochet and Tirole (1996b)) and those that model the propagation of single failures through both the two channels. This last class of models can be further divided in models of contagious runs and domino effect through the interbank market (Aghion, Bolton and Dewatripont (1999), Allen and Gale (2000)) and models of contagious runs and domino effect through payment systems (Freixas and Parigi (1998), Freixas, Parigi and Rochet (1999)).

Chen (1999) analyses the contagious nature of runs in a model where banks do not interact

[13]The described models of individual bank runs (Diamond and Dybvig (1983), Jacklin and Bhattacharya (1988), Chari and Jagannathan (1988)) as well as Allen and Gale (1998), Alonso (1996), Postlewaite and Vives (1987) can all be interpreted as models of system-wide crises, that is of systemic runs caused by an aggregate shock.

through the payment system or the interbank market. He shows that runs at some banks may cause panic among depositors and generate runs at other banks. Depositors at one bank are heterogeneously informed about the outcome of the bank's investments. Some depositors receive perfect interim information on the bank's prospects while some others remain uninformed. The informed depositors have an advantage over the uninformed depositors as they can withdraw earlier when they receive bad bank-specific information and not be rationed from the sequential service constraint. Given this, the uninformed depositors have incentives to respond to any information available in the economy before the bank-specific information is revealed. Failures of other banks can be one information source. A large number of bank failures implies that the prospects of the remaining banks are likely to be poor. Although this information is noisy, uninformed depositors may respond to it and withdraw, thus forcing informed depositors to withdraw as well. When this is the case, a noisy (firm specific) contagion occurs. The contagious nature of runs relies on the imposition of the sequential service constraint rule in depositors' repayment and on the information externalities generated by banks' failures.[14]

Rochet and Tirole (1996b) address the issue of contagion in a model of interbank lending, in which they emphasise the trade-off between the risk of propagation of individual bank failures and peer monitoring among intermediaries. Interbank lending arises because of banks heterogeneity: Some banks are good at collecting deposits but have poor investment opportunity while some others have plenty of investment opportunities. The former have then an incentive to lend to the latter and also to monitor them. This creates the scope for interbank lending and scope economies, and represents the potential source of systemic risk. If the borrowing bank incurs a distress because of a liquidity shock, the lending bank may be affected and be forced to shut down. In particular, the lending bank's continuation depends on the liquidity shock it may encounter and the realised profit or the loss generated by the interbank loan which, in turn, is a function of the borrowing bank's outcome. The higher the liquidity shock faced by the borrowing bank, the more likely the closure of the lending bank.[15]

Aghion, Bolton and Dewatripont (1999) focus on the propagation of runs in a model where

[14] Chen (1999) addresses also the issue of whether bank runs are an efficient incentive devices for bankers' moral hazard problems, as I will discuss in section 5.3.

[15] As stressed by the authors, the exact details of contagion occurence depend on a number of institutional features, such as how the borrowing bank meets the liquidity shock and priority rules on borrowing bank's profits.

multiple banks interact in the interbank market. Banks are subject to a liquidity shock that makes them unable to repay depositors. In order to prevent depositors from running, illiquid banks can engage in interbank lending. If the system has enough liquidity, no failures occur: Illiquid banks are bailed out by their liquid counterparts and all depositors are repaid. When an individual bank fails, a contagious run arises: Depositors at other banks interpret the failure of a specific bank as signal of global illiquidity in the system and precipitate a run. The presence of an interbank market reduces the likelihood of failure of an individual bank but, at the same time, may trigger the propagation of runs.

The structure of interbank markets is the key factor for the risk of contagion also in Allen and Gale (2000). Banks hold inter-regional claims (deposits) on other banks to insure against liquidity preference shocks. If there is no aggregate uncertainty, the banking system is stable and the first best allocation of risk sharing is achieved. If, on the other hand, there is an excess aggregate demand for liquidity, the financial linkages among banks may turn out to be a disaster: In order to provide more consumption to depositors and to avoid the liquidation of long-term assets, banks start to withdraw deposits in other regions. This mutual liquidation denies liquidity to the troubled region that may then experiment a run. Depending on the structure of the cross holdings of deposits among banks, the individual crisis propagates through the economy. If regions are well connected (complete interbank market), the contagion is avoided. If connections among regions are limited (incomplete interbank market) and liquidity shocks are strong enough, then contagion may arise.

Contagious run through the payment system is the focus of Freixas, Parigi and Rochet (1999). They construct a model in which both liquidity and solvency shocks may affect banks located in different locations. Geographical consumption preferences cause liquidity shocks: Depositors may want to consume in a location different from where they have their deposits. They can satisfy their geographical consumption needs either by withdrawing their funds and transferring cash (holding of liquidity assets) or by transferring deposits from one bank to another (payment system). When banks are subject only to liquidity shocks, two equilibria are possible. In one equilibrium, depositors do not run and payment systems are efficient in reducing the opportunity costs of holding liquid assets (credit line equilibrium). In the other equilibrium, all depositors run and banks have to liquidate all their assets (speculative gridlock equilibrium).

In case of both liquidity and (idiosyncratic) solvency shocks, the stability of the banking system depends on the architecture of payment flows: The closure of an insolvent institution is less likely to propagate to the entire system when payment systems are well diversified, that is when credit lines are uniformly distributed among banks. However, diversified systems are more unstable with respect to the insolvency of an institution, that is they are less capable to absorb losses of insolvent active banks without generating systemic withdrawals (the so-called resiliency).

In a former paper, Freixas and Parigi (1998) tackle the question of the optimal design of payment systems. The focus is on the trade-off in terms of risk of contagion and efficiency associated with real-time gross settlement and net settlement. The framework is a Diamond-Dybvig type model enriched with stochastic investment returns and geographical consumption preferences. Payments across banks located in different regions can be made either by directly transferring liquidity (gross system) or by transferring claims on the assets of the bank in the other location (net system). The gross system entails high liquidity costs but it is free of contagion; the net system economises on liquidity but exposes banks to contagion because of the transfer of asset claims from one location to the other.

On the empirical level, studies aiming to quantify the likelihood and the intensity of contagion in the banking sector appear controversial. This is somewhat surprising, given the critical importance that the existence of systemic risk has in the debate on the need for banking regulation and supervision.[16]

With reference to contagious run, Kaufman (1994) argues that only bank-specific contagion is significant in the spread of an individual distress over the banking system. Using the loss to shareholders or deposit rates paid by banks other that the failing one as measures of the breadth of contagion, he finds that contagious run occurred only for banks in the same market or product area as the initially affected one in the pre-Federal Deposit Insurance Corporation (FDIC) and in the post-FDIC era.[17]

Regarding the credit channel, Kaufman (1994) does not find strong evidence on the im-

[16] In the following I review only the main empirical literature on bank contagion. For an extensive survey of the empirical literature of systemic risk, see De Bandt and Hartmann (1999).

[17] Other papers analysing contagious run in the Pre-FDIC era are Saunders and Wilson (1998) and Calomiris and Mason (1997). Neither of the two papers finds evidence of contagion, that is of generalised runs triggered by the release of bad news on the solvency of a bank or on the failure of a single institution.

portance of interbank exposures as mechanisms transmitting shocks from failing to solvent banks. In the Continental Illinois failure in 1984, for example, no correspondent bank suffered solvency-threatening losses. But the no propagation effect in Continental Illinois crisis may have depended on the low losses the bank suffered at the time it failed. Continental Illinois was indeed timely closed and the value of its assets was almost preserved. Schoenmaker (1996b) argues that if the Federal Reserve and the FDIC had not protected uninsured deposits at Continental Illinois, its failure might have caused a chain of bank failures. Using data from a period without a central bank acting as a lender of last resort, he shows that there is risk of contagion in banking since bank failures are dependent.[18] This finding suggests a role for central banks in assisting distressed banks, whose failure may have a systemic impact.

## 1.3 The Safety Net: Deposit Insurance Schemes and the Lender of Last Resort

The potential vulnerability of the banking system to runs and systemic crises is one of the major factors leading many scholars and policymakers to conclude that banks are unique and need to be regulated. In particular, systemic risk constitutes the fundamental rationale for the introduction and the development of the safety net arrangements, namely a form of deposit insurance and a lender of last resort facility.[19] [20]

Although both the two instruments represent a form of insurance for the banking system, they differ in their task, scope, time of application and contractual arrangement. Traditionally, deposit insurance is assigned the task of protecting individual depositors by granting them the reimbursement of their claims in case of bank distress. As formally shown by Diamond and Dybvig (1983), this contractual arrangement is effective in preventing the occurrence of

[18] Schoenmaker (1996a) derives an autoregressive Poisson model, which adresses explicitly the possible spillover effects from one troubled bank to other banks and apply it to a data set of monthly bank failures under the US National Banking System from 1880 to 1936.

[19] This is clearly argued, among many others, by Gerald Corrigan, a former President of the Federal Reserve Bank of New York, and reported in Kaufman (1994), "It is the systemic risk phenomenon [...] - more than any other - that constitutes the fundamental rationale for the safety net arrangements that have evolved in this and other countries".

[20] Investors' protection is often advocated as the other main reason for the need of regulation in the banking system, especially for the introduction of deposit insurance.

runs without reducing banks' ability to transform short-term liabilities into long-term assets. A demand deposit contract with government deposit insurance achieves optimal risk sharing among depositors as unique Nash equilibrium. Government's ability to levy non-distortive taxes and deposit insurance guarantee induce depositors not to withdraw prematurely. Consequently, bank asset liquidation policy is independent from the volume of withdrawals, no strategic issues of confidence arise and no bank runs take place.

The lender of last resort (LOLR) facility is assigned the task of preventing the emergence of systemic crises (in the narrow sense) by supplying liquidity to individual banks in distress.[21] The exact scope and form of central bank intervention are highly controversial in the academic literature. The debate expresses four different views:[22] the classic, the modern-pragmatic, the monetarist and the free banking. The main controversy among these schools centres on the trade-off between the benefits (prevention of contagion) and the costs (distortion of incentives-moral hazard problem) of bailing out distressed banks.[23]

According to the classic view,[24] central banks have a role in lending freely at time of crises in order to avert panics. Loans should be made at a penalty rate and only against good collateral, so to be extended to illiquid but solvent banks. LOLR rules should be well defined and publicly announced. This should discourage banks from using central bank facilities to finance current operations and should prevent an indiscriminate rescue of all institutions. The ideas of market failure and of central banks having superior information on bank solvency are the main arguments for the necessity of LOLR interventions.[25]

[21] Note that crisis management includes three courses of action: "taxpayer money solution", "private money solution" and "central bank money solution" (Padoa-Schoppa (1999)). The first refers to the injection of taxpayers' money by Finance Minister; the second consists of the injection of private money by banks or other market participants; the third one refers to the injection of money created by central banks. The central banks money solution represents the LOLR function, in the strict sense. However, central banks often play a coordinating role for the private money solution to materialise. In terms of the models of contagion presented in section 2.2.1, the central bank should simply act as a coordinating device to prevent the contagion due to the incompleteness of markets in Allen and Gale (2000) and that due to a speculative gridlock in Freixas et al. (1999). Differently, the central bank should act as LOLR to prevent the fundamental gridlock which occurs in Freixas et al. (1999) when the closure of an insolvent bank jeopardises the stability of the entire system.

[22] See, for example, Bordo (1990).

[23] The moral hazard problem, as well as the too big to fail doctrine, will be examined in details in the next section.

[24] The classical school is associated with the works of Thornton (1802) and Bagehot (1873).

[25] According to Bagehot (1873), the market mechanism is unable to deal with bank liquidity shocks because of the presence of asymmetric information about bank solvency. This causes intermediaries not to be able to transmit credible information on the true asset value during a crisis.

The modern-pragmatic view of LOLR focuses on the uncertainty about the true value of bank assets. Goodhart (1987) argues that it is virtually impossible, even for the central bank, to distinguish illiquidity from insolvency at the time the LOLR should act. Further, banks demanding such assistance are under a suspicion of insolvency since they could otherwise raise funds from the market. If illiquidity is inextricably connected with likelihood of insolvency, then central bank ability to lend only to solvent institutions may be hindered. Still, there is a role for LOLR intervention: Central banks should extend the emergency facility to individual banks whose distress may propagate to the entire system. Whenever the social cost of a bank failure is larger than its private cost, the central bank should enlarge discount window loans to individual banks. This does not have to imply a systematic and indiscriminate rescue of all banks: As the private cost of risk taking is reduced if the bank is rescued, the LOLR, as any insurance scheme, induces banks to take greater risk. Consequently, it is crucial that central banks prevent only the failures of individual banks, which are expected to have a systemic impact.

The monetarist school suggests a more restricted use of LOLR facility. Goodfriend and King (1988) argue that there is no need for central bank's discount window loans to individual banks since open market operations are sufficient to deal with systemic liquidity crises. In other words, LOLR should intervene at the macroeconomic level but not at the microeconomic level. The rationale is the idea that the central bank is neither better informed nor more capable to deal with information problems than the private sector. Also, it may be under political pressure to extend loans to weak banks.

Proponents of free banking have an even more extreme position.[26] They argue that no LOLR facility is needed and that public interference in the monetary system is the main cause of instability. Competitive forces would lead to an efficient and stable banking system where the possibility of systemic crises would be remote.

Despite the variety of opinions expressed in the academic debate, central banks actually act as LOLR in most countries and follow a rather uniform policy: They do bail out distressed financial institutions on the ground of eliminating the risk of contagion; they do not commit to a specific line of action but do use "constructive ambiguity" in making their decisions on which

[26]See, among others, White (1984) and Dowd (1989).

banks they are likely to rescue. The use of discretion in LOLR policy should prevent banks from taking full benefit of the LOLR support and from increasing their risk.

The academic literature has recently provided some theoretical support to the LOLR policy implemented by central banks. Goodhart and Huang (1999) formalise the importance of the risk of contagion in central bank closure/rescuing decisions from a macromonetary perspective. There are two crucial aspects in the model: Policymakers cannot discern the state of solvency of institutions demanding liquidity; the cost of bank failure rises more rapidly with the size of the failing bank than the cost of bank rescue. In a static setting, the optimal behaviour of the LOLR is to support bigger banks and to let the smaller fail. This result justifies the use of constructive ambiguity: In order to prevent commercial banks from increasing their risk position, the threshold size for the LOLR support should not be publicly announced. In a dynamic setting, where the risk of contagion and/or the moral hazard problem of LOLR are introduced, the optimal LOLR policy is more complex. If contagion is the main concern, the central bank has an excessive incentive to rescue banks and the equilibrium risk level is consequently high. If moral hazard is the main concern, central bank incentives reduce largely and do not depend on bank size. When both contagion and moral hazard are considered, the central bank incentives to rescue banks are stronger than in the single period setting but weaker than in the dynamic setting with only contagion.

Freixas (1999) provides a rationale for the use of constructive ambiguity from a microeconomic perspective. In a cost-benefit analysis for bailing out banks, he finds that the optimal LOLR policy depends on the liability structure of the bank in distress -the amount of uninsured debt- and on central bank ability to credibly commit to a given policy. In the non-commitment case, the LOLR follows a pure strategy where support is provided to all banks with a low level of uninsured debt. In the commitment case, the LOLR may be a mixed strategy where LOLR is extended randomly to banks fulfilling the uninsured debt requirement. The use of a mixed strategy in the optimal LOLR policy is interpreted as the foundation of the constructive ambiguity policy.

Central bankers' claim that the provision of LOLR helps in preventing contagion finds some support in the empirical evidence. Miron (1986) shows that the LOLR provision in the US has effectively limited the frequency of systemic crises. He finds that the creation of the Federal

Reserve Board in 1914 reduced the probability of having a panic -in a given year- from 0.316 to only 0.005 in the period 1914-1928. A similar result is found by Bordo (1990). Comparing the US and the UK in the years 1870-1913, he finds that the former experienced four massive crises and the latter had none in spite of the evident similarities in the business cycles of the two countries.[27]

## 1.4 The "Moral Hazard" Problem

After about half a century of relative world-wide stability, massive bank failures occurred in 1980s and 1990s in both industrialised and developing countries. Systemic crises arose in such different economies as the United States, Nordic countries, Russia, Japan, Chile, Argentina, Indonesia and Mexico. Regulators in United Kingdom, Italy, Germany and France intervened to rescue individual institutions in distress. Crisis management required huge injections of public resources: The cost of the bail-out policy amounted to 3.2 percent of GDP for the Saving&Loans in US,[28] ranged from 2.8 to 4.0 percent in Scandinavian, was equal to 4 percent in Norway, to 6.4 percent in Sweden and to 8 percent in Finland; the rescue of Japanese banks costed more than $100 billion; the rescue of the Credit Lyonnais, whose losses were unofficially estimated about $10 billion, was the most massive single rescue up to that time experienced by the French government.[29]

The new wave of crises has spurred numerous studies on the causes of the problems and has contributed to renew the debate on the optimal design of financial regulation. According to most economists, one of the main causes of these crises was the excessive risk taking on the part of banks.

Concerning US experience, for example, Edwards and Mishkin (1995) show that banks became more fragile in the 1980s since they expanded their traditional lending activities in riskier areas and started to pursue new, off-balance-sheet, activities. The massive losses on

[27] The LOLR system was created in 1866 in the UK while only in 1914 in the US.

[28] In the period 1984-1991 the United States experienced the failure of more than 1400 Saving & Loans and 1300 commercial banks. James (1991) estimates that total losses on assets occured in the shorter period 1985-88 averaged 30 percent of the failed bank's assets. Losses on assets are measured as the difference between the book value of assets and the recovery value net of the direct expences associated with the failure.

[29] Data are from Caprio and Klingebiel (1996).

loans realised during the 1980s and the peak reached in 1987 witness the decline in bank asset quality. Similar evidence is also found by Boyd and Gertler (1993) in an analysis of bank assets and liabilities over the post-war period.

Several elements can explain the increased risk exposure of most banking systems in the last two decades. One of most recognised is the moral hazard problem induced by ill-designed deposit insurance schemes and by the de facto complete protection offered by regulators implementing forbearance and (often) systematic bail out policies.

The moral hazard problem, as by-product of safety net arrangements, entails two forms of excessive risk taking:

- deposit insurance induces banks, especially those poorly capitalised, to undertake greater risks since depositors do not have any incentives to monitor their banks' asset value;[30]

- a systematic use of LOLR facilities encourage banks to take more risk since they can rely on future bailout in case of distress.

Several studies analyse the potential distortions of the safety net, whether in the form of deposit insurance or LOLR facilities, from a theoretical perspective. Merton (1977) formalises the moral hazard risk of fixed-rate deposit insurance by using option theory. He argues that if deposit premia are risk-insensitive, banks may take on higher risks in order to maximise the put option value implicit in the deposit insurance. Boot and Greenbaum (1993) show that a fixed-rate deposit insurance induces banks to monitor their borrowers less. Under the assumptions of convex monitoring costs and risk neutral agents, a profit maximising bank faces a concave objective function that reaches its maximum at an inefficient level of monitoring whenever the bank finances itself with fully insured demand deposits.[31]

The severity of the moral hazard problem induced by the LOLR facility depends on the likelihood with which distressed financial institutions are bailed out. If central banks could credibly commit to rescue only illiquid banks, the moral hazard problem would be minimal: Bankers would have no incentives to take on excessive risk because they would anticipate they

[30] Empirical evidence on the thrift industry during the 1980s shows that capital deficient and insolvent Saving & Loans were the mostly involved in riskier non-traditional activities in order to maximise the value of deposit insurance (see, for example, Kane (1985)).

[31] Note that already Diamond and Dybvig (1983) recognize that their result on the optimality of a complete deposit insurance relies heavily on the assumption of riskless technology. Indeed, the authors observe that once the choice of bank portfolio risk is taken into account, a moral hazard problem would exist and the introduction of a complete deposit insurance might distort bank incentives.

would fail in case of insolvency. However, such a policy may be neither credible nor optimal. Firstly, as mentioned before, the central bank may not be able to distinguish the state of solvency of banks demanding assistance; secondly, the central bank might choose to provide assistance to a potential insolvent institution if its failure would endanger systemic stability. In other words, in order to limit the risk of contagion, insolvent institutions might be rescued and this, in turn, reintroduces the risk of moral hazard. As formally showed by Goodhart and Huang (1999), when LOLR decisions are based on both risk of contagion and moral hazard problem, the optimal policy is characterised by a positive risk of moral hazard. This is the minimum level consistent with the minimisation of the risk of contagion.

If the provision of LOLR assistance to an individual distressed institution, even if insolvent, depends on the potential effects that its failure would have on the system, it is more likely that large-size banks and banks occupying key positions in the channels of contagion would be rescued. This is the rationale behind the so-called "too big to fail" (TBTF) policy often implemented by central banks. The term TBTF refers to a menu of policies, varying from assistance provision at the discount window to direct infusion of capital and protection of uninsured depositors, that central banks pursue in favour of large or important banks. The anticipated disparity of treatment between small and large (or important) banks in distress might result in an indiscriminate subsidy in favour of these latter.[32] Nevertheless, TBTF may be an optimal LOLR policy to limit the social costs entailed by individual bank failures. This is formally shown by Goodhart and Huang (1999) in the case of big banks in distress, and by Freixas, Parigi and Rochet (1999) in the case of distressed banks having key positions in the payment system.

Another source of moral hazard induced by LOLR intervention is the so- called "too many to fail" policy (TMTF), which refers to the simultaneous bail out of a large number of institutions. Mitchell (1998) analyses the optimality and the distortions of the TMTF policy in formal terms. The crucial element of the model is the interplay between banks and regulators: Banks have to choose between passively rolling over loans in default or actively pursuing their claims; the regulator has to choose whether to monitor bank financial state and how to handle bank

[32] In a study of US banks' behaviour in the 1980s, Boyd and Gertler (1993) find supporting evidence of the risk taking effect induced by TBTF policy.

failures. If banks are discovered to have been passive or to be insolvent, the regulator has to decide whether to monitor them further, and then decide the proper action, or just to rescue them. The choice depends on the cost of the two policies, which, in turn, is a function of the number of institutions in distress: For a given number of distressed banks, a TMTF policy is implemented whenever recapitalisation is the least costly option. Given the possibility of a future massive bailout, banks have more incentives to be passive. In particular, also banks that are financially distressed but still solvent may choose to be passive if they find it optimal to collude in order to trigger TMTF. In equilibrium, the regulator's behaviour will depend on the fraction of distressed banks expected to be insolvent, on the cost of recapitalising banks and on the likelihood that solvent institutions will implicitly collude to trigger TMTF policy.

The controversy on the two distortionary LOLR policies can be considered as part of the debate on "rules versus discretion" in regulatory bailout and, more generally, in crisis management. Supporters of a clearly stated set of rules consider the TBTF and TMTF policies just as a negative consequence of the ambiguity of central bank intervention. This discretionary power would result in a transfer of wealth from small to large banks and in an unconditional rescue of a large number of institutions in distress. On the contrary, supporters of regulatory discretion consider an ambiguous policy as necessary to bring in some market discipline and to deal better with the trade off between systemic risk and moral hazard implicit in crisis management.[33]

Another aspect of the "rules versus discretion" issue relates to the well known dichotomy "flexibility versus laxity". As pointed out by Rochet and Tirole (1996a), a public system has an undeniable advantage relative to a private one in that, by levying tax or issuing money, it does not encounter confidence crises. Such a flexibility, however, may become laxity if the public system is not rigorous enough. When regulatory standards, such as bank closure criteria, are weakened, the so-called forbearance policy can take place. This is defined as the decision of allowing an insolvent institution to remain open. Similarly to the case of systemic LOLR intervention, if banks anticipate they will be allowed to operate even if insolvent, they have strong incentives to take on high risks. As the Saving & Loans debacle confirms, this may result in greater losses and more failures than if regulators implemented tougher closure policies.[34]

[33]See, for example, Goodhart (1987).

[34]See, for example, Kane (1990).

Dewatripont and Tirole (1994) argue that forbearance policy is nothing but the failure of the so-called "representative hypothesis". They study the role of the allocation of control rights to external investors as an incentive device to deter bank moral hazard. In the optimal rule, control should be allocated to debtholders when the bank performs poorly. However, since depositors are small, uninformed and free-riders, they need to be represented by an agent, the regulator, who should act as a large uninsured depositor and implement the optimal closure policy. The "representative hypothesis" works successfully only if regulator's objective function is the minimisation of depositors' losses. Whenever the regulator pursues other interests, such as reputation, or he is resource constrained, he may become too passive and the "representative hypothesis" fails. The regulator may undertake the so-called "regulatory gambling" policy, that is he may conceal banks difficulties in the hope that a positive shock will recapitalise them in the future without any further intervention.

## 1.5 The Responses to the Moral Hazard Problem

The new wave of bank crises has contributed to renew the debate on the optimal design of financial regulation among both policymakers and academics. Several proposals have been put forth to reduce the moral hazard problem induced by the safety net arrangements and the protective attitude of regulators.

The debate proposes two approaches: One, which can be defined as regulation-oriented, focuses on how to reform deposit insurance, how to induce regulators to intervene optimally, and how to strengthen prudential regulation, particularly in the form of capital requirements; the other, which can be as defined market-oriented, suggests to scale back deposit insurance and impose greater market discipline on banks in the form of uninsured claimholders' (either depositors or other debtholders) monitoring.

Proposals on the reform of deposit insurance focus on the premium structure as a crucial element in the design of the optimal insurance scheme.[35] Risk sensitive premia may represent an incentive mechanism for bankers' risk taking even in the presence of asymmetric information between banks and the deposit insurer.

[35] Other crucial parameters, which I will not discuss here, are the nature, public or private, and the funding sources of the deposit insurer. On these issues, see, for example, Baltensperger and Dermine (1987).

Proposals on central bank intervention rely on the idea that effective regulatory monitoring and tough closure policy are important factors in limiting bank risk. The analysis of regulators' incentives to monitor and to close banks is crucial especially when creditors have no incentives in disciplining banks because of complete deposit insurance.

Proposals on the strengthening of prudential regulation centre on the common belief that larger capital ratios reduce bank risk taking. High capital levels should promote stability in two ways: First, they represent a cushion to absorb losses, thus reducing the likelihood of failure; secondly, with more stake at risk, bankers should have fewer incentives to take on risk. However, theoretical results on the effects of capital requirements on bank risk taking are controversial. Kareken and Wallace (1978) show that bankers' incentives to take on risk decrease with the introduction of capital requirements when banks maximise the equity value and seek to exploit deposit insurance. On the other hand, in a high mean-variance framework with utility-maximising banks, Kahane (1977) finds that capital requirements are not only ineffective in controlling risk but they may even induce bankers to choose riskier assets.[36] Similarly, Boot and Greenbaum (1993) find that the introduction of capital requirements may worsen bank moral hazard as it may reduce monitoring effort and, hence, bank asset quality.[37]

More recent contributions suggest to analyse the efficacy of capital ratios in more general frameworks, which take into account other regulatory tools, such as deposit insurance, regulatory monitoring and bank closure policy. The idea is to investigate the complementarity and/or the substitutability of capital ratios with other regulatory incentive devices. Another important strand of literature focuses on how to appropriately compute capital requirements. Two examples are the value-at-risk models for market risk and the pre-commitment approach.[38]

In the following, I review some selected theoretical contributions to the debate on how to

[36] Keeley and Furlong (1990), however, criticise Kahane's approach for its inconsistency and claim that capital requirements are effective prudential devices. With a similar approach to Merton (1978), the authors show that capital ratios prevent intermediaries from choosing excessive risky assets since they reduce the value of the implicit option available to an insured bank. Negative effects of capital requirements on bank risk taking are however found also in Koehn and Santomero (1980), Besanko and Kanatas (1996) and Gennotte and Pyle (1991).

[37] Also Hellwig (1995) regards capital requirements as an inadeguate regulatory response to banking excessive risk taking. He argues that an exaustive theoretical analysis of the implications deriving from capital requirements, both at micro and macro level, is still lacking and that capital regulation cannot substitute uninsured creditors' monitoring of bank asset value, as argued, for example, in Dewatripont and Tirole (1994).

[38] See Freixas and Rochet (1997) for a more detailed survey on the effects of capital requirement on bank risk taking (the so-called portfolio approach to solvency regulation) and Goodhart et al. (1998) for an overview of the new techniques for risk management.

ameliorate bank moral hazard problem. In particular, I focus on the following issues: i) risk sensitive deposit insurance premia; ii) regulatory monitoring and bank closure policy; iii) role of market discipline.[39]

### 1.5.1 Risk-Sensitive Deposit Insurance Premia

The idea of risk sensitive deposit insurance is that, if bank asset risk can be accurately determined, then it is possible to eliminate any advantage from increasing risk through an adequate adjustment of the insurance premium. The practical problem is that the implementation of this proposal requires information that is often not available to regulators.

Insurance pricing cannot be conditioned directly on bank risk profile if banks have private information about their asset value and the investments they undertake. In this case, the insurer can try to acquire some information either thorough costly periodic audits and examinations or through the design of a risk sensitive pricing system that induces banks to disclose their private information. Chan, Greenbaum and Thakor (1992) study the feasibility of such a system in a context where banks' type and asset choice are not observable to the deposit insurer. They show that a deposit insurance pricing linked to banks' observable reported capital can induce banks to reveal their type and to make the appropriate asset choice. However, this is not possible if the deposit insurance premium is fairly priced, that is if the deposit insurer has to break even on each individual institution, and the banking sector is perfectly competitive. A fairly priced and completely risk-sensitive deposit insurance is implementable only if banks have access to rents, either through explicit regulatory subsidies or restricted entry into banking. The reason is that if banks do not have any rents, they are indifferent to the capital structure and high-risk banks find it optimal to mimic their low risk peers when deposit insurance premia are risk-adjusted.

Thus, fairly priced deposit insurance is incompatible with a competitive credit market in which private information and moral hazard distort the equilibrium. Successful sorting instruments are charter values, induced by entry restrictions, or deposit-linked subsidies accomplished by an insurance pricing scheme inversely linked to capital requirement. Nevertheless, the design of contracts that would arise in a competitive insurance market may not be the appropriate

[39] Other interesting proposals, which I will not discuss in this chapter, include cash asset reserve requirement, limits on discount window borrowing and portfolio restrictions. On the debate about these issues see Bhattacharya, Boot and Thakor (1998).

objective for bank regulators in the presence of important externalities in the banking industry, such as informational asymmetries or risk of contagion.[40]

Freixas and Rochet (1999) question the result of incompatibility between fairly priced deposit insurance and competitive banking market argued by Chan, Greenbaum and Thakor (1992). They show that under more general assumptions on bank operating costs, fairly priced deposit insurance becomes possible in a competitive banking system, even when an adverse selection problem is present. However, such a scheme is not desirable, since the optimal premium schedule would entail a subsidisation of the less efficient banks by the more efficient ones.

Gianmarino, Lewis and Sappington (1993) study the optimal design of a risk-adjusted deposit insurance scheme in a context where monopoly profit-maximising banks have private information on their environment and activities, and the regulator maximises social welfare, which is given by bank profits less the social costs of government involvement and of financial failures. The key result is that the first best level of asset quality cannot be achieved in the socially optimal deposit insurance scheme. This is because of the trade-off between information asymmetries and costly government intervention: Since deposit insurance is financed through distortive taxes, the regulator has to limit the informational rents accruing to banks, which, in turn, will not have adequate incentives to choose the first best level of asset quality. So, an incentive-compatible deposit insurance pricing scheme is not optimal: The benefits in terms of higher asset quality are not counterbalanced by the distortions in terms of costly regulatory intervention necessary to achieve incentive-compatibility.[41]

### 1.5.2 Regulatory Monitoring and Bank Closure Policy

With full deposit insurance, creditors do not play any active role in monitoring and disciplining banks. Regulators act on behalf of depositors to limit bank risk taking. The analysis of how to provide regulators with the appropriate incentives to implement the optimal level of monitoring and to take the proper actions becomes crucial. As noted by Tirole (1994a), the design of an optimal incentive scheme for public regulators is not straightforward because some of their performance variables are not easily measurable and their actions are often driven by

[40] The sacrifice of fair price deposit insurance for the achievement of the optimal level of safety and soundness is, for example, suggested by Berlin, Saunders and Udell (1991).

[41] Similar results are obtained by Bensaid, Pages and Rochet (1995).

informal incentives, such as political and career concerns. Therefore, the regulator may be easily distorted away from undertaking the optimal action.

This is formally demonstrated by Boot and Thakor (1993). They consider a two-period model in which (i) in each period the bank chooses the investment risk (ii) the regulator monitors the bank asset choice in the first period and selects the level of bank capital at which to close the bank. The regulator's monitoring is imperfect and depends on his quality. The result is that if there is even a little uncertainty about the regulator's ability to monitor, he will not implement the optimal bank closure policy. Indeed, when the regulator cares also about his monitoring reputation, he chooses to close the bank at a lower capital level than what would be socially optimal. This is because the closure of a bank induces the market to down-grade its beliefs about the regulator's monitoring ability. The more lax closure policy induces bank managers to choose a higher level of risk in both the two periods.

To sum up, a self-interested regulator purses a sub-optimal bank closure policy, which, in turn, induces banks to take on more risk. A similar result arises when monitoring banks is costly and the regulator is not benevolent. In this case, bank risk taking can be controlled by providing the regulator with adequate incentives to monitor, or, alternatively, by using complementary and/or alternative regulatory tools. Campbell, Chan and Marino (1992) study the optimal incentive contract between depositors and regulators. They use a static framework where the banking sector is perfectly competitive, depositors choose the capital requirements and the regulator has to monitor the bank asset choice. Monitoring technology is costly effort-related and the greater the effort, the greater the probability that the monitoring is perfectly informative. If the regulator is benevolent and his monitoring effort is observable, monitoring and capital requirements are at the first best levels, capital requirements are lower than without monitoring and depositors gain from the increased liquidity service of deposits. Conversely, if the regulator is self-interested and his actions are not observable, monitoring and capital requirements depart from the first best levels. Thus, depositors have to design an appropriate incentive scheme for the regulator, taking into account his personal resource constraint. When the regulator's limited liability constraint is not binding, the first best solution is still attainable. The regulator expends the right monitoring effort in order to avoid the penalty imposed in the case of bank insolvency. When the regulator's limited liability constraint is binding, only the second best solution is achievable since providing the regulator with the right incentives is too costly.[42] The first best level of bank moral hazard deterrence can be achieved only through higher capital requirements and lower monitoring activity. So, if the regulator is benevolent or his limited liability constraint is not binding, direct monitoring and capital requirements are alternative mechanisms to limit bank risk taking incentives. Otherwise, they become complementary.

Apart from informational problems and/or inadequate incentives for regulators, other factors can preclude the implementation of optimal regulatory policies. Mailath and Mester (1994) investigate regulators' incentives to close intermediaries in a dynamic framework, in which they introduce credibility concerns and a social opportunity cost of closing a bank. The analysis focuses on the interaction between banks' current and future risk taking and closure rules. Profit-maximising banks have to choose between a riskier and a less risky project and their choice is observable to the regulator. The authors show: (i) the implementation of the closure policy depends on the regulator's objective function and, in particular, on whether he is welfare-maximising or cost-minimising;[43] (ii) both types of regulator fail to implement a "clear cut" closure policy, that is they do not close all insolvent banks and do not let all solvent banks operate. This is because the closure threat may be not credible and, therefore, ineffective in deterring bank risk taking.

Thus, forbearance may have different connotations. It needs not indicate that a regulator is passive or fraudulent. It may also indicate that the regulator faces a credibility problem with being tough if closing a bank is costly. Furthermore, forbearance may emerge as part of an optimal closure policy when bank assets are subject to market risk. This is argued by Nagarajan and Sealey (1995) in a context where banks operate under a moral hazard problem and their portfolio returns are subject to both specific and market risk. The authors show that the optimal regulatory policy involves promptly closure of banks that fail when good market conditions

[42] The regulator's personal resource constraint limits the penalty that can be imposed. So, when this constraint is binding, the first best monitoring level can be induced only with a compensation package composed of a higher reward in the case of bank solvency and a lower penalty in the case of bank insolvency. Then, depositors optimally trade off the value of monitoring against the cost of inducing the monitor to exert effort and find it profitable to rely more on capital requirements than on monitoring to deter bank risk taking.

[43] The welfare-maximising regulator cares about the payoff of all agents and, therefore, he takes into account the social opportunity cost of closing a bank, in the form of forgone intermediation service. The cost-minimising regulator is only concerned with his insurance payments to depositors. There is not a univoque correspondence between thoughness of closure policy and regulator's objective. Indeed, the cost-minimising regulator may have either stronger or weaker incentives to leave the bank open than the welfare-maximising regulator.

prevail and forbearance (although with a certain probability) of banks that fail when market conditions are poor. The optimal policy include also a minimum capital requirement so to prevent banks from undertaking a "gambling for resurrection" behaviour, but it is independent of the deposit insurance pricing.[44] This latter only influences the regulator's profit condition: The regulator can break-even under a risk-sensitive deposit insurance scheme, whilst he may be obliged to offer a subsidy under a flat-rate pricing scheme. The irrelevance of deposit insurance pricing for the design of the optimal regulatory policy suggests that the risk-shifting problem induced by a fixed-rate deposit insurance may be overestimated if not analysed within a more comprehensive regulatory package.

The optimality of the forbearance policy remains however a controversial issue in the academic debate. More research is needed. The scope of the forbearance policy needs to be investigated in richer frameworks, which consider also other relevant parameters, such as market structure. A first attempt in this direction is Acharya and Dreyfus (1989). They find that the closure policy of a cost-minimising regulator depends on the spread between the expected rate of return on bank assets and the interest rate on deposits, which is in turn a function of market structure. In particular, they show that when this spread is low, banks face a positive probability of being closed when the ratio of assets-to-deposits falls below an endogenously determined threshold level. Since this level is greater than one, healthy banks may be closed in a competitive environment. Differently, when the spread is sufficiently large, the regulator always finds it optimal to let banks operate. Thus, in a monopoly setting all banks are allowed to operate, whilst in a competitive setting bank solvent banks may be closed.

### 1.5.3 The Role of Market Discipline

The basic idea of the market-oriented approach is that a system without deposit insurance could credibly restore discipline on banks. Some policymakers have argued that a market-based monitoring system, where depositors discipline banks by threatening to withdraw their funds, would limit regulatory forbearance and would provide a credible threat of bankruptcy.[45]

[44]The "gamble for resurrection" behaviour refers to banks' incentives to maximise risk when they are let operate even if insolvent.

[45]See, for example, Federal Reserve Bank of Chicago (1990) and Broaddus (1994). Also Boyd and Gertler (1993) sympathize with the idea of scaling back deposit insurance in order to reintroduce market forces.

[46] Depositors at uninsured banks would have an incentive to run as soon as they have doubts about the solvency of the intermediary, thus forcing a tough and rapid closure.

Some contributions have analysed the incentive effects of demandable debt on bank risk taking. Calomiris and Kahn (1991) show that the threat of bank liquidation disciplines the banker when he can fraudulently divert resources ex post. Their analysis focuses on the issue of costly acquisition of information by depositors and gives a rationale for the imposition of the "sequential service constraint" in the repayment of depositors. Calomiris and Kahn regard bank runs as always beneficial since they prevent fraud and allow the salvage of some of the bank value. The result is then that any extra market intervention, such as deposit insurance or central bank facilities, is both unnecessary and undesirable because it would only lower depositors' welfare.

This may not hold any more when bank runs arise from the co-existence of heterogenous depositors (informed, uninformed and 'consumption-oriented' depositors ). As I will show in chapter 3, market discipline can be effective in resolving the moral hazard problem that arises when depositors do not know whether bankers are monitoring the projects they finance. However, market discipline may come at a cost. When depositors are not equally informed about the future value of bank assets, withdrawals caused by a liquidity shock may be confused with future insolvency and cause uninformed depositors to precipitate a run. Likewise, withdrawals due to upcoming insolvency may be confused with a liquidity shock and dissuade depositors from running. Bank runs may be, therefore, costly and imperfect disciplinary devices for bankers. This result suggests a role for extra market interventions: Any attempt to make market discipline work should entail adequate regulatory measures aimed to eliminate its inefficiencies.

Another possible inefficiency of market discipline stems from the contagious nature of bank runs. Chen (1999) shows that depositors' monitoring can prevent the banker from liquidating the long-term project at the interim date and investing in an inefficient short-term project. However, market discipline may be inefficient because depositors at one bank may withdraw their funds from their bank in response to failures of other banks. That is, runs at some banks may cause runs at others. Since the contagion is based on noisy information, it entails social

[46] Debtholders can exert discipline on banks also through the pricing of their claims. See, for example, Benston et al. (1986) and Calomiris (1998).

costs.

As I will discuss further in chapter 3, the two explanations of the inefficiency of market discipline differ in the mechanism generating such inefficiency. While in the model I develop in chapter 3, the source of inefficiency originates inside the bank and disruptive runs can occur even if some depositors are *already* informed on the value of their bank's assets, in Chen's model it originates outside the bank and contagion takes place *before* depositors at one bank receive precise bank-specific information. In this sense, the sources of market discipline inefficiency in two contributions are complementary.

One plausible objection to the proposal of relying on depositors' monitoring to discipline bank risk taking is that small retail depositors, even if uninsured, would have little incentives to monitor. However, recent empirical evidence supports the market-oriented approach. Park and Peristani (1998) and Peria and Schmukler (1998), among others, show that both in developed and developing economies depositors do react to the deterioration of banks' balance sheet. They find that depositors, whether small or large, punish risky banks by withdrawing their funds or by requiring higher interest rates.

## 1.6 Competition and Fragility of the Banking System

In the previous sections, I have reviewed the literature on bank fragility, excessive risk taking and regulation. In particular, I have analysed contributions aiming to explain why the banking sector is vulnerable to individual runs and systemic crises, why it needs to be regulated, how public intervention creates distorsions in terms of excessive risk taking, and which measures can be adopted to ameliorate bank moral hazard problems. Risk-sensitive capital requirements and deposit insurance premia, and tough bank closure policy have been analysed as potential devices to control bank excessive risk taking. Also, market discipline in the form of uninsured depositors' monitoring has been suggested as effective measure.

One aspect that has been largely ignored by the banking literature concerns the relationship between competition, stability and regulation. Most of the contributions reviewed so far pay very little attention to the strategic interaction among banks and ignore the effects of market structure on bank stability and on the effectiveness of the measures aiming to preserve it.

Most of the models assume that banks operate in a perfectly competitive environment or in a monopoly setting. For example, models of bank runs analyse the effects of demandable debt offered by banks operating either in a competitive or in a monopoly setting. In both circumstances, runs emerge in equilibrium either as a consequence of depositors' coordination failure or as depositors' rational response to the arrival of negative information about bank future solvency. These models do not tell in which market structure the banking system is more likely to be unstable. Neither they do explore the effectiveness of the safety net arrangements and of other regulatory measures in different market settings.

In the following, I review the few theoretical contributions that address the relationship between competition, stability and regulation. In line with the previous sections, I present firstly the models addressing the link between market structure and fragility, and, secondly, those focusing on market structure and excessive risk taking. Finally, I describe the models on competition and regulatory reforms. Note that only few of the contributions I will discuss attempt to endogenise aspects of industrial organisation. The majority of them just compare the equilibria achievable in different market setting without taking into account any strategic interaction among intermediaries.

The link between financial fragility and competition among banks is analysed by De Palma and Gary-Bobo (1996). The model focuses on the relationship between Cournot competition on the loan market and depositors' withdrawal decisions. Intermediaries issue demandable deposits and grant loans to limited liability firms. Loans are subject to macroeconomic shocks and can be liquidated prematurely only at a cost. After depositing their funds at a bank, depositors receive information on the future bank solvency and decide whether to withdraw their deposits or to wait. If they decide to withdraw, an information-based run occurs. Depositors' decisions depend on their probabilistic beliefs about the uncertain returns of bank investments. Such beliefs are described by a bimodial density function, which represents their hesitation between two views of the world in a state of crisis. The Cournot competition on the loan market, together with depositors' bimodial belief distribution, generates multiple equilibria: In the safe equilibrium, banks offer a small amount of loans at a high interest rate and bear no bankruptcy risk. In the risky equilibrium, banks supply a large amount of loans but are subject to a positive probability of runs when depositors receive a bad signal.

Since depositors are uninsured and there no capital requirements in the model, De Palma and Gary-Bobo suggest to interpret their analysis as a theory of deregulated banking competition. Consequently, the results suggest that a deregulated system with market imperfection is potentially highly fragile. However, coordination problems among depositors can emerge also independently of competition and, consequently, can occur in any market configuration. This is shown by Matutes and Vives (1996) in a model that introduces product differentiation and network externalities in the classic framework of Diamond (1984). Unlike in Diamond, banks cannot fully diversify their portfolios, even if there are economies of scale, and, as a consequence, they can fail. The distress probability of a bank is endogenously determined by depositors' expectations, which are self-fulfilling due to scale economies. A bank perceived to be safer commands a higher margin and a larger market share, which, in turn, makes it actually safer because of better diversification. The self-fulfilling character of depositors' expectations implies multiple equilibria. Possible equilibria include corner solutions, where one bank is out of the market, and even no banking. This event is interpreted as a systemic crisis or, consistently with Diamond and Dybvig (1983), as a sunspot run. The bad equilibrium is due to a coordination problem among depositors, which arises for reasons similar to those encountered in the network literature, irrespective of the competition on the deposit market.

The relationship between competition and bank fragility is also analysed by Smith (1984). He uses a Diamond and Dybvig (1983) framework where banks compete to attract depositors that have different probability distributions over the dates of withdrawal. When information is perfect, there exists a Nash equilibrium that achieves the optimal contract: Banks attract depositors by announcing state contingent vectors of first and second period interest rates and break even on each type of deposit offered. Conversely, if an adverse selection problem is present, that is if depositors only know their own probability of withdrawals, there may not exist a Nash equilibrium. This is due to the fact that the equilibrium contract, either pooling or separating, is destroyed by the possibility of banks offering positive profit contracts to a specific segment of depositors. When this is the case, the banking system in not viable or, in other terms, is unstable. Thus, competition for deposits makes banks fragile in an environment characterised by adverse selection problems.

## 1.7 Market Structure and Excessive Risk Taking

Competition in banking has traditionally been blamed for excessive risk taking. High bank charter values have been regarded as a strategic regulatory instrument, justifying the imposition of structural limitations and/or the allowance by governments of collusive agreements among banks.

The deregulation wave that occurred in the 1980s has been considered as a major cause of the moral hazard problem that contributed to trigger the massive bank crises in the last two decades. The Saving and Loans debacle showed how the competitive pressure from mutual funds on the deposit market and the subsequent release of some regulatory constraints induced many thrift institutions to undertake riskier activities, thus increasing their probability of failure. Edwards and Mishkin (1995) argue that the excessive risk taking observed in the U.S. system in the 1980s was nothing but banks' response to their diminished profitability in an attempt to maintain their position as financial intermediaries. Banks' lower profitability was, in turn, a consequence of greater competition in financial markets, which decreased the cost advantage that banks had in acquiring funds and undercut their positions in the loan market.

Keeley (1990) provides empirical support to the hypothesis that enhanced competition induces banks to take greater risk through a reduction of their charter values. In a study on US banks over the period 1970-1986, he finds that those with more market power had a lower default risk, as reflected in lower risk premia on large and uninsured CD's. Furthermore, he argues that the banking system was stable until the 1980s, despite the presence of a fixed-rate deposit insurance scheme, because high charter values were effective in countervailing its perverse effects. Once the regulatory constraints were eased in the 1980s, the charter values decreased and banks started to assume greater risks in the pursuit of the risk engendered by deposit insurance.

Moving from Keeley's results, some papers (the so-called 'charter value' literature) have analysed, from a theoretical perspective, the incentive effects of high charter values for bank risk taking. Besanko and Thakor (1993) build a model of relationship banking to examine the effects of interbank competition on bank portfolio choice. Lending relationships between intermediaries and borrowers provide banks with informational advantages over other lenders and, consequently, with informational rents. To the extent that the bank and the borrowers share these rents, both parties have an incentive in continuing the relationship. Thus, despite the presence of risk-insensitive deposit insurance, banks are induced to limit their risk exposure in order to enjoy the value of the relationship. However, as the banking industry becomes more competitive, relationship banking decreases in value and, in turns, banks take more risk.

Boot and Greenbaum (1993) reach similar results. They develop a two-period model in which banks can affect the payoff distribution of investment projects through costly monitoring. Banks have different monitoring abilities, which are not observable. Monitoring increases banks' expected profits in two ways: On the one hand, it decreases project risk; on the other hand, it improves banks' reputation as capable monitors, thus lowering their subsequent funding costs. Funding-related reputational benefits and rents are then substitute mechanisms for limiting bank risk exposure. Incentives based on reputation are available only to banks that are uninsured. By fixing banks' future funding costs, risk-insensitive deposit insurance destroys the funding benefits related to reputation, thus discouraging monitoring and inducing excessive risk. This is especially undesirable when increased competition reduces monopoly rents.

However, the negative link between competition and excessive risk taking becomes blurred in richer frameworks where banks use more than one instrument in dealing with asymmetric information. Caminal and Matutes (1997a) develop a static model in which banks can use monitoring and credit rationing to deal with a moral hazard problem on the part of the entrepreneur. The two instruments are imperfect substitute incentive devices: Monitoring is costly, whilst credit rationing reduces the potential gain from trade. If the bank does not monitor, it reduces credit in order to induce entrepreneurs to choose the appropriate investment project. The model compares the outcomes of two extreme market structure, namely monopoly and Bertrand competition on the loan market. Two countervailing forces determine the loan size and, consequently, asset risk in equilibrium: On the one hand, when banks enjoy high market power, they tend to set higher lending rates and higher levels of monitoring, thus decreasing the proportion of credit-constrained borrowers. On the other hand, given a level of monitoring, a higher interest rate worsens the firm incentive problem, which in turn tightens the credit constraints. If the first effect is sufficiently strong, monopoly leads to higher interest rates and greater loan size. Due to the assumption of multiplicative aggregate shocks, this implies that a monopoly bank faces a higher failure probability than a competitive bank. As a consequence, the relationship between market power and failure probability is ambiguous: Since a monopoly bank acquires more information and uses less credit rationing, it may be more exposed to macroeconomic uncertainty, leading to a higher bankruptcy probability.

## 1.8 Competition and Regulatory Structure

As discussed insofar, the existing theoretical models are not sufficiently robust to deliver clear conclusions on the link between competition and stability, in both dimensions of fragility and excessive risk taking. Some models show that competition undermines stability, whilst some others conclude that instability arises irrespective of competition and, even, that monopoly settings may be more unstable than competitive environments.

As a consequence, also the normative implications of the relationship between competition and stability are not well understood yet. Existing contributions tend to indicate that competition and regulation influence each other and that, as a consequence, the effectiveness of a particular regulatory measure cannot be assessed independently on the market structure, which, in turn, may change with the new regulatory environment.

Matutes and Vives (1996) extend their analysis on competition and fragility, by investigating the welfare effects of deposit insurance. They show that deposit insurance prevents the coordination problem among depositors, thus eliminating the risk of fragility. However, deposit insurance implies a welfare trade-off: On the one hand, it prevents bank collapses and tends to enlarge the market. On the other hand, by ensuring that all banks are active, it may preclude the realisation of desirable diversification and may induce fiercer competition for deposits, which, in turn, increases the failure probability of banks. The net welfare effects of deposit insurance are ambiguous and cannot be assessed independently of the market structure. Also, by extending the market, the introduction of deposit insurance has the potential effect of changing the market structure from one where banks have local monopoly power to one where they compete.

Turning to the impact of competition on excessive risk taking, regulation can affect the way in which charter values are generated and, in turn, affect their impact on banks' incentives to take risk. Nagarajan and Sealey (1995) argue that when higher margins are the result of a forbearance policy that extends the expiration of equity holders' call option, they may not result in higher quality of bank assets. In particular, high charter values provoke excessive risk taking when they are generated by a non-optimal forbearance policy.

These results suggest also that the perverse link between competition and stability may be corrected by adjusting regulation to the structural changes in banking. This could be done, for instance, through effective closure policy or risk-adjusted deposit insurance. Cordella and Yeyati (1998) investigate the impact of competition on the determination of deposit interest rates and on bank risk taking behaviour under different deposit insurance arrangements. They develop a model of spatial competition where banks choose privately their portfolio risk and face an imperfectly elastic demand for financial services. Under fixed-rate deposit insurance, enhanced competition increases deposit rates and risk. Indeed, a lower product differentiation among banks increases the interest rate elasticity of deposit supply and induces tougher price competition and lower margins, thus reducing banks' incentives to limit risk. Conversely, under risk-adjusted deposit insurance, deposit rates and asset risk are lower than under a flat-rate pricing scheme. When risk information is disclosed to the deposit insurer who can charge a risk-based premium on deposits, banks can credibly commit to reduce asset risk, thus reducing the cost of funds and improving their overall performance.

The welfare implications of enhanced competition on bank risk taking under different deposit insurance schemes are also examined by Matutes and Vives (2000). They develop a model of product differentiation where banks subject to limited liability compete for deposits and their failure entails social costs. Banks choose the risk of their portfolio and the deposit rate, whilst investors decide how much to deposit at each bank. The risk of banks' portfolios can be either observable or unobservable by investors; deposits can be either uninsured or insured. In the latter case, deposit insurance pricing can be either flat or risk-sensitive. In an uninsured market, high failure costs and intense competition lead to excessive deposit rates. The more competitive the market, the larger the set of failure costs for which the deposit rates are excessive. In the limit case of perfect competition, deposit rates are always excessive, independently of the size of failure costs. Uninsured market performance can be improved through deposit rate regulation, which can maximise welfare when bank asset risk is observable, whilst needs to be complemented with investment restrictions when bank asset risk is not observable.

Deposit insurance modifies banks' incentives to set deposit rates and to take on asset risk. Flat deposit insurance makes banks more aggressive in deposit rate setting, thus inducing them to maximise asset risk, irrespective of the costs of failure. Both deposit regulation and asset restrictions are needed to improve welfare. Risk-sensitive deposit insurance generates lower equilibrium levels of deposits rates and asset risk than uninsured markets. However, when failure costs are high, welfare may still be improved by introducing deposit rate ceilings.

## 1.9 Conclusions

This chapter has reviewed the main theories on the issue of bank stability and regulation. In particular, it has highlighted two connotations of the term instability, fragility and excessive risk taking, and has reviewed the literature accordingly. The vulnerability of banks to runs and systemic crises results from the specificity of intermediaries as liquidity providers and from the informational asymmetries characterising bank activities. The moral hazard problem of excessive risk taking is a distorsion created by an inadequate structure of the safety net, namely the way in which deposit insurance and the lender of last resort function are designed.

The bank crises that occurred in the last two decades have prompted a rethinking on how to preserve the soundness of the banking system, with particular concern on how to structure deposit insurance, regulatory interventions and capital requirements. The contributions surveyed in this chapter stress the potential of a risk-based deposit insurance and of a reduction of deposit insurance coverage in order to reintroduce market discipline in the form of uninsured depositors' monitoring. Further, they stress the importance of regulators' objective functions and the difficulty of providing regulators with adequate incentives to avoid laxity in the circumstances where forbearance is not part of the optimal regulatory design.

The literature on bank stability largely disregards the implications of different banking structures for the safety of the sector. The general argument is that competition worsens stability. For example, a perfectly competitive setting may be incompatible with a fairly-priced and incentive-compatible risk-adjusted deposit insurance, since its implementation requires banks to earn positive rents. Market power is seen as a mitigating factor of bank risk taking, since high margins act as a buffer against portfolio risk and increase the cost of bankruptcy.

Nevertheless, the relationship between competition and stability becomes elusive in richer models, which consider imperfect competition and endogenise aspects of industrial organisation. Recent contributions suggest that coordination problems among depositors causing bank fragility can emerge independently of competition. Also, they show that a monopoly bank may face a higher failure probability than a competitive banking industry.

The relationship between competition and stability has important normative implications. Policy instruments and market structure influence each other: The effectiveness of a particular regulatory measure depends on the industrial setting, which, in turn, may change once the regulatory measure is implemented. The few contributions addressing the optimal regulation design in models of imperfect competition suggest that, even if competition hurts stability, its negative effects can be ameliorated by designing financial regulation appropriately.

To conclude, despite the general feeling that competition induces higher instability in banking, the theoretical literature is still far from being conclusive. The results on the link between competition and fragility, and on market structure and risk taking are still ambiguous. To achieve a better understanding both on the positive and normative aspects of the relationship between competition and stability, additional research is needed in several directions. First, the link between market structure and bank fragility is worth further study: models of runs and panics should be extended to situations of imperfect competition. Second, the effects of imperfect competition on bank risk taking should be examined in richer frameworks, which consider competition on both the loan and the deposit market. Third, on the normative side, further analysis is needed to evaluate the effectiveness of regulatory reform proposals.

# Chapter 2

# The Bank Run Phenomenon: A Selective Review of the Literature within a Uniform Framework

## 2.1 Introduction

This chapter analyses the phenomenon of bank runs, examining in detail the main theoretical literature on this topic.

The term "run" refers to the situation in which many or all depositors at one bank attempt to withdraw their funds simultaneously, forcing the bank to liquidate all its assets and to close down. Banks' fragility originates in the peculiar structure of their balance sheets. They issue liquid liabilities in the form of demandable deposit contracts and invest in illiquid assets in the form of loans, which are costly to liquidate before maturity. This maturity transformation explains banks' role as providers of liquidity to depositors and, at the same time, their vulnerability to runs.

The literature on bank stability has suggested different explanations for depositors' withdrawal decisions, which have implied different conclusions concerning the efficiency of runs. They have been regarded as "unfortunate and undesirable side-effect of demandable debt" in the literature of the early '80s, as a "rational response of depositors to the arrival of negative information, although sometimes erroneous" at the end of the '80's, as "perfect discipline devices" in the early '90s and as "inefficient discipline device" very recently. These different connotations of the term "runs" derive from the different issues addressed in the literature in the course of time.

Two sets of questions have dominated the literature on bank stability: (1) Why are banks unstable? What are the costs of their instability? (2) Why have banks used demandable debt as the primary means of funds if it entails higher costs than other available means of financing, such as standard debt or equity?

The chapter is organised around these two sets of questions. The contributions on bank runs are divided into two categories, depending on the issues they address. The first category, defined as the "first generation" literature, focuses on the role of banks as providers of flexibility to depositors in their uncertain timing of consumption and regards runs as a consequence of the maturity mismatch between assets and liabilities. Depending on the mechanism triggering depositors' decisions to withdraw prematurely, contributions belonging to this category see runs as irrational ("sunspot" approach) or information-based ("asymmetric information" approach) events. The second category, defined as the "second generation" literature, examines the agency problems existing between a bank and its depositors and considers the role of runs as discipline devices for bankers' moral hazard problems. Depending on whether depositors base their withdrawal decisions on perfect or noisy information, runs are now seen as efficient or costly incentive mechanism.

Before reviewing the literature, I describe a general model which I use as common framework in describing the contributions of both generations. This should facilitate the understanding of the different approaches of the literature, and the relation between these and the model I will develop in chapter 3.

The chapter proceeds as follows. Section 2 describes the general framework used for the description of the literature. Section 3 analyses the main first generation models, making a distinction between the "sunspot" approach and the "asymmetric information" approach. Section 4 presents the second generation models, emphasizing whether runs are seen as efficient or inefficient discipline devices.

## 2.2 The General Model

In this section I develop a model which I use as general framework in describing the bank run literature.

### Assumptions

1) Consider a three-date economy (T=0,1,2) with a single good. At date 0 investment is undertaken. At dates 1 and 2, investment returns are realised and agents consume.

2) *TECHNOLOGY:* There are two technologies available: a short-term storage technology, which transforms each unit of good at date 0 into one unit of good at date 1; a long-term technology, which converts one unit of good at date 0 into $R > 1$ units of good at date 2. The long-term technology can be either deterministic or stochastic. In the latter case $R$ takes up two values, $H$ and $L$, with probability $p$ and $1 - p$, respectively, with $H > 1$, $0 \leq L < 1$ and $pH + (1 - p)L > 1$.[1] The investment is perfectly divisible.

The long-term production process can be interrupted prematurely at a cost: Each initial unit of investment in the long-term technology yields a return $\ell \leq 1$ if liquidated at date 1. This assumption captures the idea that investments are illiquid. In particular, it finds its explanation in the investment irreversibility in the case of deterministic technology and in the information advantage that investors have on the return $R$ in the case of stochastic technology.

Table 1 summarises the technologies available in the economy.

| TECHNOLOGY | $T = 0$ | $T = 1$ | $T = 2$ |
|---|---|---|---|
| Short term | $-1$ | $1$ | $0$ |
| Long term | $-1$ | $\ell \leq 1$ | $R = \begin{cases} R > 1 & \text{if deterministic} \\ \begin{cases} H > 1 & p \\ 1 > L \geq 0 & 1-p \end{cases} & \text{if stochastic} \end{cases}$ |

Table 1: Technology

3) *DEPOSITORS*: There is a continuum of depositors (consumers) who are endowed with one unit of good at date 0 and none at the other times. Consumers are all ex ante identical but

[1]The case of deterministic return corresponds to the limit case $p \longrightarrow 1$ and $H = R$.

become either of type 1 or of type 2 at date 1 after a preference shock is realised. Type 1 agents are impatient, as they value today's consumption more relative to tomorrow's consumption. In the extreme case, they are assumed to die at date 1 and, therefore, to care only about consumption at that date. Conversely, type 2 agents prefer tomorrow's consumption to today's consumption and, in the extreme case, they derive utility only from consumption at date 2. More specifically, consumers' utility functions are given by:

$$\begin{aligned} U^1(c_{11}, c_{21}) &= \quad u(c_{11}) + \rho_1 u(c_{21}) \quad \text{for type 1 agents} \\ U^2(c_{12}, c_{22}) &= \quad u(c_{12}) + \rho_2 u(c_{22}) \quad \text{for type 2 agents} \end{aligned}$$

where $c_{ij}$ is the consumption at date $i$ of an agent of type $j$ and $\rho_i$ is the intertemporal discount factor with $1 > \rho_2 > \rho_1 \geq 0$. Thus, $\rho_i$ describes the degree of agents' impatience. The utility functions are *smooth* if $\rho_i > 0$ with $i = 1, 2$, and they are *corner* when $\rho_1 = 0$ and $\rho_2 > 0$. In both cases, the utility functions $U^i : R_{++} \rightarrow R$ (with $i = 1, 2$) are assumed to be twice differentiable, increasing, and satisfying the Inada conditions $u'(0) = \infty$ and $u'(\infty) = 0$. Further, $U^i$ is assumed to be either strictly concave ($u'' < 0$) or linear ($u'' = 0$). In the former case, agents are risk averse and their relative risk aversion coefficient, $RRA = -cu''(c)/u'(c)$, can be either bigger or small than one. In the latter case, agents are risk neutral.

Each agent has a probability $t$ of becoming of type 1 and $1 - t$ of becoming of type 2. The probability $t$ can be either a constant (no aggregate uncertainty) or a stochastic variable (aggregate uncertainty). In this case, $\tilde{t}$ is assumed to be discrete with probability function $q$.

As agents are ex ante identical, their ex ante utility function is given by:

$$U = tU^1(c_{11}, c_{21}) + (1 - t)U^2(c_{12}, c_{22}).$$

Storage of goods is not allowed between dates 0 and 1, whilst it is allowed between dates 1 and 2.

4) *INFORMATION:* When the long-term technology is stochastic, the realisation of the return $R$ is publicly known only at date 2. However, a fraction $\alpha$ of type 2 agents receives at date 1 a signal $s$ on the future value of $R$. The signal is the same for all depositors and can be either perfect or partial. In the former case, $s$ corresponds to the future realisation of $R$,

that is $s \in \{H, L\}$; in the latter case, $s$ is described by the distribution of ex post beliefs to which it could lead. The posterior beliefs, $\widehat{p}$ and $1-\widehat{p}$, are consistent with the priors, that is $p = \sum_s prob(s)\widehat{p_s}$, where $\widehat{p_s}$ is the value of $\widehat{p}$ given that $s$ is observed.

Observing the signal $s$ can be either free or costly. When it is free, the fraction $\alpha$ of informed depositors is either constant or stochastic, in which case it assumes the values $\overline{\alpha}$ or 0 with probability $r$ and $1-r$, respectively. When observing the signal is costly, the fraction $\alpha$ of informed depositors is an endogenous variable of the model.

At date 0, agents do not know which type they will become and whether they will receive the signal at date 1 or not.[2] At date 1, the preference shock and the signal $s$ are privately observed. Further, the realisations of the variables $t$ and $\alpha$, if stochastic, are not publicly observable in the economy.

5) *INTERMEDIARY:* There is an intermediary in the economy, which collects individuals' endowments and invests them in the technology available. Let $Z$ be the fraction of funds invested in the short-term technology and $1-Z$ that in the long-term technology. The bank can be either a mutual fund, which operates in a competitive sector and make zero profits, or a profit maximising institution, which operates in a monopoly industry.

In exchange for deposits, the bank offers individuals a demandable contract that gives them the right to withdraw per unit of investment either the amounts $x_1$ at date 1 and $x_2$ at date 2 or the amounts $y_1$ at date 1 and $y_2$ at date 2. The amounts $x_i$ and $y_i$ with $i = 1, 2$ depend on the assumptions regarding the structure of the banking sector and depositors' risk aversion, as specified below. Even if $x_i$ and $y_i$ are designed at date 0 for the different depositors' types, type 2 depositors can imitate type 1 agents and ask for type 1 withdrawal whenever they find it optimal to so.

6) *DEPOSITORS' REPAYMENT:* At date 1 depositors decide whether to withdraw or leave their funds at the bank until date 2. All individuals withdrawing prematurely submit their withdrawal demands either sequentially or simultaneously and the bank uses the investment returns to repay them. If the bank is solvent, depositors receive the amount promised in the contract. Otherwise, depositors are repaid according to either a sequential service constraint

[2]This assumption will not hold in Chen (1999), where depositors know at date 0 whether they will receive the signal at date 1 or not.

rule or a pro-rata rule. In the former case, the bank serves its depositors sequentially on a first-come, first-served basis until it exhausts its resources. In the latter case, the bank deals with customers simultaneously and resources are distributed proportionally among them.

7) *TIMING:* Figure 1 summarises the timing of the model. At date 0 individuals deposit their funds at the bank in exchange for a demandable debt, which entitles them to withdraw a predetermined amount in each period. Then, the bank invests. At date 1 each depositor discovers his type and, in the case of stochastic long-term technology, the fraction of informed type 2 agents receives the signal $s$. Following this, all depositors make their withdrawal decisions.

If all depositors behave according to their true type (that is if each of them demands the withdrawal designed in the deposit contract for his type), the bank can satisfy their withdrawal demands and continue until date 2, when the long-term technology produces its return and claims are settled. Conversely, if depositors misreport their types and demand for type 1 withdrawal, a bank run is originated: The bank liquidates all the investment and is closed down, whilst depositors are repaid according to either the sequential service constraint or to the pro-rata rule.

| T=0 | | T=1 | | T=2 |
|---|---|---|---|---|
| deposit contract determined; agents deposit their funds and the bank invests | preference shock realised; eventual signal received | depositors make their withdrawal decisions; the bank may liquidate investment | investment returns realised | claims are settled |

Figure 1: Timing

### Equilibrium

The equilibrium consists of two elements: the deposit contract at date 0 and depositors' withdrawal decisions at date 1. The contract is signed at date 0 and is conditional neither on depositors' withdrawal decisions at date 1 nor on the signal $s$. It depends only on how many agents will turn to be of type 1 and of type 2, if known at date 0.[3]

a) *THE DEPOSIT CONTRACT:* The optimal contract choice problem for the deposit contract is solved by the vector $\Psi = \{c^*_{11}, c^*_{12}, c^*_{21,} c^*_{22}, Z\}$, where $c^*_{ij}$ is the optimal consumption for an agent of type $j$ at date $i$ and $Z$ is the fraction of deposits invested in the short-term technology. Under the assumptions of perfectly competitive banking system and deterministic long-term technology, the vector $\Psi$ satisfies:

$$U^* = \max_{\{c_{ij}\}} tU^1(c_{11}, c_{21}) + (1-t)U^2(c_{12}, c_{22}) \tag{1}$$

subject to:

$$tc_{11} + (1-t)c_{12} \leq Z \tag{2}$$

$$tc_{21} + (1-t)c_{22} \leq R(1-Z) \tag{3}$$

$$U^j(c_{1j}, c_{2j}) \geq U^j(c_{1k}, c_{2k}) \text{ with } j,k = 1,2 \text{ and } j \neq k. \tag{4}$$

Expressions (2) and (3) are the resource balance constraints for periods 1 and 2 respectively. Expressions (4) is the incentive compatibility constraint that guarantees that type $j$ depositors will prefer type $j$ withdrawal $(c_{1j}, c_{2j})$ to type $k$ withdrawal $(c_{1k}, c_{2k})$.

The maximisation problem changes slightly when the long-term technology is stochastic as the amounts $c_{21}$ and $c_{22}$ and the expected utility function depend on the realisation of the return $R$. Expression (1) becomes:

$$U^* = \max_{\{c_{ij}\}} \underset{\tilde{R}}{E} \left\{tU^1(c_{11}, c_{21}(R)) + (1-t)U^2(c_{12}, c_{22}(R))\right\} \tag{1'}$$

and the constraint (3) is given by:

$$tc_{21H} + (1-t)c_{22H} \leq H(1-Z) \tag{3'}$$

$$tc_{21L} + (1-t)c_{22L} \leq L(1-Z). \tag{3''}$$

[3] As it will be clear below, the inflexibility of the deposit contract to the signal $s$ can play an important role in determining the occurence of runs. The bank could indeed avoiding runs by making the contract conditional on the signal $s$ (Alonso (1996)).

The solution to the above maximisation problem depends on the assumptions concerning agents' risk aversion and the shape of their utility function. When the problem is solved, the bank sets its contract terms so that $x_1 = c^*_{11}$, $x_2 = c^*_{21}$, $y_1 = c^*_{12}$ and $y_2 = c^*_{22}$.

In the case of *smooth* utility functions and *risk averse* agents with *RRA>1*, the optimal contract gives:[4]

$$\begin{aligned} c^*_{12} &< c^*_{11} \\ c^*_{11} &> 1 \\ c^*_{21} &< c^*_{22} < R. \end{aligned}$$

In the case of *smooth* utility functions and *risk averse* agents with *RRA<1*, the optimal contract gives:[5]

$$\begin{aligned} c^*_{12} &< c^*_{11} < 1 \\ c^*_{21} &< c^*_{22}. \end{aligned}$$

Clearly, with *corner* utility functions, and *independently* of RRA, $c^*_{12} = c^*_{21} = 0$ obtains.

Finally, when depositors are *risk neutral*, they just consume their initial unit of deposit at date 0 and an amount equal to $R$ at date 2.[6]

The different assumptions concerning agents' risk aversion, the liquidation value of long-term technology $\ell$ and the rule used to repay depositors are the crucial "technical" factors triggering a bank run.

The determination of the deposit contract changes when the assumption of perfectly competitive banking system is removed and the bank operates so as to maximises its profits. In this case, depositors are offered a contract which allows them just to break even and the bank retains all the surplus.

b) *DEPOSITORS' WITHDRAWAL DECISIONS*: Depositors make their withdrawal decisions

[4]This resembles the solution in Diamond and Dybvig (1983), except for the fact that preferences are smooth here instead of corner.

[5]This is the solution in Jacklin and Bhattacharya (1988).

[6]With risk neutral depositors, there is no reason to solve the maximisation problem since there is no optimal risk sharing among depositors (see, for example, Chari and Jagannathan (1988)).

at date 1, after the realisation of the preference shock and the observation of the signal $s$ by informed type 2 depositors in the case of stochastic long-term technology. Type 1 agents report truthfully their type to the bank and demand the repayment $c_{11}^*$ at date 1 and $c_{21}^*$ at date 2 in the case of *smooth* preferences. The decision of type 2 agents is more complex and depends crucially on the assumptions concerning the long-run technology (deterministic versus stochastic, and the amount of the liquidation value), depositors' risk aversion and the repayment rule (sequential service constraint or pro-rata). Given the importance of depositors' withdrawal decisions in triggering bank runs and their differences across the different approaches, I will describe them in detail in the analysis of the single models.

## 2.3 The "First Generation" Literature

This literature emphasizes the role of banks as providers of flexibility to consumers who are uncertain about the timing of consumption. Demandable debt permits depositors to satisfy their unexpected consumption needs by giving them the right to withdraw at any point in time. However, demandable debt makes banks vulnerable to runs. The short-term characteristic of liabilities together with the illiquidity of long-term assets imply that banks may not have enough funds to satisfy all depositors' withdrawal demands. When this is the case, a bank run takes place and intermediaries liquidate all the assets and close down.

The first generation literature can be divided into two streams depending on the mechanism triggering depositors' decisions to withdraw prematurely. The first category, started by Diamond and Dybvig (1983), considers bank runs as an unfortunate and undesirable effect of demandable debt, which aims to provide depositors with insurance against liquidity shocks. Runs are purely random events ("sunspots"), which occur if depositors lose confidence in the bank's solvency despite the fact that its investments are not risky.

In contrast, the second category, to which Jacklin and Bhattacharya (1988) and Chari and Jagannathan (1988) belong, considers runs as triggered by a bilateral asymmetric information between the bank and its depositors. On the one hand, the bank does not observe depositors' type and therefore their true consumption needs; on the other hand, the bank invests in a stochastic technology and depositors are asymmetrically informed on the return of its investment.

In this framework, runs become systematic events triggered by the arrival of negative interim information on the bank's future solvency. More precisely, they are both information-induced and panic phenomena. The latter refers to the situation in which runs occur even though no one has received any negative information about the bank's investment returns.

The precise nature of runs derives from the assumptions on the liquidation value of the long-term technology and on depositors' risk aversion. If, as in Jacklin and Bhattacharya (1988), the long-term technology is totally illiquid and depositors are not very risk averse, bank runs are triggered by the decision of *informed* type 2 depositors to misreport their type and demand type 1 withdrawal profile, when they receive sufficiently negative information. Absent any information, a run would not occur. Conversely, if the long-term technology is only partially illiquid and depositors are risk neutral, as in Chari and Jagannathan (1988), the bank is forced to liquidate all its assets and close down only if *all* depositors demand their funds back at date 1. Depending on the realisation of the stochastic variables of the model, it can happen that a run occurs even though no one has received any information about the bank's future solvency.

To sum up, the first generation literature provides different reasons for the occurrence of bank runs. In Diamond and Dybvig (1983), a run is a self-fulfilling phenomenon which occurs when each depositor anticipates, for whatever reason, that the others will run. In contrast, the "asymmetric information" approach considers runs as more rational phenomena, triggered, as in Jacklin and Bhattacharya (1988), by the arrival of sufficiently negative interim information about the bank's future solvency, or even, as in Chari and Jagannathan (1988), even by the *fear* that some depositors have received such a negative information.

In the following sections, I analyse the three main first generation models in detail, mentioning some other works that have originated from them. I outline these models within the general model described in the previous section, pointing it out only the aspects which differ significantly from it.

### 2.3.1 The "Sunspot" Approach

The work by Diamond and Dybvig (1983) provides the first formal analysis of the role of banks as providers of insurance to depositors against the risk of a liquidity shock. The focus is on the characteristics of demandable debt. This is the means through which banks can offer liquidity to depositors, at the same time becoming vulnerable to runs.

In particular, the authors show that the deposit contract supports two equilibria. If depositors have confidence in the bank's solvency, the economy reaches the "good" equilibrium in which banks provide allocations superior to those of competitive markets. They can implement the optimal insurance contract and reach the optimal risk sharing among depositors. However, if depositors lose confidence in the bank's solvency, the economy reaches the "run" equilibrium in which all depositors panic and withdraw their funds prematurely as they anticipate the bank will run out of funds. The run equilibrium entails a real cost in terms of welfare reduction as it forces the bank to interrupt production, thus breaking the optimal risk sharing among depositors.

### Assumptions

1) T=0,1,2 and a single good.

2) *TECHNOLOGY:* There is a short-term storage technology and a long-term technology with a deterministic return $R$ and liquidation value $\ell = 1$. Table 2 summarises the technology available.

| TECHNOLOGY | $T=0$ | $T=1$ | $T=2$ |
|---|---|---|---|
| Short-term | $-1$ | 1 | 0 |
| Long-term | $-1$ | $\ell=1$ | $R>1$ deterministic |

Table 2: Technology

3) *DEPOSITORS:* Each depositor has a probability $t$ of becoming of type 1 and a probability $1-t$ of becoming of type 2. As $t$ is deterministic, there is no aggregate uncertainty and $t$ and $1-t$ correspond to the actual fractions of agents of type 1 and 2, respectively. Depositors have *corner* utility functions given by:

$$\begin{aligned} U^1(c_{11}) &= u(c_{11}) \text{ for type 1 agents} \\ U^2(c_{22}) &= u(c_{22}) \text{ for type 2 agents} \end{aligned}$$

where $c_{ij}$ is the consumption at date $i$ of an agent of type $j$. Depositors are risk averse ($u'' < 0$) with RRA>1, that is $-cu''(c)/u'(c) > 1$.

The ex ante utility function is therefore given by:

$$U = tu(c_{11}) + (1-t)u(c_{22}).$$

4) *INFORMATION:* There is no interim information regarding the value of the return $R$ since it is deterministic.

5) *INTERMEDIARY:* The bank is structured as a mutual fund and operates in a perfectly competitive sector. In exchange for deposits, it offers individuals a contract that gives them the right either to withdraw, per each unit of investment, $x_1$ at date 1 or $y_2$ at date 2.

6) *DEPOSITORS' REPAYMENT:* Depositors' withdrawal demands at date 1 are satisfied on a first-come, first-served basis as long as the bank has funds. Those at date 2 are instead subject to a pro-rata rule. Formally, the effective repayments to depositors are given by:

$$V_1(f_j, x_1) = \begin{cases} x_1 & \text{if } f_j < x_1^{-1} \\ 0 & \text{if } f_j \geq x_1^{-1} \end{cases}$$

$$V_2(f, x_1) = \max\{R(1 - x_1 f)/(1-f), 0\}$$

where $V_i$ is the repayment that depositors receive at date $i$, $f_j$ is the number of withdrawers' deposits serviced before agent $j$ as a fraction of total deposits; $f$ is the proportion of deposits withdrawn at date 1.

7) TIMING: Figure 2 summarises the timing of the model.

| T=0 | | T=1 | T=2 |
|---|---|---|---|
| \| | \| | \| | \| |
| deposit contract | \| | depositors | \| |
| determined; | preference | make their | remaining |
| agents deposit | shock realised | withdrawal | depositors |
| their funds; | | decisions; | repaid conditional |
| the bank invests | | bank may | on the aggregate |
| | | liquidate | withdrawals |
| Figure 2: Timing | | investiment | at date 1 |

**Equilibrium**

The equilibrium consists of the deposit contract at T = 0 and of depositors' withdrawal decisions at date 1.

a) *THE DEPOSIT CONTRACT:* The optimal contract choice problem for the deposit contract is solved by the vector $\Psi = \{c^*_{11}, c^*_{22}, Z\}$ satisfying:[7]

$$U^* = \max_{\{c_{ii}\}} tU^1(c_{11}) + (1-t)U^2(c_{22}) \tag{5}$$

subject to:

$$tc_{11} \leq Z \tag{6}$$

$$(1-t)c_{22} \leq R(1-Z) \tag{7}$$

$$U^1(c_{11}) \geq U^1(c_{22}) \tag{8}$$

$$U^2(c_{22}) \geq U^2(c_{11}) \tag{9}$$

where $c^*_{ij}$ is the optimal consumption for an agent of type $j$ at date $i$, $Z$ is the fraction of deposits invested in the short-term technology, constraints (6) and (7) are the resources constraints at dates 1 and 2, and constraints (8) and (9) are the incentive compatibility constraints.[8]

The bank sets the contract terms $x_1$ and $y_2$ equal to the optimal consumption levels $c^*_{11}$ and $c^*_{22}$, which satisfy:

$$u'(c^*_{11}) = Ru'(c^*_{22}) \tag{10}$$

$$tc^*_{11} + [(1-t)c^*_{22}/R] = 1 \tag{11}$$

$$1 < c^*_{11} < c^*_{22} < R. \tag{12}$$

Condition (10) equates the marginal utility to the marginal productivity; condition (11) is the total resource constraint. The relations in (12) derive from the assumptions of RRA>1 and $R > 1$.

[7]The assumption of *corner* utility function implies $c^*_{21} = c^*_{12} = 0$ in the optimal contract.
[8]Note that constraint (8) is redundant as, given the shape of the utility functions, it is always satisfied.

Expression (12) shows that the deposit contract offers insurance to depositors and is Pareto improving relative to the autarchy situation in which individuals invest directly and have a consumption stream $c_{11} = 1$ and $c_{22} = R$.[9]

b) *DEPOSITORS' WITHDRAWAL DECISIONS:* Depositors' withdrawal decisions at date 1 are determined as Nash equilibria in pure strategy. The deposit contract supports two Nash equilibria. In the first one, defined as the "good" equilibrium, depositors choose the withdrawal decisions embedded in the contract at date 0: The $t$ agents of type 1 withdraw at date 1, obtaining the repayment $x_1 = c_{11}^*$, and the $1-t$ agents of type 2 wait until date 2, as they anticipate that only type 1 agents withdraw prematurely. These withdrawal decisions constitute a Nash equilibrium since they satisfy the incentive compatibility constraints (8) and (9). Given others' decisions, each agent finds it optimal to choose the consumption stream designed for his own type.

In the other equilibrium, defined as the "run" equilibrium, all depositors panic and withdraw at date 1. Independently of their true types, all depositors report to be of type 1 and demand the repayment $c_{11}^*$. As $c_{11}^* > 1$ and the liquidation value of the long-term technology is equal to one ($\ell = 1$), the bank has to liquidate all its assets at date 1. Given the sequential service constraint, depositors joining the queue before $1/c_{11}^*$ others obtain the full amount $c_{11}^*$, and the $1 - 1/c_{11}^*$ agents arriving late are rationed and obtain nothing. This situation is a Nash equilibrium since each depositor finds it optimal to withdraw given that all others will do it. The incentive compatibility constraint for type 2 agents (expression (9)) is now violated because they get a higher utility by withdrawing at date 1 than by waiting until date 2, that is $1/c_{11}^* u^2(c_{11}^*) > u^2(0)$.

Runs are an equilibrium only if $x_1 = c_{11}^* > 1$. If $x_1 = 1$ and $\ell = 1$, runs would never occur as type 2 agents would not worry about others' behaviour and the bank's future solvency. Indeed, independently of how many depositors withdraw early, the bank would have enough resources

[9] In autarchy, a market opens each period where individuals can exchange their consumption good. Since all agents are identical at $T = 0$, they may want to exchange goods only at $T = 1$ and at $T = 2$. The price at $T = 1$ of consumption at $T = 2$ is equal to $R^{-1}$. At this price agents will never exchange goods and they cannot do better than producing goods for their own consumption. Therefore, agents choose $c_{11} = 1$ and $c_{22} = R$ in autarky, as type 1 agents will always interrupt production at $T = 1$ while type 2 agents will continue it until $T = 1$.

to fully satisfy type 2 depositors' demands at date 2.[10] In such equilibrium, the deposit contract would reach the same allocation as in the case of autarchy and the intermediary wouldn't deliver any Pareto improvement relative to competitive markets.

### Conclusions

The deposit contract allows the economy to reach two different equilibria. The first one, defined as the "good" equilibrium, is characterised by optimal risk sharing among depositors and constitutes a Pareto improvement relative to the autarchy situation. The second equilibrium, defined as the "run" equilibrium, occurs when all depositors panic and withdraw their funds prematurely. As a consequence, the bank is emptied of funds and the optimal risk-sharing mechanism breaks down. This happens because the face value of deposits is larger than the liquidation value of the bank's assets. The run equilibrium is a worse outcome for both types of depositors than the autarchy equilibrium since certain returns of 1 and $R$ for each type are replaced by uncertain returns of mean 1. Intermediaries are vulnerable to runs because they transform liquid deposits into illiquid assets and provide depositors with an insurance against their preference shock. Without these activities, they would not be vulnerable to runs but would not play any role in the economy. Thus, a deposit contract without the risk of runs does not provide any liquidity service.

Since the good equilibrium dominates holding assets directly, individuals choose to deposit their funds at the bank, provided that the run equilibrium occurs with a low enough probability. The selection between the two equilibria is assumed to depend on some commonly observed random variables, such as sunspots. If the sunspot does not occur, the economy reaches the good equilibrium. Otherwise, it reaches the run equilibrium. If the sunspot occurs with low probability, depositors are still willing to deposit their funds at the bank. This explanation of equilibrium selection is somewhat incomplete as it relies on some exogenous variables, the sunspots, which are absent in the model. Postlewaite and Vives (1987) investigate the equilibrium selection problem further and show the existence of a unique equilibrium that involves a positive probability of runs. In their analysis the bank run is associated with a Prisoner's Dilemma situation in which depositors withdraw at date 1 for self-interest rather than for con-

[10] Formally, with $x_1 = 1$, it holds $V^1(f_j, 1) < V^2(f, 1)$ for any $0 \leq f_j \leq f$.

sumption needs. This approach has the merit of eliminating the exogenous sunspot elements and the multiplicity of equilibria. However, recognising the suboptimality of demandable debt, it fails to explain its use in the real world.

The existence of a run equilibrium in Diamond and Dybvig is not a robust result. Type 2 depositors' decision to withdraw prematurely depends crucially on the rule used for depositors' repayment (sequential service constraint), on the assumptions regarding depositors' relative risk aversion coefficient ($RRA > 1$ and hence $c_{11}^{*} > 1$), the liquidation value of long-term technology ($\ell = 1$) and the shape of depositors' utility function (corner preferences). The run equilibrium can be eliminated by modifying any of these assumptions. Bhattacharya and Thakor (1993) show that using a pro-rata rule for depositors' repayment at date 1, instead of the sequential service constraint rule, would remove the run equilibrium since type 2 depositors would not have any incentive to withdraw prematurely;[11] Jacklin and Bhattacharya (1988) argue that the run equilibrium is eliminated if depositors are not too risk averse, that is if the relative risk aversion coefficient is less than 1. In this case, maintaining the assumption of $\ell = 1$, the bank would not run out of funds at date 1 even if more than $t$ depositors withdrew, since the repayment promised to type 1 depositors, $c_{11}^{*}$, would be less than 1.

Bank runs in Diamond and Dybvig can be eliminated also by some modifications of the deposit contract. If the fraction $t$ of type 1 depositors is deterministic, the bank can predict exactly how many type 1 depositors there will be and can therefore promise to redeem on demand at date 1 only $t$ withdrawal demands and postpone the others at date 2. This 'suspension of convertibility' clause guarantees optimal returns and eliminates the occurrence of runs as its anticipation prevents type 2 depositors from withdrawing prematurely.

The suspension of convertibility arrangement breaks down if the fraction $t$ becomes stochastic. This is because the bank cannot predict how many depositors turn to be of type 1 and is therefore unable to select an appropriate threshold of withdrawals at which to suspend payments at date 1. However, runs can still be eliminated when $t$ is random by introducing government deposit insurance. This arrangement guarantees that the promised amount will be paid to all agents, thus making it convenient for type 2 depositors not to withdraw prematurely. The deposit guarantee is honored through taxes imposed on depositors withdrawing early. Unlike

[11] On the importance of the sequential service constraint, see also Wallace (1988).

the bank, which must provide sequential service constraint, the government can impose taxes on an agent *after* he has withdrawn. Therefore, the amount of the tax depends on the realised total value of withdrawals $f$. Depositors withdrawing early are taxed if $f > \bar{t}$, that is if the total withdrawals are greater than the maximum realisation of the fraction $t$ of type 1 depositors, and not otherwise. The amount of the tax is such that the after-tax proceeds at date 1 never exceed those at date 2. Since depositors are concerned with after-tax payoffs, no type 2 agents withdraw at date 1. Thus, deposit insurance satisfies the incentive-compatible constraints and achieves the optimal outcome as unique equilibrium if the government can finance it through non-distorsive taxes.

### 2.3.2 The "Asymmetric Information" Approach

The seminal paper of Diamond and Dybvig (1983) provides only a partial justification for the occurrence of runs. As the investment returns are certain, there is no reason why depositors should lose confidence in the bank's solvency and panic. Therefore, runs have a pure speculative origin. The story that agents observe some random variable in the economy, as sunspots, is not well founded and does not find support in the empirical evidence. Gorton (1988) finds that runs occurred in the US before the introduction of the federal deposit insurance occurred in time of recessions and were caused by changes in fundamentals. In particular, the author finds that bank crises were tied to low performance of banks' loan portfolios or high failure rates of small firms. Thus, runs were triggered by the arrival of some negative information regarding the future performance of banks.

The asymmetric information approach aims to explain fundamental bank runs (which are justified by a poor performance of the bank's investment) and to combine them with speculative runs (which are due to fears of a poor performance of the bank's investments). The main element of the analysis is the uncertainty of the bank's investments and the information that some type 2 depositors may receive at date 1. A fundamental run takes place when, as in Jacklin and Bhattacharya (1988), it is caused by *informed* type 2 depositors' decision to demand type 1 withdrawals. A run combines fundamental and speculative elements when, as in Chari and Jagannathan (1988), it is triggered by *all* (informed and uninformed) type depositors' decision to demand type 1 withdrawal profile.

In what follows, I analyse the papers of Jacklin and Bhattacharya (1988) and of Chari and Jagannathan (1988), describing in detail the different factors triggering a run in the two models.

### Information-based Bank Runs

Jacklin and Bhattacharya (1988) regard runs as triggered by the arrival of negative interim information regarding the bank's future performance. The main elements of the analysis are the assumptions concerning depositors' utility functions, the uncertainty of the bank's investments and the interim information on the bank's future performance available to some depositors.

### Assumptions

1) T=0,1,2 and a single good.

2) *TECHNOLOGY:* The long-term technology is now stochastic and cannot be liquidated early. Table 3 summarises the technologies available and their returns.

| TECHNOLOGY | $T=0$ | $T=1$ | $T=2$ |
|---|---|---|---|
| Short-term | $-1$ | 1 | 0 |
| Long-term | $-1$ | $\ell=0$ | $\widetilde{R}=\begin{cases} H>1 & p \\ 1>L>0 & 1-p \end{cases}$ stochastic |

Table 3: Technology

3) *DEPOSITORS:* Each depositor has a probability $t$ of becoming of type 1 and $1-t$ of becoming of type 2. The variable $t$ is deterministic. Both types of depositors have *smooth* preferences given by:

$$\begin{aligned} U^1(c_{11},c_{21}) &= u(c_{11})+\rho_1 u(c_{21}) \text{ for type 1 agents} \\ U^2(c_{12},c_{22}) &= u(c_{12})+\rho_2 u(c_{22}) \text{ for type 2 agents} \end{aligned}$$

where $c_{ij}$ is the consumption at date $i$ of an agent of type $j$, $\rho_i$ is the intertemporal discount factor with $1>\rho_2>\rho_1>0$. Individual are risk averse ($u''<0$) with RRA<1, that is $-cu''(c)/u'(c)<1$.

The ex ante utility function is given by:

$$U = tU^1(c_{11}, c_{21}) + (1-t)U^2(c_{12}, c_{22}).$$

4) *INFORMATION:* At date 1 a fraction $\alpha$ of type 2 agents, defined as informed, observes a signal $s$ on the future value of bank's assets. The signal is partial and is described by the distribution of posterior beliefs $\widehat{p}$ about the probability of success, i.e. that the bank's asset return is $R$.

5) *INTERMEDIARY:* The bank is structured as a mutual fund and operates in a perfectly competitive sector. It offers depositors a contract which gives them the right to withdraw per unit of investment either the amounts $x_1$ at date 1 and $\widetilde{x_2}$ at date 2 or the amounts $y_1$ and $\widetilde{y}_2$ at dates 1 and 2, respectively. The uncertainty of the promised repayment at date 2 is due to the stochastic character of the long-term technology.

6) *DEPOSITORS' REPAYMENT:* Depositors withdrawing at date 1 are repaid according to the sequential service constraint rule. Those at date 2 are repaid proportionally.

7) *TIMING:* Figure 3 summarises the timing of the model.

| T=0 | | T=1 | | T=2 |
|---|---|---|---|---|
| deposit contract determined; agents deposit their funds; the bank invests | preference shock realised; signal received by informed type 2 depositors | depositors make their withdrawal decisions; the bank may liquidate its assets | return $R$ realised | remaining depositors repaid conditional on the bank's solvency |

Figure 3: Timing

### Equilibrium

The equilibrium consists of the deposit contract at date 0, and of the withdrawal decisions of type 1 depositors and of informed type 2 depositors at date 1.

a) *THE DEPOSIT CONTRACT:* The optimal contract choice problem in the absence of interim information is solved by the vector $\Psi = \left\{c^*_{11}, c^*_{12}, c^*_{21,} c^*_{22}, Z\right\}$, where $c^*_{ij}$ is the optimal consumption at date $i$ of an agent of type $j$ and $Z$ is the investment in the short-term technology, satisfying:

$$U^* = \max_{\{c_{ij}\}} \underset{\widetilde{R}}{E} \left\{tU^1(c_{11}, c_{21}(R)) + (1-t)U^2(c_{12}, c_{22}(R))\right\} \tag{13}$$

subject to:

$$tc_{11} + (1-t)c_{12} \leq Z \tag{14}$$

$$tc_{21}(R) + (1-t)c_{22}(R) \leq R(1-Z) \qquad \forall R \tag{15}$$

$$U^j(c_{1j}, c_{2j}) \geq U^j(c_{1k}, c_{2k}) \text{ for } j, k = 1, 2 \text{ and } j \neq k \tag{16}$$

where constraints (14) and (15) are the resources constraints at dates 1 and 2, and constraint (16) is the incentive compatibility constraint. The optimal consumption levels determined in the contract are $x_1 = c^*_{11}$, $x_2 = c^*_{21}$, $y_1 = c^*_{12}$, and $y_2 = c^*_{22}$. If $R = H$ is realised at date 2, the bank is able to pay depositors the promised amounts $x_2$ and $y_2$; conversely, if $R = L$ is realised at date 2, the bank is insolvent and able to pay only a fraction $L/H$ of the promised amounts. Since depositors have RRA<1, the optimal consumption levels satisfy:

$$1 > c^*_{11} > c^*_{12} \text{ and } c^*_{22} > c^*_{21}. \tag{17}$$

Unlike Diamond and Dybvig (1983), type 1 agents are now promised an amount smaller than one at date 1. However, given the assumption of total illiquidity of the long-term technology, the bank is still vulnerable to runs. If more than $t$ depositors ask for the type 1 withdrawals at date 1, then there will not be enough resources to satisfy them fully and the bank will be forced to close down. This situation is possible because the deposit contract is determined at date 0 ignoring the impact of any interim information that some depositors receive on their preferred withdrawal profiles (see below for details).

b) *DEPOSITORS' WITHDRAWAL DECISIONS:* At date 1 depositors choose between one of the two consumption profiles, $(x_1, x_2)$ and $(y_1, y_2)$, promised in the deposits contract. Their choice depends on the realisation of the preference shock and on the information they receive.

The $t$ agents that turn out to be of type 1 choose the profile $(x_1, x_2) = (c_{11}^*, c_{21}^*)$. The $\alpha$ informed type 2 depositors update the success probability of the bank's project from $p$ to $\widehat{p}$ on the basis of the signal $s$ they receive. Given this revised probability assessment, they update their expected utility with type 2 withdrawal and type 1 withdrawal and choose the profile which makes them better off. In other words, they update their incentive compatibility constraint, which is given by:

$$\widehat{E}\left[U^2(c_{12}, \widetilde{c}_{22})\right] \gtrless \widehat{E}\left[U^2(c_{11}, \widetilde{c}_{21})\right] \tag{18}$$

where $\widehat{E}$ indicates the expectation calculated using the posterior $\widehat{p}$. Type 2 depositors choose the type 1 withdrawal profile whenever the inequality in (18) holds with the sign $<$. When this is the case, an information-based run takes place. The bank does not have enough funds to satisfy the demand for type 1 withdrawals at date 1 since the long-term technology is illiquid. Depositors are repaid randomly: The first $t$ individuals arriving in the queue receive the full amount $c_{11}^*$ while the remaining $1-t$ agents only get the amount $c_{12}^*$. This resembles a suspension of convertibility clause, after $t$ depositors have been dealt with.

Unlike in Diamond and Dybvig (1983), a run would not occur without the arrival of interim information. With a liquidation value of the long-term technology equal to zero and a relative risk aversion coefficient less than one, which implies $1 > c_{11}^* > c_{12}^*$, the incentive compatibility constraint in (18) would never be violated and, thus, type 2 depositors would never prefer type 1 withdrawal.

### Conclusions

Introducing a random return on bank's investments, Jacklin and Bhattacharya (1988) explain the occurrence of runs as the rational response of depositors to the arrival of negative information on the bank's future performance. A run occurs when informed type 2 depositors observe a sufficiently negative signal $s$. In other terms, when the return $R$ is observed to be negative with probability $\widehat{p} < \overline{p}$ where $\overline{p}$ is the threshold level which makes the expression (18) hold as an equality. This is because the arrival of a bad signal modifies informed type 2 depositors' incentive compatibility constraint and induces them to prefer the type 1 withdrawal. The threshold level $\overline{p}$ is positively correlated with the variance of the return $R$, that is with the dispersion between $H$ and $L$.[12] Thus, the higher the variance of $R$, the higher the threshold $\overline{p}$, and the more likely a run is to occur.

When a run occurs, the welfare of both types of agents decreases. This leads the authors to address the question of the relative performance of demand deposit economies versus equity economies. The comparison of the two economies for different parameterizations of the model shows that demand deposits would perform better for a low variance of $R$, whereas the equity economy would be preferred for a large dispersion of $R$. This is because when the variance of $R$ is high, the probability of runs increases, thus decreasing welfare.

The inflexibility of the deposit contract to the arrival of interim information is a major reason for the occurrence of runs in Jacklin and Bhattacharya. If the bank took such information into account, it could prevent runs by making the contract incentive compatible after type 2 agents have become informed. However, banks may prefer not to avoid runs if the costs of modifying the contract are too high. Alonso (1996) shows that the choice between contracts with runs and contracts without runs depends on the parameters of the model. Deposit contracts with runs are socially preferable if the probability of informed type 2 depositors receiving a bad signal is low enough. This is because, in order for informed type 2 depositors not to withdraw after receiving a bad signal, depositors' payoffs have to be significantly modified in all states of nature. Thus, when the occurence of receiving a bad signal is not very likely, a high loss is incurred with high probability whilst the gain of avoiding runs is only realised with low probability.

The possibility of structuring deposit contracts without runs also sheds new light on the comparison between demandable deposit and equity economies. Alonso (1996) shows that deposit arrangements are superior to equity arrangements in most cases. This is because when deposit contracts achieve allocations which are inferior to those generated by equity contracts, the bank can modify the deposit contract to avoid runs and dominate equity performance (for example in the case of a high variance of the return of the bank's investments).

[12] Note that in the original paper, the threshold level $\overline{p}$ is inversely correlated with the variance of the returns. This is because, in line with the general model described in the first sectsion, I consider the posterior success probability of $R$ while the paper refers to the posterior failure probability of $R$.

### Panic Bank Runs

The paper of Chari and Jagannathan (1988) represents a synthesis of the models examined up to now, since it combines fundamental and speculative runs. The focus of the analysis is the signal-extraction problem faced by uninformed depositors in their withdrawal decision. The key assumptions are that the return of the bank's investment is stochastic as well as the fraction of depositors turning to be of type 1 and the fraction of type 2 depositors becoming informed at date 1. Type 2 depositors who remain uninformed observe the total amount of withdrawals at date 1 and base on it their withdrawal decision. However, since the realisation of the random variables characterising the model is not observable, uninformed type 2 depositors may not be able to infer correctly the bank's future performance. In particular, they may not be able to distinguish whether a long queue is due to some informed type 2 depositors withdrawing early or to a large proportion of type 1 depositors only. The equilibrium of the model has the property that runs can be both fundamental and speculative. A speculative, or "panic", run occurs when all depositors withdraw prematurely for *fear* that some other depositors have received a bad signal on the bank's investment returns in circumstances in which there is no information in the economy.

In what follows, I describe the model in detail. In doing this, I will not follow the original paper of Chari and Jagannathan, in which individuals invest directly in the technology, but, in line with the general framework, I consider the existence of an intermediary which collects deposits and invests them in the technology available. Notice that the intermediary does not really play any role in this model. Indeed, as depositors are risk neutral and the banking sector is perfectly competitive, the intermediary does not provide any liquidity service and optimal risk sharing. Neither it extends loans nor monitors them. In line with the first generation literature, this implies that runs are still seen as disruptive phenomena when they are not motivated by negative information on the bank's future prospects. In the next chapter, I will develop a model in which the banker acts as delegated monitor and has the possibility to make an improper use of depositors' money. Demandable debt will then be considered as a possible incentive device for the banker to act in line with depositors' interests. The introduction of an agency problem between the bank and its depositors will lead to different conclusions on bank run efficiency and policy implications.

**Assumptions**

1) T=0,1,2 and a single good.

2) *TECHNOLOGY:* There is now only one long-term technology available with stochastic return. Its liquidation value depends on how many depositors withdraw early: Each unit invested yields a return $\ell = 1$ if the fraction of depositors withdrawing at date 1, $W$, is lower than a certain threshold $\overline{W}$, whilst it yields a return $\overline{\ell} < 1$ otherwise. Table 4 summarises the technology available.

| TECHNOLOGY | $T=0$ | $T=1$ | $T=2$ |
|---|---|---|---|
| Long-term | $-1$ | $\ell = \begin{cases} 1 & \text{if } W < \overline{W} \\ \overline{\ell} < 1 & \text{otherwise} \end{cases}$ | $\widetilde{R} = \begin{cases} H > 1 & p \\ L = 0 & 1-p \end{cases}$ |

Table 4: Technology

3) *DEPOSITORS:* A stochastic fraction $\widetilde{t}$ of depositors turn out to be of type 1 at the beginning of date 1; the remaining $1-\widetilde{t}$ depositors are of type 2. The variable $\widetilde{t}$ is discrete, assuming three values $t \in \{0, t_1, t_2\}$ with probability $q_0, q_1$ and $q_2$, respectively. Both types are risk neutral and their preferences are given by:

$$\begin{aligned} U^1(c_{11}) &= c_{11} \quad \text{for type 1 agents} \\ U^2(c_{22}) &= c_{22} \quad \text{for type 2 agents} \end{aligned}$$

where $c_{ij}$ is the consumption of an agent of type $j$ at date $i$.

4) *INFORMATION:* A stochastic fraction $\widetilde{\alpha}$ of type 2 depositors, defined as informed, observes at date 1 a signal $s$ on the future value of the bank's assets. The signal is perfect, that is $s = \{H, 0\}$. The variable $\widetilde{\alpha}$ takes on two values, $\widetilde{\alpha} = \{\overline{\alpha}, 0\}$, with probability $r$ and $1-r$, respectively. The remaining $1-\widetilde{\alpha}$ type 2 depositors remain uninformed. The realisations of the three random variables $\widetilde{t}$, $\widetilde{\alpha}$ and $\widetilde{R}$ as well as the signal received by informed type 2 depositors are not observable. The only public variable in the economy is the amount of aggregate withdrawals at date 1, defined as $W$. This is not always perfectly informative of the signal $s$ received by the informed depositors if the following restrictions hold:

$$t_1 = \overline{\alpha} \tag{19}$$

$$t_2 = t_1 + \overline{\alpha}(1 - t_1). \tag{20}$$

5) *INTERMEDIARY:* The bank operates in a perfectly competitive sector. In exchange for deposits, it offers individuals a contract which gives them the right to withdraw, per unit deposited, either the amount $x_1$ at date 1 or the amount $y_2$ at date 2.

6) *DEPOSITORS' REPAYMENT:* Depositors withdrawing at date 1 submit their demands simultaneously. If the bank has not enough resources, depositors are repaid according to a pro-rata rule. Bank resources at date 1 depend on the liquidation value of the technology $\ell$, which in turns depends on the threshold level $\overline{W}$. The latter is assumed to be equal to $t_2$.

7) TIMING: Figures 4 illustrates the timing of the model.

| T=0 | | T=1 | | T=2 |
|---|---|---|---|---|
| deposit contract determined; agents deposit their funds; the bank invests | preference shock realised; fraction of informed depositors realised; signal received by informed type 2 depositors, if any | depositors make their withdrawal decisions; the bank may liquidate its assets | return $R$ realised | remaining depositors repaid conditional on the bank's solvency |

Figure 4: Timing

### Equilibrium

The equilibrium consists of the deposit contract at date 0, and of depositors' withdrawal decisions at date 1.

a) *THE DEPOSIT CONTRACT :* The assumption of risk neutral individuals together with that of a competitive bank simplifies greatly the analysis of the deposit contract.[13] The bank fixes $x_1 = c_{11}^* = 1$ and $y_2 = c_{22}^* = H$. Given $H > 1$, the promised repayments satisfy the incentive-compatibility constraints at date 0.

b) *DEPOSITORS' WITHDRAWAL DECISIONS:* At date 1 depositors choose between one of the two consumption profiles, $x_1$ and $y_2$, offered in the deposit contract. Their decisions depend on their type and on the information they have.

Type 1 depositors demand $x_1 = c_{11}^* = 1$ to satisfy their consumption needs. Informed type 2 depositors decide on the basis of the signal they receive: They leave their funds at the bank until date 2 if $s$ is positive, whilst they withdraw at date 1 if $s$ is negative.

Uninformed type 2 depositors observe the aggregate withdrawals at date 1, $W$, and update their expected utility accordingly. Given this revised assessment, they may prefer to withdraw at date 1 instead of waiting until date 2. Since $W$ is not perfectly correlated with $s$, however, uninformed type 2 depositors are not always able to infer the real state of their world. Thus, they may commit mistakes in their withdrawal decisions.

A run occurs in the rational-expectations equilibrium of the model when all depositors withdraw at date 1, that is when $W = 1$. A *panic* run occurs when all depositors withdraw even if no one has received any information about the bank's future solvency. Formally, a run is a panic if the equilibrium aggregate withdrawals $W$ at date 1 are equal to 1 for at least one state in which $\alpha = 0$. Under restrictions (19) and (20), a panic run can occur only if uninformed type 2 depositors are confounded, when observing a long queue, between a large number of depositors withdrawing for consumption needs and the possibility that some informed depositors have received poor information. A panic run is inefficient since, under the assumption $pH + (1-p)0 > 1$, it would be socially optimal to let the bank continue the investment.

## Conclusions

Chari and Jagannathan model runs as an equilibrium phenomenon in a framework where all equilibria are characterised by runs. The essential of the paper is the confusion faced by

[13]Since there is only one technology available, the bank does not have to choose how to allocate the funds between short- and long-term investments. Thus, the deposit contract consists only of the amounts promised to depositors.

uninformed depositors. Individuals decide to withdraw prematurely for different reasons: Some do it for consumption needs, some others because they receive bad information on the future returns of bank's investment, some others, denominated uninformed, withdraw prematurely because they fear the bank will be insolvent.

Note that a run can never occur in states in which some type 2 depositors receive the signal (that is $\alpha = \overline{\alpha}$) and the signal is positive ($s = H$). A run can only occur in this framework when informed type 2 depositors receive a negative information or when there is no information in the economy. Thus, only panic runs are inefficient and, since they occur in the absence of information, they resemble the sunspot runs in Diamond and Dybvig. Such an observation will be useful in the next chapter where I will develop a model in which inefficient runs will occur *despite* the fact that some depositors know with certainty that the bank will be solvent.

Chari and Jagannathan do not provide a justification for the existence of banks. Indeed, they neither provide liquidity insurance as depositors are risk neutral, nor they monitor their debtors. All the bank can do is to prevent panic runs by suspending convertibility at date 1 after more than $t_1$ depositors have withdrawn. This arrangement saves on the liquidation costs associated with the interruption of valuable assets but induces a loss in terms of possible continuation of valueless assets and of rationing of type 1 agents when their proportion turns to be higher than $t_1$. The suspension of convertibility improves upon the equilibrium allocations only when the first effect dominates. This depends, in turn, on the model's parameters and especially on the liquidation value of the long-term technology and the probability of a fraction of depositors larger than $t_1$.

### 2.3.3 A Numerical Example

I now describe a numerical example that illustrates how runs are generated in the first generation models.

I consider a simple framework similar to Diamond and Dybvig (1983) in which the long-term technology has a return $R$, depositors have *corner* utility functions with RRA>1 and there is a fraction $t$ of agents of type 1 and $1-t$ of agents of type 2.[14]

[14]For simplicity, I use this framework to analyse all the models presented above even if it is not always coehrent with them. For example, concerning the shape of depositors' utility functions, I use *corner* preferences also to

**1) The "sunspot" approach**

I also assume:

- $R = 1.5$ deterministic and $\ell = 1$;

- $t = 0.25$ deterministic.

In line with the data described, the deposit contract specifies:

$$c_{11}^{*} = 1.2$$

$$Z = 1.2 \cdot 0.25 = 0.3$$

$$c_{22}^{*} = \frac{(1-Z)R}{1-t} = \frac{0.7 \cdot 1.5}{0.75} = 1.4.$$

If a fraction $f$ of depositors greater than $\widehat{f} = 0.5$ withdraws prematurely, all other depositors find it optimal to withdraw at date 1 as well and a run takes place. Indeed, as $\widehat{f} = 0.5$ is the number of withdrawals at date 1 that makes type 2 depositors indifferent between withdrawing and waiting, any $f > 0.5$ makes it optimal for them to join the queue at date 1 and trigger a run.

Formally, for $f = 0.5$, the bank has to liquidate $1.2 \cdot 0.5 = 0.6$ units of investment and, therefore, only 0.4 units are left at date 2 to repay remaining depositors, who would get:

$$c_{22} = \frac{0.34 \cdot 1.5}{0.4} = 1.2$$

and would therefore be indifferent between withdrawing at date 1 or at date 2.

**2) The "asymmetric information" approach**

*(i) Information-based runs*

I assume now:

- $\widetilde{R} = \begin{cases} H = 1.5 & p \\ L = 1.2 & 1-p \end{cases}$ stochastic and $\ell = 0$;

- $t = 0.25$ deterministic.

---

analyse Jacklin and Bhattacharya (1988) in which, instead, they are *smooth.*

Type 2 depositors receiving a negative signal $s$ on the future value of the bank's assets prefer to misreport their type and choose the type 1 withdrawal profile. This is because, if they waited, they would get:

$$c_{22} = \frac{(1-Z)L}{1-t} = \frac{0.7 \cdot 1.2}{0.75} = 1.12 < 1.2.$$

More precisely, as the signal is not perfectly informative of the bank's future solvency but only indicates the posterior probability $\widehat{p}$ of success of the bank's assets, informed type 2 depositors prefer to withdraw when $\widehat{p}$ is lower than the threshold level $\overline{p}$, which makes them indifferent between withdrawing early and waiting until date 2. The threshold level is the solution to:

$$\overline{p} : U(c_{11}) = E\left[U(\frac{\widetilde{R}(1-Z)}{1-t})|s\right].$$

*(ii) Panic bank runs*

I now assume the following:

- $\widetilde{R} = \begin{cases} 1.5 & p = 0.9 \\ 1.1 & 1-p = 0.1 \end{cases}$ stochastic and $\ell = 1$;
- $\widetilde{t} = \begin{cases} 0.25 & q = 0.9 \\ 0.4 & 1-q = 0.1 \end{cases}$ stochastic;
- $\widetilde{\alpha} = \begin{cases} 0.15 & r = 0.5 \\ 0 & 1-r = 0.5 \end{cases}$ stochastic.

If uninformed type 2 depositors observe a queue $W = 0.4$, they do not know whether it is composed by the sum of $t = 0.25$ and $\alpha = 0.15$ or only by $t = 0.4$. Thus, they expect to obtain the following repayments:

$$R = 1.1, t = 0.25, \alpha = 0.15 \quad \rightarrow \quad \frac{(1-0.48)1.1}{0.6} = 1.3$$

$$R = 1.5, t = 0.4, \alpha = 0 \quad \rightarrow \quad \frac{(1-0.48)1.5}{0.6} = 0.953$$

where $0.48 = 0.4 \cdot 1.2$. Then, given $r = 0.5$, uninformed type 2 depositors' expected repayment from waiting until date 2 is given by:

$$0.5 \cdot 1.3 + 0.5 \cdot 0.953 = 1.126 < c_{11} = 1.2.$$

Given this, uninformed type 2 agents prefer to misreport their type and withdraw at date 1 when they observe $W = 0.4$. This triggers a bank run which is a panic run if the queue is formed only by type 1 agents and no one has received any information on $R$.

Finally, it is interesting to note that, absent any interim information, there would be no sunspot runs as in Diamond and Dybvig (1983). For example, for the same queue $W = 0,4$, type 2 depositors would obtain at date 2:

$$c_{22} = \frac{(1 - 0.4 \cdot 1.2)}{1 - 0.4} [0.9 \cdot 1.5 + 0.1 \cdot 1.1] = \frac{0.52 \cdot 1.46}{0.6} > 1.2$$

and therefore they would not have any incentive to withdraw at date 1.

## 2.4 The "Second Generation" Literature

The first generation literature provides an explanation of both the existence of banks and their vulnerability to runs. Banks provide flexibility to depositors who are uncertain in the timing of consumption. The maturity-mismatch between investments and deposits exposes banks to the possibility of massive withdrawals. Runs involve social costs as they break the risk sharing mechanism among depositors and/or force the liquidation of assets which may be valuable.

A question then arises spontaneously: Why have banks always used demandable debt if it involves high costs? Is it because of its liquidity characteristics only? Calomiris and Kahn (1991) and Chen (1999), provide an answer to these questions. The focus of these models is to explain the role of demandable debt as part of an incentive scheme for disciplining bankers. This will also be the focus of the model I will develop in chapter 3.

The starting point is the agency relation that arises between the banker and his depositors after the deposit contract is signed. Agents entrust their money to the banker that choose how to allocate them among different uses. The banker has an informational advantage in determining which investments are most profitable but he has also the possibility to act against depositors' interests. For example, he can invest in projects which are not socially desirable or he can misuse investment returns for his own purposes. In this context, demandable debt can provide an incentive-compatible solution to the banker's moral hazard problem. The right to take money out of the bank at any time gives depositors the possibility to register their lack of confidence in the activities of the banker. The threat of a run induces the banker to act in line with depositors' interest, so to minimise asset liquidation and avoid the bank's closure. In other words, demandable debt allows the banker to precommit to a behaviour he would not follow otherwise.

As in the asymmetric information approach, runs are depositors' rational responses to the arrival of negative information on the bank's future solvency and are therefore efficient if they force the liquidation of assets that would yield lower returns if continued until maturity. Further, runs have now the additional purpose of taking assets away from the banker, thus preventing him from misbehaving. However, runs can still induce costs and act as imperfect incentive devices. As in the asymmetric information approach, this emerges when depositors are imperfectly informed on the future value of the bank's assets at the time of making their withdrawal decisions.

In the following, I present the papers by Calomiris and Kahn (1991) and Chen (1999). In the first one, demandable debt is part of an optimal arrangement for monitoring banks. Runs are efficient mechanisms to prevent the banker from absconding in states of the world where he would do it otherwise. They occur only when informed depositors have received negative information on the bank's future performance. In Chen (1999), demandable debt acts as an incentive device for the banker but implies some costs as depositors at one bank can start to run when they observe many other banks failing. Even if this is not a precise source of information about their bank's future solvency, uninformed depositors have strong incentives to respond to it, in order not to be rationed by the sequential service constraint.

In short, market discipline is inefficient in Chen (1999) because runs are contagious. In chapter 3 of this thesis, I will discuss another possible source of market discipline inefficiency, which originates *within* one bank.

The question of whether runs are efficient control devices for moral hazard problems is relevant for the debate on whether the banking sector needs to be regulated or not. Extra market arrangements, such as the suspension of convertibility, deposit insurance and lender of last resort, are desirable only if needed to correct market imperfections.

### 2.4.1 Bank Runs as Perfect Discipline Device

The paper of Calomiris and Kahn (1991) aims to explain the emergence of demandable debt as the main means of financing banks although it entails a cost in terms of bank suspension and liquidation. A plausible explanation is that demandable debt provides the correct incentives for the banker to act in line with depositors' interests. The banker has the possibility of absconding with the investment returns and not repaying loans. Depositors have the possibility of acquiring a costly signal on the future value of the bank's investment and of using it in making their withdrawal decisions. If the signal is negative, they prefer to withdraw and liquidate the bank as they anticipate that the banker would abscond otherwise. The threat of runs induces the banker to act in line with depositors' interests and to attract deposits.

The assumption of costly acquisition of information provides a rationale for the sequential service constraint rule imposed in depositors' repayment. As depositors are repaid according to their position in the queue, they have an incentive to monitor the bank and be the first in line if necessary. Thus, the sequential service constraint avoids the free riding problem in information gathering.

I now describe the model in detail. Given its complexity and its significant difference with the other papers presented, I restrict the attention to the simple case of a single agent who deposits at a monopoly bank. In this case, a run coincides with the depositor's decision to withdraw at date 1 and liquidate the bank.

#### Assumptions

1) T=0, 1, 2 and a single good.

2) *TECHNOLOGY:* There is an investment opportunity that yields a random return $\widetilde{R} = \{H, L\}$ at date 2 for each unit invested, with $H > L > 0$. The probability of the high return is $p$. The project is liquidated at date 1 if the bank is liquidated. In this case, the value of the investment is reduced by a proportion $\eta \in [0,1]$, which can be interpreted as the tax due to liquidation. Table 5 summarises the technology available.

| TECHNOLOGY | $T=0$ | $T=1$ | $T=2$ | |
|---|---|---|---|---|
| Long-term | $-1$ | $\ell = 1-\eta$ | $\widetilde{R} = \begin{cases} H & p \\ L & 1-p \end{cases}$ | stochastic |

Table 5: Technology

3) *DEPOSITORS:* There are many risk neutral agents, each endowed with one unit of the good. One of them deposits his funds at the bank at date 0 if he is promised a repayment at date 2 at least equal to his reservation level $S$.

4) *INFORMATION:* After the agent deposits his funds at the bank, he has the possibility of acquiring, at the cost $k$, a signal $s$ on the future return $R$. The signal can take on two values $s = \{g, b\}$. The probability of $R = H$ contingent on $s$ is $p_s$ with:

$$p_g > p > p_b.$$

The indicator variable $e \in \{0,1\}$ represents the depositor's choice: $e = 1$ if he invests in the signal, $e = 0$ otherwise. The informed depositor's decision to withdraw at date 1 after observing $s$ corresponds to a run and implies the liquidation of the bank.

The decision of whether to invest in the signal or not and the value of the signal is private knowledge. Also, the realised return of the bank's investment at date 2 is privately observed by the bank manager. Thus, the deposit contract cannot be directly conditional on the value of $R$.

5) *INTERMEDIARY:* There is a bank managed by a risk neutral monopoly banker. He collects deposits at date 0 in exchange for a contract which allows the depositor to withdraw, per unit of investment, either the amount $x_1$ at date 1 or the amount $y_2$ at date 2. The banker operates subject to a moral hazard problem. At date 2, after the realisation of the investment return but before the depositor's repayment, he can abscond with the funds beyond the reach of the law. Absconding allows the banker not to repay loans but reduces the realised value of the investment $R = \{H, L\}$ by a proportion $A \in [0,1]$. The loss from absconding can be interpreted as the cost of engaging in fraud. The banker chooses the action (absconding or not) that maximises his profits. Thus, he prefers to abscond when the tax on absconding, $AR_i$, is lower than the promised repayment to the depositor at date 2, $y_2$. The banker's action depends on depositors'

reservation level, $S$, and on the tax on absconding, $AR_i$. For example, under the assumption $S > AL$, the banker absconds if $R = L$ as he has to repay the depositor at least $S$.

Some other assumptions are imposed. First, the investment is socially desirable (even considering the loss from absconding if $R = L$), that is $pH + (1-p)(1-A)L > S$. Second, liquidation is less wasteful than absconding, that is $\eta < A$. Third, the maximum amount $x_1$ that the bank can feasibly pay to the depositor in the case of liquidation is greater than what it can repay in the case of non liquidation, that is $AH > x_1 > AL$.

6) *DEPOSITORS' REPAYMENT:* Depositors withdrawing at date 1 are repaid according to the sequential service constraint rule. This is not relevant, however, in the case of a single depositor.

7) *TIMING:* Figure 5 summarises the timing of the model.

| T=0 | | T=1 | | T=2 |
|---|---|---|---|---|
| \| | \| | \| | \| | \| |
| deposit contract determined; an agent deposits his funds; the bank invests | \| the depositor can acquire the signal | the depositor makes his withdrawal decision; the bank may be liquidated | \| return $R$ realised | depositor repaid unless the banker absconds |

Figure 5: Timing

### Equilibrium

The equilibrium consists of the deposit contract at date 0. The banker offers the profit-maximising contract among those which yield the depositor at least $S$ in expectations. A contract is a function from a space of announcements $\sum$ into outcomes. An outcome is a pair $(y_2, \Omega)$, where $y_2$ is the repayment promised to the depositor at date 2 and $\Omega \in \{0,1\}$ is an indicator variable equaling 1 if the bank is liquidated and 0 otherwise.[15] The contract is called

[15]The contract should also specifies the banker's response to the depositor's announcement $\hat{s}$ and to the

"simple" if it specifies only one outcome; it is called "compound" if it specifies two outcomes. The simple contracts are the liquidating and the nonliquidating contracts. A compound contract consists of the quartet $(y_b, \Omega_b, y_g, \Omega_g)$. This is because in the single depositor case there can be only two outcomes as the signal can take only two values.

Each contract originates a sequential game in which the depositor chooses whether to invest in the signal and the announcement as a function of the signal received. The banker chooses his action (abscond or stay, repaying the depositor) as a function of the depositor's announcement and the realised return $R$. An optimal contract is one for which there exists a sequential equilibrium that generates maximum profits for the bank and gives the depositor an expected return equal to $S$. The optimal contract can be either a simple or compound, depending on the value of $S$. If this is relatively high, then the optimal contract takes on the form of a compound contract that resembles demandable debt. The depositor acquires the signal and takes his decision conditional on it. If $s = b$, the depositor liquidates the bank; if $s = g$ he lets it operate. This arrangement is optimal since it prevents the banker from absconding when $s = b$ (the banker would abscond if not liquidated as $AL < y_{2,b}$, where $y_{2,b}$ is the amount effectively repaid to the depositor in the bad state) and avoids costly liquidation when $s = g$ (the banker does not have incentives to abscond in this case as $y_{2,g} < AH$, where $y_{2,g}$ is the amount effectively repaid to the depositor in the good state).

### Conclusions

Calomiris and Kahn show that demandable debt can be part of an incentive-compatible intermediation in a context of asymmetric information and potential cheating on the part of the banker. Under some circumstances, the depositor acquires interim information and uses it to decide whether to liquidate the bank at date 1 or not. If the signal is good, the depositor lets the bank operate and the banker does not abscond; if the signal is bad, the depositor liquidates the bank and prevents the banker from absconding. Thus, demandable debt turns out to be the optimal arrangement.

The results obtained in the single depositor case can be generalised to the more complex

realised return. However, this can be easily derived: The banker absconds if $y_{2,\tilde{s}} > gAR_i$ and does not abscond if $y_{2,\tilde{s}} < AR_i$.

case of multiple depositors acquiring independent and identically distributed signals on the future value of the bank's investment. A run would then occur when enough depositors have received a bad signal. The sequential service constraint rule arises endogenously in the case of many depositors to discourage the free riding problem among them.

In focusing on the agency relation between the banker and depositors, the analysis abstracts from the role of banks as providers of liquidity to depositors. Further, as in Jacklin and Bhattacharya (1988), it considers only the withdrawal decisions of informed depositors without looking at the effects of the arrival of interim information on uninformed depositors' decisions. The result is that runs occurs when informed depositors only decide to liquidate the bank. Therefore, run efficiency depends only on the informativeness of the signal. If the signal is perfect, depositors always choose the correct action and runs are socially efficient. They solve the banker's moral hazard problem without entailing any unnecessary costly liquidation. In terms of policy implications, any extra market intervention in this framework is undesirable since it prevents the occurrence of efficient runs and its discipline effects on the banker.

### 2.4.2 Bank Runs as Imperfect Discipline Device

Chen's (1999) paper investigates the efficiency of market discipline further. As in Calomiris and Kahn, demandable debt constitutes a discipline device for moral hazard problems in banking. The threat of runs induces the banker not to speculate with depositors' money and not to invest in inefficient short-term projects.

However, market discipline can be an inefficient discipline device. Runs at some banks may cause panic and generate wasteful runs at other banks. At one bank, some depositors receive perfect interim information about the value of the bank's investment. This gives them an advantage as they can withdraw earlier than uninformed depositors when the bank's prospects are poor and avoid to be rationed. Given this, uninformed depositors have an incentive to respond to any information arriving before the value of the bank's assets is revealed, such as the failure of other banks. When uninformed depositors observe a large number of other banks failing, they infer that the prospects for the economy are poor and respond to it. They withdraw their deposits from their bank, thus forcing informed depositors to run. Runs triggered by bank-specific information (information-based runs) are efficient as they occur only if informed depositors receive negative interim information. Runs triggered by general-economic information (panic runs) entail higher social costs, as they are based on noisy information.

### Assumptions

1) T = 0,1,2 and a single good.

2) *TECHNOLOGY:* There are many different long-term investments available; each of them requires an outlay of one unit of capital at date 0 and yields a stochastic return $\widetilde{R} = \{H, L\}$ at date 2, with $H > 1 > L \geq 0$. The probability of $R = H$ is $p$ and that of $R = L$ is $1 - p$.

The variable $p$ is a random variable that depends on the prospects of the economy: $p = p_g$ when the prospects are favorable and $p = p_b$ when the prospects are poor, with $p_g > p_b$. All the investments have the same expected return but, given the realisation of $p$, the returns of two different projects are independent. Let $\eta_0$ be the prior probability of favorable prospects of the economy. Then, the expected value of $p$ at date 0, denoted as $p_0$, is given by:

$$p_0 = \eta_0 p_g + (1 - \eta_0) p_b.$$

If liquidated at date 1, the investment yields a return equal to 1. Table 6 describes the returns from each investment available.

| TECHNOLOGY | $T = 0$ | $T = 1$ | $T = 2$ |
|---|---|---|---|
| Long-term | $-1$ | $\ell = 1$ | $\widetilde{R} = \begin{cases} H & p \\ L & 1-p \end{cases}$ |

Tabella 6: Technology

3) *DEPOSITORS:* There are many depositors, each of them endowed of one dollar at date 0. A fraction of them equal to $t$ becomes of type 1 at date 1; the remaining fraction $1 - t$ becomes of type 2. The variable $t$ is deterministic. Depositors are risk averse, with $RRA > 1$ and corner preferences:

$$U^1(c_{11}) = u(c_{11}) \quad \text{for type 1 agents}$$
$$U^1(c_{22}) = u(c_{22}) \quad \text{for type 2 agents}$$

where $c_{ij}$ is the consumption at date $i$ of an agent of type $j$.

4) *INFORMATION:* At date 1 a fraction $\alpha$ of depositors at each bank, defined as informed, receives a perfectly informative signal $s = \{H, L\}$ on the future return of the bank's investment. The remaining fraction $1-\alpha$ remains uninformed. Unlike the previous models, depositors know at date 0 whether they will be informed at date 1 and both informed and uninformed depositors have the same chance of becoming of type 1 or of type 2.

5) *INTERMEDIARY:* There are $N$ competitive banks in the economy. Each of them is owned and managed by a risk-neutral banker. Banks collect deposits at date 0 in exchange for a contract which allows depositors to withdraw for each unit deposited either the amount $x_1$ at date 1 or the amount $y_2$ at date 2. Then, each banker invests in one of the investments available. There is a moral hazard problem between bankers and depositors. At date 1, after receiving perfect information on the return of the investment at date 2, the banker can decide to liquidate it and invest in a negative NPV short-term project. The new investment yields a random return $\widetilde{R}_s = \{H_s, L_s\}$ with $H_s > H$ and $L_s = 0$. The probability of $R = H_s$ is $\epsilon$ with $\epsilon H_s < 1$.

6) *DEPOSITORS' REPAYMENT:* Depositors' withdrawal demands at date 1 are repaid according to the sequential service constraint rule.

7) *INFORMATION TRANSMISSION AMONG BANKS*: Agents can deposit their funds at one bank only. Banks do not have any interrelations, that is they are not linked through the payment systems and/or interbank markets.[16] However, as depositors' withdrawals at one bank can be observed by depositors at the same bank and at other banks, there can be informational externalities. These work as follows. The timing of the revelation of the liquidity shock and of the interim information at date 1 differs among banks. Revelation occurs first at $N_1$ banks, randomly chosen among the existing $N$, and later at the remaining $N - N_1$. Depositors know whether their bank is in the first $N_1$ banks before date 1.

The revelation of bad interim information at a bank ("bank-specific information") may cause its failure. Assume $K_1 \leq N_1$ the number of failed banks among the first $N_1$ banks. A large $K_1$ implies that the prospects in the economy are poor. The failure of one bank is publicly

[16]The lack of interrelations among banks implies that if there is a propagation of failures from a group of banks to another, this takes place through the information channel. Using the terminology introduced in chapter 1, the propagation of failures in this model is an industry specific contagious run.

observable. Thus, depositors at the $N - N_1$ remaining banks observe $K_1$ ("general-economic information") before the revelation of their liquidity shock and of the interim information specific to their bank, and may respond to it.

8) *TIMING:* Figure 6 summarises the timing of the model.

| T=0 | T=1 | | | | T=2 |
|---|---|---|---|---|---|
| agents deposit their funds at one bank; banks invest | preference shock realised, information revealed at $N_1$ banks; depositors make withdrawal decisions; $K_1$ banks fail | depositors at remaining $N - N_1$ banks observe $K_1$ and make withdrawal decisions; $N - N_1$ banks may fail | preference shock realised, information revealed at the $N - N_1$ banks; depositors make withdrawal decisions | the banker may decide to speculate if his bank has not failed | return $R$ realised; remaining depositors repaid |

Figure 6: Timing

### Equilibrium

The equilibrium consists of the deposit contract at date 0, and depositors' withdrawal decisions at date 1. The model is solved backwards. Given the amounts $x_1$ and $y_2$ promised in the deposit contract, depositors' withdrawal decisions are determined; then, given depositors' decisions, the contract is determined.

a) *DEPOSITORS' WITHDRAWAL DECISIONS:* The decision of depositors at each of the first $N_1$ banks is as follows. Type 1 depositors withdraw at date 1. Informed depositors withdraw at date 1 if they receive $s = L$ and wait otherwise. Uninformed depositors observe the queue at the bank and join it if this is longer than $t$. If all depositors demand $x_1$ after the arrival of bank-specific information, then an *information-induced* run takes place: The bank liquidates all the assets and closes down. Informed depositors obtain the promised repayment $x_1$ while uninformed depositors are rationed.

$K_1$ banks fail among the first $N_1$ banks. Depositors at the other $N - N_1$ banks observe $K_1$ and use it to update the probability of favorable prospects of the economy (now equal to $\eta_1(K_1)$) and the probability of success of their bank's investment (now equal to $\widehat{p}$). The higher $K_1$ the lower $\eta_1$ and $\widehat{p}$. The general-economic information has a different value for the different categories of depositors. Informed depositors do not have any incentive to react to it as they are fully repaid if a run occurs at their bank after the arrival of bank-specific information. Uninformed depositors have incentive to react to $K_1$ as they are rationed if a run triggered by bank-specific information occurs at their bank. Thus, they compare the updated expected utility from waiting with that from withdrawing and choose to withdraw whenever convenient. This happens when the posterior probability assessment $\widehat{p}$ is low enough (that is when $K_1$ is high enough). Uninformed depositors' decision to withdraw their money forces informed depositors to withdraw prematurely as well in order to avoid being rationed and a *panic run* takes place.

b) *THE DEPOSIT CONTRACT*: The determination of the optimal contract is quite difficult as it has to take into account the probability of runs, depositors' incentive compatibility and participation constraints and the banker's moral hazard. The optimal deposit contract has the following characteristics:

$$y_2 = c^*_{22};\ x_1 = c^*_{11} \text{ with } c^*_{22} > c^*_{11} > 1.$$

The condition $c^*_{22} > c^*_{11}$ is necessary for individuals to make deposit at date 0, the condition $c^*_{11} > 1$, together with the assumptions of liquidation value $\ell = 1$ and of sequential service constraint rule, implies a payoff externality among depositors and makes the bank run out of funds at date 1 if all depositors withdraw at that date (this is analogous as in Diamond and Dybvig). The repayments promised in the contract make it convenient for the banker to speculate on the project at date 1 only in the bad state when, however, the bank will be liquidated. Thus, as in Calomiris and Kahn, the banker will never misbehave in equilibrium.

The optimal contract allows for the possibility of panic runs at date 1. A panic run could be eliminated by decreasing $x_1$ or increasing $y_2$ appropriately so to prevent uninformed depositors from reacting to the general-economic information. Decreasing $x_1$, however, reduces the insurance service of demandable debt. A lower repayment at date 1 makes depositors of type 1 worse off. Thus, the optimal contract entails a positive probability of panic runs in circumstances in which the major concern is risk sharing rather then the prevention of panics.

### Conclusions

Chen's analysis shows that market discipline is an inefficient control device for banks' moral hazard problems. Runs may be contagious, that is runs at some banks may trigger runs on others. This is due to both the sequential service constraint and the informational externalities among depositors, which stem from the different timing of information and preference shock revelation among banks. The sequential service constraint generates a payoff externality among depositors, forcing them to respond to noisy general-economic information, such as failures of other banks.

Chen's result on the inefficiency of market discipline has important policy implications. Unlike in Calomiris and Kahn (1991), now public interventions are desirable. Chen shows that there is a deposit insurance system that can induce depositors to respond to bank-specific information, thus eliminating contagious runs. This system fully protects uninformed depositors, whilst leaving informed depositors uninsured. The fully protected uninformed depositors do not need to withdraw early in response to bad general-economic information. Because uninformed depositors never start a run, the uninsured informed depositors at a bank can always wait until bank-specific information is revealed and withdraw only when this is bad. Thus, market discipline works efficiently since contagious runs are prevented.

Chen's model shares some common features with the other models presented above. As in Diamond and Dybvig(1983), agents are very risk averse (RRA>1) and the liquidation value of the long-term investment equals 1. This implies that the bank offers depositors insurance against the shock of being of type 1, as it promises them an amount $x_1 > 1$ at date 1, but it also implies a payoff externality that causes inefficient runs. As in Diamond and Dybvig, the deposit contract supports multiple equilibria, as depositors run if they fear that other depositors will start to withdraw early, independently of any information. However, the panic equilibrium in Chen does not rely on the existence of multiple equilibria. By introducing a stochastic technology and informational externalities among banks, Chen shows that a panic run can occur even when depositors choose the Pareto-dominant equilibrium (the one described above in the analysis) when there are multiple equilibria. The negative payoff externality generated is now used to analyse how depositors respond to the arrival of early and noisy information rather than to analyse how it generates multiple equilibria.

As in Chari and Jagannathan (1988), runs occur when all depositors withdraw prematurely and depositors are asymmetrically informed on the future value of the bank's assets. This is the source of run inefficiency in both models. However, the mechanism generating it differs in the two models. In Chari and Jagannathan, the inefficiency is within one bank and there is no disciplinary effect on the asset side. Uninformed depositors use the public information available (the queue at their bank) to update their ex post probability of success of the bank's investment and they withdraw when this is low enough. A panic run occurs when uninformed depositors commit a mistake in their updating process as they confuse a large realisation of the liquidity shock with an insolvency problem. In Chen, uninformed depositors at one bank make perfect inference of the signal received by the informed depositors by looking at the withdrawal queue. So, a run triggered by bank-specific information is always efficient as it occurs only if the bank will turn to be insolvent at date 2. However, because of their informational disadvantage, uninformed depositors are rationed in the case of a run triggered by bank-specific information. This makes them sensitive to the arrival of some external information and a panic run occurs in the model when they respond to it instead of waiting for more precise bank-specific information.

In the next chapter, I am studying further the character of demandable debt as imperfect discipline device, focusing on internal inefficiency instead of contagion phenomena.

# Chapter 3

# Bank Moral Hazard and Market Discipline

## 3.1 Introduction

Bank stability has always been a major concern for policymakers and a major topic of debate among economists. The issue is not purely academic, as many and recurrent bank runs and failures demonstrate. There is a stark contrast between the variety of opinions expressed in the debate and the uniformity of policy stances. Since the devastating crises which undermined the stability of the US banking system in the 1930s, policymakers in industrialised countries have taken up a supervisory role and introduced deposit insurance. Also, they have often chosen to offer a *de facto* complete insurance to the banking system by forbearing financially distressed banks and bailing out insolvent ones.

This attitude, which has prevailed for the last half century, has been shaken by the many bank failures of the 1980s and 1990s. Systemic crises occurred both in developed and developing countries, such as the United States, Finland, Japan, Chile, Argentina, Mexico, Indonesia, and Russia. Regulators in the United Kingdom, Italy, Germany and France intervened to rescue individual banks in distress.[1] According to many economists, the underlying cause of these crises was a problem of moral hazard. The *de facto* protection provided by regulators induced

[1] See Caprio and Klingebiel (1996).

excessive risk-taking on the part of banks.

These crises have renewed the debate on how to regulate the banking system. Several proposals have been put forward to reduce the distortions induced by the protective attitude of regulators. Most suggestions revolve around the ideas of strengthening supervision or increasing the price for bank protection by tightening capital adequacy ratios and deposit insurance premia.[2] Apart from problems with the internal consistency of these approaches, their implementation may be too difficult. For example, risk-sensitive prudential regulation requires an amount of information often not available to regulators. Moreover, regulators may fail–for whatever reason–to use their supervisory tools to avert banks' distress and to step in once a failure has occurred, as the Saving and Loans debacle has shown.

For these reasons, some researchers and policymakers have stressed the potential of market forces to overcome banks' moral hazard problem. The idea is that market participants have stronger incentives than regulators to acquire information on banks' risk exposure and to make use of the information they do have. Clearly, this requires that depositors or other creditors of banks do not anticipate being bailed out if their bank is in troubles. Debtholders can exert discipline on banks through the pricing and the growth of their claims. Two recommendations, usually seen as alternative, have been advocated: (*i*) introduce an uninsured subordinated debt requirement on banks;[3] (*ii*) scale back deposit insurance and reintroduce uninsured depositors' monitoring in the form of deposit withdrawals on demand.[4] Proposal (*i*) relies on the idea that subordinated debt yields should be formally used as an early warning signal in the regulatory process of identifying banks with a high likelihood of distress. High yields would indicate high risk exposure and should trigger regulatory discipline on the bank. Proposal (*ii*) suggests instead that market intervention should substitute regulators in the closure decision of banks with a high likelihood of insolvency. While regulators are often prone to forbearance, uninsured depositors have strong incentives to run as soon as they have doubts on banks' solvency. Hence, the credible threat of depositors' withdrawals would restrain banks from taking on high risks.

The effectiveness of market discipline depends crucially on how good the information available to market participants is. Concerning proposal (*ii*), for example, if depositors have im-

[2] An overview of these proposals has been presented in chapter 1.
[3] See, for example, Benston et al. (1986) and, more recently, Calomiris (1998).
[4] See, for example, Broaddus (1994), Federal Reserve of Chicago (1990) and Kaufman (1994).

perfect information on the future value of bank assets, they may commit mistakes in their withdrawal decisions: They can either precipitate a run on a solvent bank or let an insolvent bank operate. If this is the case, then market discipline can entail high social costs and needs to be accompanied by regulatory measures.

In this chapter, I analyse from a theoretical point of view the effectiveness of the proposal of restoring market discipline, focusing on how depositors assess their bank's financial conditions and react to the (asymmetrically distributed) information they have on the value of the bank's assets.[5] To address this issue, I develop a model in which a banker needs to raise external funds to invest in an illiquid and risky project. The banker operates under moral hazard since he can privately decide whether to monitor the project, thus increasing its probability of success, or not. To raise funds, he can issue either standard debt or demandable debt. With standard debt, investors deposit their funds at the bank and demand a predetermined repayment at maturity. With demandable debt, investors are allowed to withdraw at any point in time. Starting from a situation in which credit markets do not work with standard debt, I show that demandable debt can constitute a solution for the banker's moral hazard problem and induce investors to deposit their funds at the bank. This is because demandable debt allows depositors to discipline the banker by observing the value of the bank's assets and triggering a run whenever they (rationally) expect this value to be low.

However, market discipline can come at a cost. When not all depositors are equally informed on the value of the bank's assets, it may happen that those who are uninformed commit errors in deciding when to run on the bank. They may either leave their deposits at the bank when prospects are in fact low, or they may withdraw their funds when prospects are in fact high. In the former case, an insolvent bank is let operate although it would be efficient to liquidate its investments and close it down. In the latter case, the erroneous withdrawal decisions of uninformed depositors can force informed depositors to disregard their information and withdraw prematurely. Even if they know that the bank will be solvent, informed depositors find it optimal to precipitate a run since they would get nothing from leaving their funds at the bank. When this is the case, an ex post inefficient run occurs and a solvent bank is forced into liquidation.

[5] For a theoretical analysis of the effectiveness of proposal (i), see, for example, Nagarajan and Sealey (1997).

Therefore, market discipline is a feasible incentive device, but it can be imperfect and costly. The benefits of demandable debt consist of its incentive effect. The costs stem from the fact that, with demandable debt, profitable projects may be liquidated and unprofitable projects continued. These results provide a rationale for regulatory intervention, which should aim to eliminate the costs of market discipline while, at the same time, preserving its benefits.

The model I develop in this chapter relates to the literature reviewed in chapter 2. Diamond and Dybvig (1983) show that runs are an unfortunate and undesirable side-effect of demandable debt and they occur as a sunspot if depositors lose confidence in the bank's solvency. The coordination problem among depositors results from the sequential service constraint imposed on depositors' repayments. I do not need to impose such a constraint. Bank runs in my model are information-induced and the coordination problem among depositors arises from both the informational asymmetries among depositors and the costly liquidation of the bank's assets.

The mechanism through which *bank runs* occur in my model is similar, though not identical, to that in Chari and Jagannathan (1988). Both models focus on a signal-extraction problem to analyse how information is revealed to uninformed depositors by the withdrawal decisions of other depositors. However, the two models differ in terms of the bank run inefficiency. Chari and Jagannathan argue that a panic run occurs when depositors fear that some of them possess (superior) negative information on the bank's project returns. My model shows that an inefficient run can take place despite the fact that some depositors do have positive information on the bank's returns. This is because, in addition to the signal-extraction problem, I also consider how the behaviour of uninformed depositors feeds back to the withdrawal decisions of informed depositors. Furthermore, unlike Chari and Jagannathan, I examine the incentive effects of demandable debt by considering an agency problem between the bank and its depositors.

This latter aspect alone has been analysed earlier by Calomiris and Kahn (1991). They study the incentive effects of demandable debt in a context of asymmetric information between a bank and its depositors.[6] By focusing on the endogenous acquisition of costly information by depositors, they show that the threat of bank liquidation restrains the banker from fraud-

[6] Also Qi (1998) and Diamond and Rajan (1999) examine the disciplinary effect of liquid deposits but in a rather different framework. They build general equilibrium models where the role of the bank is either to issue liquid deposits against information-sensitive loan funding (Qi ) or to solve liquidity problems of both depositors and borrowers (Diamond and Rajan). Both models abstract from issues of asymmetric information and of liquidation costs of premature withdrawals.

ulently diverting resources ex post. Within this framework, market discipline turns out to be an efficient mechanism: The sequential service constraint induces some depositors to monitor the bank and to withdraw their deposits in case they receive negative information. Hence, Calomiris and Kahn always regard bank runs as socially beneficial, since they result from the withdrawal decisions of perfectly informed depositors. Similarly to Calomiris and Kahn, my results show that demandable debt acts as an incentive device, which mitigates the moral hazard problem stemming from the banker's discretionary choice of whether to monitor projects or not. However, unlike Calomiris and Kahn, I consider a context where bank runs arise from the co-existence of liquidity, informed and uninformed depositors. This allows me to explicitly model both the costs *and* the benefits of the incentive effects of demandable debt.

Chen (1999) also discusses whether depositors' monitoring is an efficient device for controlling bankers' moral hazard problems, but he focuses on an inefficiency different from the one studied here. He shows that market discipline is costly because bank runs are contagious, that is failures of a few banks may cause runs on others. In his model, a panic run occurs when depositors withdraw their funds from their bank in response to failures of other banks: The payoff externality generated by the sequential service constraint induces uninformed depositors to respond to early information and forces informed depositors to precipitate a run before more precise information on their bank is revealed. In my model, the inefficiency of market discipline originates inside the bank: Both inefficient runs and continuations can occur even if some depositors are *already* informed about the value of the bank's assets. This is because depositors make mistakes in their information updating process, which takes place after specific information on their bank is revealed.

The rest of the chapter is organized as follows. The model is described in section 2. The incentive effects of demandable debt are analysed in section 3. Section 4 concludes the chapter.

## 3.2 The Model

Consider a three-date economy (T = 0,1,2) with two types of risk neutral agents: a bank and a continuum of depositors of measure one. The bank, which acts as a monopolist, can invest in a risky and illiquid project. Since it has no capital, the bank needs to raise funds from depositors. It can offer either a standard or a demandable deposit contract for different maturities, as specified below. Both deposit contracts are noncontingent, in the sense that the amount that depositors can withdraw at each date is not contingent on the returns of the bank's project but it depends only on the timing of withdrawal. Depositors are perfectly competitive and each is endowed with one unit. The riskless interest rate is normalised to zero.

*The bank and the moral hazard problem*

The project the bank invests in requires an outlay of one unit of capital at date 0. At date 2, it yields $x = H$ per unit invested if it succeeds and $x = 0$ if it fails. The probability of success of the project depends upon the behaviour of the banker, who is the manager and also the only owner of the bank: He can simply invest in the project or he can actively monitor it, increasing the probability of success. I denote the probability of success by $p_b$ in the former case, and by $p_g$ in the latter, with $p_g > p_b$. The subscripts $b$ and $g$ denote 'bad' (no monitoring) and 'good' (monitoring) behaviour on the part of the banker. The project is economically viable only if the banker monitors: $p_g H > 1 > p_b H$. The banker may choose not to monitor in order to enjoy, at date 2, a non-transferable private benefit $B$. The private benefit can be interpreted as the opportunity cost of monitoring the project. Returns are observable but monitoring choice is not. This creates a moral hazard problem.

The severity of the moral hazard problem depends on the magnitude of the private benefit $B$ and on the cost of deposits, which, in turn, depends on the debt contract that the bank uses to raise funds. To provide a benchmark, I start by analyzing the situation in which the banker chooses between monitoring and not monitoring when he issues a standard debt contract. Then, I turn to the situation when the banker issues demandable debt.

*The standard debt contract*

A standard debt contract (S) is defined as a contract that requires one unit of investment at date 0 in exchange for the repayment, at date 2, of a sum $R^S$, which includes principal and interest. Repayment is conditional on the bank being solvent, i.e. the bank is protected by limited liability. The face value of debt, $R^S$, is determined at date 0 so as to guarantee depositors zero expected profits.

Since the project is economically viable only if the banker monitors, depositors leave their money with the bank only if they anticipate that the banker will indeed do so.[7] The banker monitors only if:

$$\Pi_g^S = p_g(H - R^S) \geq \Pi_b^S = p_b(H - R^S) + B$$

that is, only if:

$$(p_g - p_b)(H - R^S) \geq B \tag{1}$$

where $\Pi_g^S$ and $\Pi_b^S$ are the bank's expected profits from monitoring and not monitoring, respectively, and $R^S > 1$ is the face value of debt conditional on the banker choosing to monitor. For simplicity, I assume that the banker behaves well when he is indifferent between monitoring and not monitoring. Perfect competition among depositors implies $p_g R^S = 1$, that is the face value of debt guarantees zero expected profits to depositors if the banker chooses to monitor. Condition (1) is satisfied whenever the moral hazard problem is not "too severe," that is when $B \leq B^S$ where $B^S$ is defined as $B^S = (p_g - p_b)(H - R^S)$. In other words, $B^S$ is the value of $B$ above which the banker prefers not to monitor.

If condition (1) fails, the bank will be unable to raise funds using the standard debt contract because depositors anticipate that the banker will not monitor. Is there a way out to this problem? I now consider the situation in which the banker issues demandable debt and I show that the use of the demandable debt contract can solve the moral hazard problem by inducing the banker to monitor the project.[8]

*The demandable debt contract*

A demandable debt contract (D) is defined as a contract that requires one unit of investment at date 0 in exchange for the right to withdraw either *(i)* the initial unit of investment at date 1, or *(ii)* a sum $R^D$, which includes principal and interest, at date 2. The amount that depositors can withdraw at date 1 is determined by their risk neutrality.[9] The face value of debt, $R^D$, is

[7]Since $p_b H < 1$, no repayment $R^S$ can guarantee depositors zero expected profits if the banker does not monitor.

[8]Note that the moral hazard problem cannot be solved by the bank promising depositors a face value of debt greater than $R^S$. Such promise would indeed only worsen the bank's incentive compatibility costraint.

[9]An analysis of the insurance provided to depositors by the bank goes beyond the scope of this model, which concentrates on the incentive effects of demandable debt. By assuming risk neutrality, I avoid worry about the

determined at date 0 so as to guarantee strategic depositors (see below) zero expected profits.

Demandable debt allows depositors to choose their preferred consumption profile. There are two classes of depositors, liquidity depositors and strategic depositors. Both classes are a continuum of measure one half, so that the maximum total amount of funding available to the bank at date 0 is one unit.

Liquidity depositors simply respond to their liquidity needs. A random fraction of liquidity depositors, $\tilde{t} \in [0, 1]$, withdraws its deposits at date 1 in response to a shock to consumption preferences. The remaining liquidity depositors, $1 - \tilde{t}$, leave their funds at the bank until date 2. Denote these two fractions 'early' and 'late' liquidity depositors, respectively. Without loss of generality, I assume that $\tilde{t}$ can take three values, $0 < t_1 < t_2$, with probabilities $q_0, q_1$ and $q_2$, respectively. I also assume that the intermediate liquidity shock is more likely than the extreme ones: $q_1 > q_0, q_2$.[10]

Strategic depositors compete for returns and behave strategically in order to maximise their utility, which is given by $U(c_1, c_2) = c_1 + c_2$. At date 0, they deposit their funds at the bank in exchange for a repayment $R^D > 1$ at date 2, which has to guarantee them zero expected profits. At date 1, after the arrival of information on the future return of the bank's project (as specified below), they decide whether to leave their deposits until date 2 or to withdraw prematurely. Since a consumer's type is not observable, strategic depositors can always imitate early liquidity depositors and withdraw at date 1 whenever they find it optimal.

The distinction between liquidity and strategic depositors suggests different interpretations. Liquidity depositors can be thought of as small depositors who are interested only in their consumption needs. Strategic depositors can be interpreted instead as wholesale depositors (or also other banks) who have the incentive and the ability to monitor and discipline the banker by reacting promptly to the information available at date 1 on the bank's future solvency.

*Information*

A crucial element of the model is the information that strategic depositors have at date 1 on the future return of the bank's project. In the following, I focus on the situation where strate-

relationship between the shape of depositors' utility function and the project's liquidation value. On this issue, see Diamond and Dybvig (1983), Jacklin and Bhattacharya (1988) and also Allen and Gale (1998).

[10] A generalization to a continuous $t$ would require its density function $f(t)$ to be single peaked.

gic depositors are asymmetrically informed. This is the economically interesting information structure, which allows me to analyse both the benefits and the costs of the incentive effect of demandable debt.

At date 1, a fraction $\alpha$ of strategic depositors receives a perfectly informative signal, $s \in \{H, 0\}$, on the future value of the bank's assets. The signal is the same for all informed strategic depositors. The remaining strategic depositors, $1 - \alpha$, remain uninformed. Denote these two fractions 'informed' and 'uninformed' strategic depositors, respectively. At date 0, a strategic depositor does not know whether he will become informed or not.

Uninformed strategic depositors have an informational disadvantage: They cannot observe the realisation of either the fraction $\tilde{t}$ of early liquidity depositors or the return $\tilde{x}$ of the bank's assets, as well as the signal $s$ received by informed strategic depositors. The only variable that they can observe (and that also informed strategic depositors can observe) is the amount of aggregate withdrawals at date 1, $W$. In other words, uninformed strategic depositors can observe the total fraction of depositors who withdraw at date 1 but not the reason behind each individual withdrawal decision.

The amount of aggregate withdrawals $W$ is correlated with the signal $s$ received by informed strategic depositors and, therefore, with the future value of the bank's assets. By observing $W$, uninformed strategic depositors try to infer $s$. If $W$ is *perfectly* correlated with $s$, then their signal extraction problem is trivial and all strategic depositors are equally and perfectly informed on the future value of the bank's assets. If $W$ is *imperfectly* correlated with $s$, then uninformed strategic depositors have a non-trivial signal extraction problem and strategic depositors are asymmetrically informed.

In order for uninformed strategic depositors to have a non-trivial signal extraction problem, I assume:

$$t_1 = \alpha \tag{2}$$

$$t_2 = t_1 + \alpha = 2t_1. \tag{3}$$

Assumption (2) states that the fraction of informed strategic depositors equals the intermediate fraction of early liquidity depositors. Assumption (3) states that the largest fraction of early liquidity depositors equals the fraction of informed strategic agents plus the intermediate fraction of early liquidity depositors. With these restrictions in place, uninformed strategic depositors observe a noisy indicator of the signal received by informed strategic depositors, from which they may be unable to infer the bank's future solvency. Without these restrictions in place, uninformed strategic depositors would always be able to infer the signal in equilibrium. Hence, assumptions (2) and (3) are neither restrictions on the structure of the economy nor are they necessary for the existence of an equilibrium. Rather, they allow me to model confounding.[11]

*Depositors' withdrawal demands and bank's project liquidation*

At date 1, all depositors decide whether to withdraw prematurely or to leave their deposits at the bank until date 2. All depositors withdrawing at date 1 submit their requests simultaneously. Depending on the value of aggregate withdrawals $W$, each depositor receives either the initial unit of deposit or a pro-rata share of the value of the bank's assets. Similarly, depositors waiting until date 2 receive either the promised repayment $R^D$ or a pro-rata share of the value of the bank's assets.[12] Depositors who do not withdraw at date 1 are not taken into account in the splitting of the value of the bank's assets.

The bank pays off depositors who demand early withdrawals by liquidating the project. Liquidation is costly in the following sense. If only (relatively) few depositors ask for repayment at date 1, that is if $W$ is low, liquidation yields the initial unit of investment. If (relatively) many depositors withdraw at date 1, that is if $W$ is high, liquidation yields an amount $\ell$ which is less than the initial unit of investment. Let $LV$ represent the liquidation value of the project. Then, I assume:

$$LV = \begin{cases} 1 & \text{if } W \leq \overline{W} \\ \ell & \text{if } W > \overline{W} \end{cases} \tag{4}$$

where is $\overline{W}$ equal to:

$$\overline{W} = \frac{t_2}{2} \tag{5}$$

and $\ell < 1$ is exogenously specified below.

[11]Without assumptions (2) and (3), the model would be similar to Calomiris and Kahn (1991) and bank runs would always be triggered by perfect information on the bank's state of solvency.

[12]Since I do not consider the issue of costly acquisition of information about the project's returns of the on parts of depositors, I do not impose the sequential service constraint rule for depositors' repayment. Indeed, this rule has been criticised as being not an optimal arrangement when depositors' information acquisition problem is exogenous (see Calomiris and Kahn (1991) and Allen and Gale (1998)).

Assumption (4) describes the liquidation value of the project conditional on the aggregate withdrawals. Assumption (5) fixes the threshold level of aggregate withdrawals beyond which the liquidation value of the project reduces to $\ell$. Together these two assumptions attempt to capture, in a reduced form, the determination of the price of the bank's assets in a secondary market where, in the spirit of Akerlof (1970), the presence of asymmetric information on the bank's future solvency generates a "lemons" problem. The secondary market prices the assets according to the only observable variable, $W$, which is correlated—albeit imperfectly—with their value. When aggregate withdrawals exceed $\overline{W}$, the liquidation value falls, reflecting the market's expectation of a lower value of the assets. Indeed, when $W > \overline{W}$, the market infers that some, if not all, strategic depositors are withdrawing prematurely. This conveys negative information on the bank's future solvency and, consequently, induces the market to offer a lower price for its assets.

The liquidation value of the project determines the bank's liquidation policy and affects strategic depositors' behaviour by specifying the resources available for redemption at the bank at dates 1 and 2. In order to simplify the analysis, I assume:

$$\ell = \frac{1+t_1}{2}. \tag{6}$$

Assumption (6) is a further specification of the liquidation value $\ell$ in (4). It simplifies the analysis in that it completely isolates strategic depositors' withdrawal decisions from the value of $\ell$. Note that (6) is stronger than necessary. As I will discuss further below, the main results summarised in proposition 1 still hold for a much wider range of values for $\ell$.

*Timing and Notation*

The time structure of the model is summarised in Figure 1. At date 0, the bank issues demandable debt and investors deposit their funds. The promised repayment to depositors at date 2, $R^D$, is determined and the banker chooses whether to monitor the project or not.

At date 1, the fraction $t$ of early liquidity depositors is realised and the signal $s$ is observed by the fraction $\alpha$ of informed strategic depositors. Then, all depositors make their withdrawal decisions simultaneously. If (relatively) many depositors withdraw at date 1, the bank liquidates the entire project and is closed down. Otherwise, at date 2 the project's returns are realised and claims are settled. Conditional on the bank being solvent, each remaining depositor receives $R^D$ and the bank retains the surplus.

| T=0 | | T=1 | | T=2 |
|---|---|---|---|---|
| investors deposit their funds; the banker chooses whether to monitor the project | fraction of early liquidity depositors realised informed strategic depositors receive the signal | depositors make the withdrawal decisions; the bank may liquidate the project | project returns are realised | depositors are repaid; the bank retains the surplus |

Figure 1: Timing

Table 1 provides the list of notation that describes the model.

| | |
|---|---|
| $\tilde{x}$ | = random return of the project at date 2 ($x = H, 0$) |
| $p_g(p_b)$ | = success probability of the project when monitoring (no monitoring) |
| $B$ | = private benefit |
| $R^S$ | = face value of Standard debt (S) |
| $\Pi_g^S(\Pi_b^S)$ | = bank's expected profits when monitoring (no monitoring) with S |
| $R^D$ | = face value of Demandable debt (D) |
| $\Pi_g^D(\Pi_b^D)$ | = bank's expected profits when monitoring (no monitoring) with D |
| $W$ | = aggregate withdrawals |
| $\overline{W}$ | = aggregate withdrawals beyond which $LV = \ell$ |
| $\ell$ | = liquidation value when $W > \overline{W}$ |
| $\tilde{t}$ | = fraction of 'early' liquidity depositors ($t = 0, t_1, t_2$) |
| $q_i$ | = probability distribution of $\tilde{t}$, $i = 0, 1, 2$ |
| $\alpha$ | = fraction of informed strategic depositors |

Table 1: Notation

## 3.3 The Incentive Effects of Demandable Debt

The equilibrium consists of three elements: the face value of debt $R^D$ and the banker's monitoring choice at date 0, and depositors' withdrawal decisions at date 1. The model is solved backward. I first analyse depositors' withdrawal decisions at date 1, given $R^D$ and the banker's choice of monitoring the project. Then I compute $R^D$ and the banker's monitoring choice at date 0, taking into account depositors' withdrawal decisions at date 1.

### 3.3.1 Depositors' Withdrawal Decisions at Date 1

At date 1, the liquidity shock is realised and informed strategic depositors receive the signal. Then, all agents choose simultaneously whether to withdraw or not, according to their type, to the available information and to others' decisions. I analyse the decisions of each type of depositors in turn.

The decision problem of liquidity depositors is trivial: A fraction equal to the realisation of $\tilde{t}$ withdraws deposits at date 1. The remaining fraction $1 - t$ waits until date 2.

Informed strategic depositors make their withdrawal decisions after observing the signal $s$ and the aggregate withdrawals $W$. When $s = 0$, they find it optimal to withdraw at date 1, irrespective of $W$, since they know they would get nothing from waiting until date 2. When they observe a large enough $W$, they choose to withdraw at date 1, irrespective of $s$. Indeed, when enough depositors withdraw at date 1, informed strategic depositors know they will get less at date 2 if they do not withdraw. This is because the bank has to liquidate assets to satisfy withdrawals at date 1 and it may not have enough resources for redemption at date 2. Let $w^I(s, W)$ be the solution to informed strategic depositors' decision problem.

Uninformed strategic depositors make their withdrawal decisions after observing $W$. They realise that $W$ is correlated with $s$, although imperfectly. Indeed, $W$ could be high either because the realisation of the fraction of early liquidity depositors $t$ is high or because informed agents have received a signal $s = 0$ on the future value of the bank's assets. This confounding is crucial for the results. Conditional on the observed value of $W$, uninformed strategic depositors compute the expected utility from waiting until date 2 and choose to withdraw if this expected utility is greater than the utility from withdrawing early. Let $w^U(W)$ be the solution to uninformed strategic depositors' decision problem, which is given by:

$$\max_{w^U(W)} c_1 + \int c_2 dF(\theta|W)$$

subject to:

$$c_1 = \max w^U(W) \left\{1, \frac{\ell}{W}\right\}$$

$$c_2 = \max\left\{R^D(1 - w^U(W)), 0\right\}$$

where $\theta \equiv (\tilde{t}, \tilde{x})$ is the state of the world, which is described by the two independent random variables $\tilde{t}$ and $\tilde{x}$, and $F(\theta|W)$ denotes the distribution of $\theta$ conditional on $W$.

The aggregate demand for withdrawals , $W_D$, is then given by the sum of depositors' individual withdrawal decisions, that is by:

$$W_D = \frac{1}{2}[t + \alpha w^I(s, W) + (1 - \alpha)w^U(W)]$$

where the terms on the right-hand side are the aggregate withdrawal decisions of early liquidity depositors, informed strategic depositors and uninformed strategic depositors, respectively. In equilibrium $W_D = W$. Table 2 describes the state of the world $\theta$ and its probability distribution and it shows the value of $W_D$ for every $\theta$.

| State | $\theta = (t, x)$ | Probability | Aggregate demand for withdrawals $W_D$ |
|---|---|---|---|
| 1 | $0, H$ | $q_0 p_g$ | $\frac{1}{2}[\alpha w^I(H, W) + (1 - \alpha)w^U(W)]$ |
| 2 | $0, 0$ | $q_0(1 - p_g)$ | $\frac{1}{2}[\alpha w^I(0, W) + (1 - \alpha)w^U(W)]$ |
| 3 | $t_1, H$ | $q_1 p_g$ | $\frac{1}{2}[t_1 + \alpha w^I(H, W) + (1 - \alpha)w^U(W)]$ |
| 4 | $t_1, 0$ | $q_1(1 - p_g)$ | $\frac{1}{2}[t_1 + \alpha w^I(0, W) + (1 - \alpha)w^U(W)]$ |
| 5 | $t_2, H$ | $q_2 p_g$ | $\frac{1}{2}[t_2 + \alpha w^I(H, W) + (1 - \alpha)w^U(W)]$ |
| 6 | $t_2, 0$ | $q_2(1 - p_g)$ | $\frac{1}{2}[t_2 + \alpha w^I(0, W) + (1 - \alpha)w^U(W)]$ |

Table 2: Aggregate demand for withdrawals

The equilibrium concept I use for depositors' withdrawal decisions is that of rational expectations. This requires that strategic agents make their optimal decisions conditional on their information, that is the volume of aggregate withdrawals, and any signal they observe.

Formally, a rational expectations equilibrium consists of:

(i) an aggregate withdrawal function $W(\theta)$ that specifies aggregate withdrawals for each state of the world $\theta$, and

(ii) withdrawal demands $w^I(s, W(\theta))$ and $w^U(W(\theta))$ for informed and uninformed strategic depositors, respectively, such that:

(a) $W(\theta) = \frac{1}{2}[t + \alpha w^I(s, W(\theta)) + (1-\alpha)w^U(W(\theta))]$, for all $\theta$;

(b) $w^I(s, W(\theta))$ and $w^U(W(\theta))$ are the optimal solutions to informed and uninformed strategic depositors' decision problems, respectively;

(c) if $W_D(\theta) = W_D(\theta')$, that is if $\frac{1}{2}[t+\alpha w^I(s, W(\theta))+(1-\alpha)w^U(W(\theta))] = \frac{1}{2}[t'+\alpha w^I(s, W(\theta'))+(1-\alpha)w^U(W(\theta'))]$, for any two states $\theta = (t, x)$ and $\theta' = (t', x')$, then $W(\theta) = W(\theta')$.

A rational expectations equilibrium is then characterized by the vector of aggregate withdrawals in each state of the world, $W = (W_1, W_2, W_3, W_4, W_5, W_6)$, which satisfies conditions (a), (b) and (c). Condition (a) is a market-clearing condition, which requires that aggregate withdrawals equal the sum of individual withdrawals. Condition (b) requires that depositors behave optimally. Condition (c) requires that if the aggregate demand for withdrawals is the same for two states of the world, then the equilibrium outcome should also be the same. Proposition 1 characterises the equilibrium. The formal proof of the proposition is in the appendix while the intuition is discussed below.

**Proposition 1** *Under assumptions (2) and (3), there exists a rational expectations equilibrium* $W = (0, \frac{\alpha}{2}, \frac{t_1}{2}, \frac{1+t_1}{2}, \frac{1+t_2}{2}, \frac{1+t_2}{2})$, *provided that:*

$$\frac{p_g q_1 R^D}{p_g q_1 + (1-p_g)q_0} > 1 \tag{7}$$

$$p_g R^D < \frac{1+t_1}{1+t_2}. \tag{8}$$

Proposition 1 describes the aggregate withdrawals in each state of the world and it implies that a run takes place in states 4, 5 and 6. Condition (7) implies that uninformed strategic depositors do not withdraw in states 2 and 3 since they expect to receive more by leaving their funds than by withdrawing. Condition (8), instead, implies that uninformed strategic depositors withdraw in states 4, 5 and 6 since they expect to receive more by withdrawing than by leaving their funds. Proposition 1 also states that informed strategic depositors withdraw in states 2, 4, 5 and 6. The complete characterisation of the equilibrium withdrawals for each state of the world is summarised in table 3. In particular, the table describes the solutions $w^I(s, W)$ and $w^U(W)$ to informed and uninformed strategic depositors' respective decision problems (indicating with 1 the decision to withdraw at date 1 and with 0 that of waiting until date 2), the aggregate withdrawals in each state of the world and the occurrence of runs.

| State | $\theta = (t, x)$ | Probability | $w^I(s, W)$ | $w^U(W)$ | $W$ | Run |
|---|---|---|---|---|---|---|
| 1 | $0, H$ | $q_0 p_g$ | 0 | 0 | 0 | no |
| 2 | $0, 0$ | $q_0(1 - p_g)$ | 1 | 0 | $\frac{\alpha}{2}$ | no |
| 3 | $t_1, H$ | $q_1 p_g$ | 0 | 0 | $\frac{t_1}{2}$ | no |
| 4 | $t_1, 0$ | $q_1(1 - p_g)$ | 1 | 1 | $\frac{1+t_1}{2}$ | yes |
| 5 | $t_2, H$ | $q_2 p_g$ | 1 | 1 | $\frac{1+t_2}{2}$ | yes |
| 6 | $t_2, 0$ | $q_2(1 - p_g)$ | 1 | 1 | $\frac{1+t_2}{2}$ | yes |

Table 3: Summary of equilibrium situations in proposition 1

Runs occurring in states 4, 5 and 6 are *information-induced* since they stem from strategic depositors' rational response to the information they have at date 1. Runs occurring in states 4 and 6 are efficient since they induce the liquidation of assets that would yield nothing if continued until date 2. The run occurring in state 5 is however inefficient since it forces the liquidation of valuable assets. Likewise, continuation in states 1 and 3 is efficient since it allows valuable assets to mature and to yield a high return. Continuation in state 2 is inefficient since the project should be optimally liquidated at date 1.

Where do these inefficiencies come from? Both stem from the fact that uninformed strategic depositors' inference problem is noisy: Conditional on the observation of the aggregate withdrawals, uninformed strategic depositors are not always able to infer the state of the world and, therefore, whether the bank is facing an insolvency or just a liquidity shock. Given this confounding, uninformed strategic depositors may make erroneous withdrawal decisions. They may withdraw when the project's returns are high or leave their deposits until date 2 when the project's returns are low. Uninformed strategic depositors realise that the magnitude of aggregate withdrawals is positively correlated with the probability of the bank being insolvent. Therefore, they withdraw erroneously when aggregate withdrawals are large (state 5) and wait mistakenly when aggregate withdrawals are low (state 2).[13]

The erroneous decision of uninformed strategic depositors in state 5 forces informed strategic depositors to precipitate a run even if they know that the bank will be solvent. They find it optimal to join the queue and share the bank's resources at date 1 since they would get nothing from waiting until date 2. Indeed, given the liquidation value equal to $\ell$, the bank has to liquidate the entire project prematurely to satisfy the withdrawal demands of early liquidity depositors and of uninformed strategic depositors. So, in the equilibrium described in proposition 1, informed strategic depositors do not withdraw only when they receive a negative signal (states 2, 4 and 6) but also when the aggregate withdrawals are large enough, irrespective of the signal they observe (state 5).

Both the two inefficiencies in strategic depositors' withdrawal decisions (inefficient continuation and liquidation) stem from 'rational' coordination problems among depositors, but the coordination failure is different in each case. The inefficient continuation in state 2 is due to the inability of uninformed strategic depositors to realise that the queue consists only of informed strategic depositors. The inefficient run in state 5, instead, is due to the inability of uninformed strategic depositors to realise that the queue consists only of early liquidity depositors *and* to the feedback effect that uninformed strategic depositors' mistakes have on informed strategic depositors' decisions. This is different from both the coordination failure in Diamond and Dybvig (1983) and the panic equilibrium in Chari and Jagannathan (1988) and in Chen (1999). The coordination problem occurring in strategic depositors' withdrawal decisions is not triggered by sunspots, as in Diamond and Dybvig (1983), but it is caused by both informational asymmetries among depositors and the costly liquidation of the bank's project. Unlike Chari and Jagannathan (1988), in which panics occur only when all depositors remain uninformed at date 1, the inefficient run takes place despite the fact that some depositors are aware of the future solvency of the bank. Informed strategic depositors know that the bank's project is valuable but they are forced to withdraw at date 1 by uninformed strategic depositors' erroneous decision. This 'feedback' inefficiency is different from that in Chen (1999), where panics take place when uninformed depositors' response to early information forces informed depositors to

[13]Since $q_1 > q_0$, $q_2$, when uninformed strategic depositors observe high aggregate withdrawals they expect the bank being insolvent with a higher probability than when they observe low aggregate withdrawals.

run on the bank without waiting for more precise 'bank-specific' information.

The equilibrium behaviour of strategic depositors summarised in proposition 1 has been derived under assumption (6) for the liquidation value $\ell$. However, from calculations available from the author, it can be shown that it still holds for $\ell \in [\underline{\ell}, \overline{\ell}]$, where $\underline{\ell} = \frac{(1+t_2)p_g R^D}{2}$ and $\overline{\ell} = \frac{H(1+t_1)}{2H-(1-t_1)}$. The intuition can be explained as follows. For $\ell < \underline{\ell}$, bank's resources for redemption at date 1 are not enough to induce uninformed strategic depositors to withdraw when they observe a long withdrawal queue, that is in states 5 and 6. Condition (8) does not hold anymore and uninformed strategic depositors expect a higher utility from waiting until date 2 than from withdrawing early.[14] Differently, for $\ell > \overline{\ell}$, bank's resources at date 2 are enough to guarantee informed strategic depositors an utility from waiting greater than when withdrawing in state 5. Then, they do not join the queue even if uninformed strategic depositors withdraw mistakenly. In other words, for $\ell > \overline{\ell}$ informed strategic depositors condition their withdrawal decisions only on the signal they observe, irrespective of the amount of aggregate withdrawals. The inefficiency in state 5 becomes partial: The bank continues until date 2 but it has to liquidate valuable assets at date 1 to satisfy uninformed strategic depositors' erroneous withdrawal demands.

### 3.3.2 Debt Face Value and Banker's Monitoring Choice at Date 0

I now turn to date 0 when the bank offers the deposit contract and chooses between monitoring the project or not. The banker makes the decision about monitoring so as to maximise expected profits given that he must ensure zero expected returns to strategic depositors and given aggregate withdrawals as in proposition 1.

The face value of debt, $R^D$, is determined by strategic depositors' participation constraint, taking into account that they do not know at date 0 whether they will be informed or not at date 1 and that they deposit their funds only if the banker monitors the project.[15] Table 4 shows strategic depositors' payoff in each state of the world given their equilibrium with-

[14] Note, however, that uninformed strategic depositors might still withdraw in states 4 and 5, even if (8) is violated. This is the case when $\frac{p_g q_2 R^D}{p_g q_2 + (1-p_g) q_1} < \max\{\frac{2\ell}{1+t_1}, 1\}$ but $p_g R^D > \frac{2\ell}{1+t_2}$.

[15] If the banker does not monitor, strategic depositors would not leave their funds with the bank because they would expect to make losses. Strategic depositors' partecipation constraint in equilibrium is then calculated given that the banker monitors the project. Proposition 2 below will describe when the banker finds it optimal to do so.

drawal decisions at date 1, $w^I(s,W)$ and $w^U(W)$, as described in proposition 1. There is no participation constraint for liquidity depositors because they are not utility maximizers.

| State | $\theta=(t,x)$ | Depositors' payoff | | |
|---|---|---|---|---|
| | | Probability | Informed strategic | Uninformed strategic |
| 1 | $0,H$ | $q_0p_g$ | $R^D$ | $R^D$ |
| 2 | $0,0$ | $q_0(1-p_g)$ | 1 | 0 |
| 3 | $t_1,H$ | $q_1p_g$ | $R^D$ | $R^D$ |
| 4 | $t_1,0$ | $q_1(1-p_g)$ | $\frac{2\ell}{1+t_1}=1$ | $\frac{2\ell}{1+t_1}=1$ |
| 5 | $t_2,H$ | $q_2p_g$ | $\frac{2\ell}{1+t_2}=\frac{1+t_1}{1+t_2}$ | $\frac{2\ell}{1+t_2}=\frac{1+t_1}{1+t_2}$ |
| 6 | $t_2,0$ | $q_2(1-p_g)$ | $\frac{2\ell}{1+t_2}=\frac{1+t_1}{1+t_2}$ | $\frac{2\ell}{1+t_2}=\frac{1+t_1}{1+t_2}$ |

Table 4: Strategic depositors' payoffs

The participation constraint for strategic depositors is given by:

$$q_0p_gR^D+q_1p_gR^D+q_0(1-p_g)(\alpha\cdot 1+(1-\alpha)\cdot 0)+q_1(1-p_g)\cdot 1+q_2(\frac{1+t_1}{1+t_2})-1\geq 0$$

where:

i) the first two terms are the expected repayments for both informed and uninformed strategic depositors at date 2, conditional on the bank being solvent and on no run occurring at date 1 (states 1 and 3);

ii) the third term is the expected repayments when informed strategic depositors correctly withdraw their funds at date 1 after receiving a negative signal, while uninformed strategic depositors erroneously wait until date 2 (state 2);

iii) the fourth term is the expected repayments for both informed and uninformed strategic depositors when they correctly withdraw at date 1 and the bank has enough funds to repay them in full (state 4);

iv) the fifth term is the expected repayments for both informed and uninformed strategic depositors when they withdraw at date 1 and get only a pro-rata share of the bank's funds. It includes not only the case of the ex post inefficient run (state 5) but also the case of the efficient run with the highest realization of the liquidity shock (state 6).

The face value of debt, $R^D$, is then equal to:

$$R^D = \frac{1}{p_g(q_0 + q_1)} \left[1 - (1 - p_g)(q_0\alpha + q_1) - q_2(\frac{1 + t_1}{1 + t_2})\right]. \tag{9}$$

Note that condition (8) implies that $1 < R^D < R^S$, where $R^S$ is the repayment promised to depositors when the bank issues standard debt. This is because $p_g R^S = 1$, while $p_g R^D < \frac{1+t_1}{1+t_2} < 1$. Depositors have different rights under the two types of contracts. With standard debt they must keep their funds with the bank until date 2, when they get either $R^S$ or 0. With demandable debt they can protect—albeit partially and sometimes erroneously—their investment by withdrawing at date 1. Therefore, they are ready to pay a 'price' for this right. This is a benefit, for the bank, of the 'discipline' function of demandable debt.[16]

It is worth pointing out that $R^D < R^S$ despite the fact that information is imperfect. Hence, with demandable debt, the larger payoffs received by strategic depositors who withdraw correctly (states 4 and 6) and by informed strategic depositors in the case of inefficient continuation (state 2) more than compensate (ex ante) for the smaller payoffs received by strategic depositors who incorrectly withdraw prematurely (state 5) and by uninformed strategic depositors in the case of inefficient continuation (state 2).

I now turn to the banker's monitoring choice. Table 5 shows the bank's payoffs when the banker monitors the project and when he does not, depending on the state of the world and on depositors' withdrawal decisions as in proposition 1.[17]

If the project is monitored, the bank makes positive profits when there is no run and returns are high (states 1 and 3). If the project is not monitored, the bank makes positive profits whenever no run occurs. In particular, the bank gets the pecuniary payoff if there is no run and the project is successful (states 1 and 3) and the private benefit whenever it remains active until date 2, irrespective of the project's returns (states 1, 2 and 3).

[16] Of course, $R^D$ would be even lower if all strategic depositors had perfect information on the future value of the bank's assets.

[17] Note that in computing the bank's expected profits I take into account that, at date 2, the bank repays $R^D$ to all depositors—both strategic and 'late' liquidity—because it cannot distinguish among them.

| State | Monitoring | | No monitoring | |
|---|---|---|---|---|
| | Probability | Payoff | Probability | Payoff |
| 1 | $q_0p_g$ | $H-R^D$ | $q_0p_b$ | $(H-R^D)+B$ |
| 2 | $q_0(1-p_g)$ | 0 | $q_0(1-p_b)$ | $(1-\frac{\alpha}{2})B$ |
| 3 | $q_1p_g$ | $(1-\frac{t_1}{2})(H-R^D)$ | $q_1p_b$ | $(1-\frac{t_1}{2})\left[H-R^D+B\right]$ |
| 4 | $q_1(1-p_g)$ | 0 | $q_1(1-p_b)$ | 0 |
| 5 | $q_2p_g$ | 0 | $q_2p_b$ | 0 |
| 6 | $q_2(1-p_g)$ | 0 | $q_2(1-p_b)$ | 0 |

Table 5: Bank's payoffs

The bank's expected profits are then given by:

$$\Pi_g^D = q_0p_g(H-R^D) + q_1p_g(1-\frac{t_1}{2})(H-R^D) \tag{10}$$

when depositors correctly anticipate that the banker will monitor the project, and by:

$$\Pi_b^D = q_0p_b\left[(H-R^D)+B\right] + q_0(1-p_b)(1-\frac{\alpha}{2})B + q_1p_b(1-\frac{t_1}{2})\left[(H-R^D)+B\right] \tag{11}$$

when depositors anticipate that the banker will monitor but he does not.

The banker finds it optimal to monitor the project, and the moral hazard problem is solved, if $\Pi_g^D \geq \Pi_b^D$. For simplicity, I assume that the banker behaves well when he is indifferent between monitoring and not monitoring. Given $R^D$ and depositors' withdrawal decisions as in proposition 1, the banker's choice is determined by the magnitude of the private benefit $B$. The condition under which the banker monitors the project is shown in proposition 2.

**Proposition 2** *Given $R^D$ and depositors' optimal withdrawal decisions, there exists a level of the private benefit, $\hat{B}$, such that demandable debt solves the banker's moral hazard problem for all $B \leq \hat{B}$, where $\hat{B}$ is equal to:*

$$\hat{B} = \frac{[q_0 + q_1(1-\frac{t_1}{2})](p_g-p_b)(H-R^D)}{[q_0 + q_1p_b(1-\frac{t_1}{2}) - q_0\frac{t_1}{2}(1-p_b)]}. \tag{12}$$

**Proof** I define $\hat{B}$ as the value of $B$ which makes the banker indifferent between monitoring and not, that is $\Pi_g^D = \Pi_b^D$. After substituting the expressions for the bank's expected profits given in (10) and (11), I obtain (12). It follows immediately that, for all $B \leq \hat{B}$, $\Pi_g^D \geq \Pi_b^D$, while for all $B > \hat{B}$, $\Pi_g^D < \Pi_b^D$. □

Proposition 2 states that market discipline is effective in resolving the banker's moral hazard problem, provided this problem is not too severe. The incentive effect of demandable debt depends on the different consequences of the threat of a bank run on the bank's expected profits. When the banker does not monitor, he is always penalized when a run takes place in that he loses his private benefits in states 4, 5 and 6 and his pecuniary profits in state 5. However, when he monitors, he is penalized less often since he only loses his pecuniary profits (in state 5) when a run forces him to liquidate the (valuable) project. He does not lose anything when efficient runs take place (states 4 and 6). This is the incentive mechanism through which the threat of bank runs induces the banker to monitor the project.

Demandable debt may constitute a solution to the consequences of asymmetric information between the banker and depositors. When credit markets do not work with standard debt, the introduction of demandable debt may be the solution, as stated in proposition 3.

**Proposition 3** Since $B^S < \hat{B}$, for all $B \in (B^S, \hat{B}]$, demandable debt solves the banker's moral hazard problem but standard debt does not.

**Proof** By comparing expressions (12) and $B^S = (p_g - p_b)(H - R^S)$, it follows that $B^S < \hat{B}$. Therefore, for all $B \in (B^S, \hat{B}]$, it follows that $\Pi_g^D \geq \Pi_b^D$, but $\Pi_g^S < \Pi_b^S$. This implies the second part of the claim. □

Proposition 3 shows that demandable debt is attractive because, unlike standard debt, it allows depositors to react, although imperfectly, to the arrival of information, in the interim period, on the future value of the bank's assets. However, if the banker's moral hazard is too severe, that is for $B > \hat{B}$, market discipline is not effective in inducing the banker to monitor the project. Then, the market collapses: Depositors do not make deposits at date 0 since they expect to make negative profits, and investment does not take place.

## 3.4 Conclusions

In this chapter, I have addressed two questions: (i) Can bank runs discipline bankers who face a moral hazard problem? (ii) What are the costs and benefits of demandable debt? With regard to the first question, I have shown that the threat of bank runs may constitute an effective incentive device. In particular, I have shown that demandable debt can induce bankers to monitor the projects they finance in situations in which the standard debt contract cannot. With regard to the second question, I have shown that market discipline is costly. If some depositors are imperfectly informed on the value of the bank's assets, they may make mistakes when deciding whether to run on a bank or not: An insolvent bank may then be allowed to continue or a solvent bank may be erroneously forced into liquidation.

These results suggest a new perspective on bank regulation. Since market discipline works, but imperfectly, any sensible attempt to make it have the desirable effects should entail adequate regulatory measures aimed to eliminate its inefficiencies. One possibility would be to increase the amount of information available to depositors. If depositors are perfectly informed, market discipline would work perfectly. The recent experience of Argentina, Chile and Mexico, for example, can be read in this perspective (Peria and Schmukler (1998)).

Alternatively, regulators should secure the survival of solvent banks that are subject to a run. The historical experiences of suspension of convertibility can be looked at from this angle. For instance, before the creation of the Federal Reserve system in 1914, the US banking system often relied on the existence of clearinghouses to reduce the probability of inefficient bank runs. Clearinghouses checked the solvency of member banks when a run occurred, acting as a delegated certifier on the part of depositors and punishing banks that suspended convertibility without being solvent while, at the same time, allowing solvent institutions to stop inefficient massive withdrawals (Gorton (1985), Gorton and Mullineaux (1987)).

Recent empirical studies provide some support for the hypothesis that market discipline works. Park and Peristiani (1998) and Peria and Schmukler (1998) show that, in both developed and developing economies, depositors do react to the deterioration of banks' balance sheets. They find that depositors, whether small or large, punish risky banks by withdrawing their funds or by requiring higher interest rates. However, the ex ante effect of possible depositor withdrawals on banks' propensity to take risk remains difficult to measure.

My framework could be extended in several directions. The first concerns the analysis of information. I have focused on depositors' ability to respond to the information they have. In this light, I have abstracted from the information acquisition problem by simply assuming that some depositors receive a costless signal on the value of bank assets. Future research could make the acquisition of information costly and endogenous. Second, one could further examine the role of depositors' risk aversion in order to assess the extent to which demandable debt can be successful in solving moral hazard, while at the same time being able to guarantee optimal risk sharing among depositors. Third, one could consider a framework with multiple banks and analyse the effects of both 'specific' and 'outside' information on depositors' withdrawal decisions. Finally, future research could extend the framework by introducing explicitly a role for a financial regulator. This would allow to study in greater depth the complementarity between market discipline and supervision, which arises from my results.

## Appendix: Proof of Proposition 1

The proof consists of three steps:

Step (i): I start by assuming that informed strategic depositors' decisions depend only on the signal $s$, that is $w^I(s)$. Given $w^I(s)$, I compute uninformed strategic depositors' withdrawal decisions, $w^U(W^{(i)})$, conditional on the conjectured equilibrium at this point, $W^{(i)}$;

Step (ii): I then show that $w^I(s)$ is not optimal in the conjectured equilibrium $W^{(i)}$ and I show that informed strategic depositors' optimal withdrawal decisions, $w^I(s, W^{(ii)})$, depend also on the amount of aggregate withdrawals $W^{(ii)}$, where $W^{(ii)}$ is the conjectured equilibrium at this point;

Step (iii): I check finally that $w^U(W^{(i)})$ is still the optimal solution to uninformed strategic depositors' decision problem in the conjectured equilibrium $W^{(ii)}$ and I show that $W^{(ii)}$ is then the actual equilibrium $W$.

Note that since depositors make their withdrawal decisions simultaneously in the rational expectations equilibrium, steps (i) and (ii) never take place. Nevertheless, I go through them in order to highlight the interaction between depositors' withdrawal decisions and the mechanism through which ex post inefficient bank runs take place. I now go through each step in detail.

Step (i): The withdrawal decision of each informed strategic depositors conditional only on $s$ is:

$$w^I(s) = \begin{cases} 0 & \text{if } s = H \\ 1 & \text{if } s = 0. \end{cases}$$

The aggregate demand for withdrawals $W_D^{(i)}$ for every state of the world $\theta$ and the conjectured equilibrium $W^{(i)}$, given $w^I(s)$, are shown in table 6.

| State | $\theta = (t, x)$ | Probability | $W_D^{(i)}$ | $W^{(i)}$ |
|---|---|---|---|---|
| 1 | $0, H$ | $q_0 p_g$ | $\frac{1}{2}(1-\alpha)w^U(W^{(i)})$ | $0$ |
| 2 | $0, 0$ | $q_0(1-p_g)$ | $\frac{1}{2}[\alpha + (1-\alpha)w^U(W^{(i)})]$ | $\frac{\alpha}{2} = \frac{t_1}{2}$ |
| 3 | $t_1, H$ | $q_1 p_g$ | $\frac{1}{2}[t_1 + (1-\alpha)w^U(W^{(i)})]$ | $\frac{t_1}{2}$ |
| 4 | $t_1, 0$ | $q_1(1-p_g)$ | $\frac{1}{2}[t_1 + \alpha + (1-\alpha)w^U(W^{(i)})]$ | $\frac{1+t_1}{2}$ |
| 5 | $t_2, H$ | $q_2 p_g$ | $\frac{1}{2}[t_2 + (1-\alpha)w^U(W^{(i)})]$ | $\frac{t_2+(1-\alpha)}{2} = \frac{1+t_1}{2}$ |
| 6 | $t_2, 0$ | $q_2(1-p_g)$ | $\frac{1}{2}[t_2 + \alpha + (1-\alpha)w^U(W^{(i)})]$ | $\frac{1+t_2}{2}$ |

Table 6: Aggregate demand for withdrawals and conjectured equilibrium in step (i)

The information partitions of uninformed strategic depositors in the conjectured equilibrium $W^{(i)}$ are:

$$\begin{array}{ll} W^{(i)} = 0 & \text{implies } \theta = \{1\} \\ W^{(i)} = \frac{\alpha}{2} = \frac{t_1}{2} & \text{implies } \theta = \{2,3\} \\ W^{(i)} = \frac{1+t_1}{2} = \frac{t_2+(1-\alpha)}{2} & \text{implies } \theta = \{4,5\} \\ W^{(i)} = \frac{1+t_2}{2} & \text{implies } \theta = \{6\}. \end{array}$$

Aggregate withdrawals are perfectly informative in states 1 and 6 but not in states 2, 3, 4 and 5. Since $R^D > 1$, uninformed strategic depositors find it optimal not to withdraw when the observe $W^{(i)} = 0$, that is in state 1. Since the bank's assets are valueless, that is $x = 0$, uninformed strategic depositors find it optimal to withdraw when they observe $W^{(i)} = \frac{1+t_2}{2}$, that is in state 6.

The left hand side of (7) is the expected utility from waiting until date 2 when $\theta = \{2,3\}$. Since this is greater than 1, uninformed strategic depositors find it optimal not to withdraw when they observe $W^{(i)} = \frac{t_1}{2}$, that is in states 2 and 3.

The expected utility from waiting until date 2 when $\theta = \{4,5\}$ is $\frac{p_g q_2 R^D}{p_g q_2 + (1-p_g) q_1}$. Since this is smaller than 1 (indeed condition (8) implies $\frac{p_g q_2 R^D}{p_g q_2 + (1-p_g) q_1} < p_g R^D < \frac{1+t_1}{1+t_2} < 1$), uninformed strategic depositors find it optimal to withdraw when they observe $W^{(i)} = \frac{1+t_1}{2}$, that is in states 4 and 5.

The solution to uninformed strategic depositors' decision problem in the conjectured equilibrium $W^{(i)}$ is then equal to:

$$w^U(W^{(i)}) = \begin{cases} 0 & \text{if } W^{(i)} = 0, \frac{\alpha}{2}, \frac{t_1}{2} \\ 1 & \text{if } W^{(i)} = \frac{1+t_1}{2}, \frac{1+t_2}{2}. \end{cases}$$

Step (ii): The conjectured equilibrium $W^{(i)}$ is not an equilibrium since $w^I(s)$ is not the optimal solution to informed strategic depositors' decision problem. In state 5, in fact, they would get nothing from waiting until date 2, even if $s = H$. Given $W^{(i)} = \frac{1+t_1}{2} > \frac{t_2}{2}$ and $\ell = \frac{1+t_1}{2}$, the bank liquidates the project in order to pay off depositors withdrawing early. It is then optimal for informed strategic depositors to join the withdrawal queue so as to share the bank's resources at date 1. The optimal solution to informed strategic depositors' decision problem in the conjectured equilibrium at this stage $W^{(ii)}$ is then:

$$w^I(s, W^{(ii)}) = \begin{cases} 0 & \text{if } s = H \text{ and } W^{(ii)} < \frac{1+t_1}{2} \\ 1 & \text{if } s = 0 \text{ or } W^{(ii)} \geq \frac{1+t_1}{2}. \end{cases}$$

Step (iii): I now verify that $w^U(W^{(i)})$ still constitutes uninformed strategic depositors' optimal withdrawal decisions in the conjectured equilibrium $W^{(ii)}$. Table 7 describes the aggregate demand for withdrawals $W_D^{(ii)}$ for every state of the world $\theta$ and the conjectured equilibrium $W^{(ii)}$, given $w^I(s, W)$. Note that $W_D^{(ii)}$ and $W^{(ii)}$ differ from $W_D^{(i)}$ and $W^{(i)}$, respectively, only in state 5.

| State | $\theta = (t, x)$ | Probability | $W_D^{(ii)}$ | $W^{(ii)}$ | Run |
|---|---|---|---|---|---|
| 1 | $0, H$ | $q_0 p_g$ | $\frac{1}{2}(1-\alpha)w^U(W^{(ii)})$ | $0$ | no |
| 2 | $0, 0$ | $q_0(1-p_g)$ | $\frac{1}{2}[\alpha + (1-\alpha)w^U(W^{(ii)})]$ | $\frac{\alpha}{2} = \frac{t_1}{2}$ | no |
| 3 | $t_1, H$ | $q_1 p_g$ | $\frac{1}{2}[t_1 + (1-\alpha)w^U(W^{(ii)})]$ | $\frac{t_1}{2}$ | no |
| 4 | $t_1, 0$ | $q_1(1-p_g)$ | $\frac{1}{2}[t_1 + \alpha + (1-\alpha)w^U(W^{(ii)})]$ | $\frac{1+t_1}{2}$ | yes |
| 5 | $t_2, H$ | $q_2 p_g$ | $\frac{1}{2}[t_2 + \alpha + (1-\alpha)w^U(W^{(ii)})]$ | $\frac{1+t_2}{2}$ | yes |
| 6 | $t_2, 0$ | $q_2(1-p_g)$ | $\frac{1}{2}[t_2 + \alpha + (1-\alpha)w^U(W^{(ii)})]$ | $\frac{1+t_2}{2}$ | yes |

Table 7: Aggregate demand for withdrawals and conjectured equilibrium in step (ii)

The information partitions of uninformed strategic depositors in the conjectured equilibrium $W^{(ii)}$ are:

$$\begin{aligned} W^{(ii)} &= 0 && \text{implies } \theta = \{1\} \\ W^{(ii)} &= \tfrac{\alpha}{2} = \tfrac{t_1}{2} && \text{implies } \theta = \{2, 3\} \\ W^{(ii)} &= \tfrac{1+t_1}{2} && \text{implies } \theta = \{4\} \\ W^{(ii)} &= \tfrac{1+t_2}{2} && \text{implies } \theta = \{5, 6\}. \end{aligned}$$

Aggregate withdrawals are now perfectly informative only in states 1 and 4. As in step (i), uninformed strategic depositors find it optimal not to withdraw in states 1, 2 and 3. Since $x = 0$, uninformed strategic depositors find it optimal to withdraw when $W^{(ii)} = \frac{1+t_1}{2}$, that is in state 4.

The left hand side of (8) is the expected utility from waiting until date 2 when $\theta = \{5, 6\}$.

Since this is smaller than the utility they would get by withdrawing early, that is than $\frac{1+t_1}{1+t_2}$ (since $W^{(ii)} > \overline{W}$ and $\ell = \frac{1+t_1}{2} < 1$), uninformed strategic depositors find it optimal to withdraw when they observe $W^{(ii)} = \frac{1+t_2}{2}$, that is in states 5 and 6.

The solution to uninformed strategic depositors' decision problem in the conjectured equilibrium $W^{(ii)}$ is then:

$$w^U(W^{(ii)}) = \begin{cases} 0 & \text{if } W^{(ii)} = 0, \frac{\alpha}{2}, \frac{t_1}{2} \\ 1 & \text{if } W^{(ii)} = \frac{t_1+1}{2}, \frac{1+t_2}{2}. \end{cases}$$

Given $w^U(W^{(ii)})$, it is straightforward to verify that $w^I(s, W^{(ii)})$ is still the optimal solution to informed strategic depositors' decision problem (indeed $w^U(W^{(i)}) = w^U(W^{(ii)})$). Therefore, the conjectured equilibrium $W^{(ii)}$ coincides with the actual equilibrium $W$. □

# Chapter 4

# The Structure of Bank Relationships, Endogenous Monitoring and Loan Rates

## 4.1 Introduction

The modern theory of financial intermediation suggests that banks are valuable because of their superior ability to reduce costly information asymmetries between borrowers and lenders relative to other financial institutions. In the seminal paper by Diamond (1984), banks arise as an efficient way of delegating project monitoring, as they avoid the problems of duplication of monitoring and of free-riding that characterise direct lending. Building on this idea, several contributions in the literature argue that close bank-firm relationships facilitate banks' acquisition of information through screening and monitoring and improve firms' financing possibilities through other special features, such as contractual discretion and flexibility, collateral and covenants.[1]

Most of this literature proceeds by assuming that firms borrow from a single bank, since there is no gain in having more than one delegated monitor. Multiple banking is inefficient, as it involves high transaction costs, duplication of screening and monitoring as well as free-riding.

[1]See Freixas and Rochet (1997) for a review of the modern theory of financial intermediation and Boot (2000) for a more specific review of relationship banking.

This immediately raises a question: Why do we observe that in many countries even small firms maintain multiple bank relationships if they are costly?[2] How can this be reconciled with theory?

The existing literature has proposed two potential explanations for the benefits of multiple bank lending, the hold up and the soft budget constraint problems. Sharpe (1990) and Rajan (1992) argue that after acquiring private information about its borrowers, a relationship bank may use this informational monopoly to extract (ex post) rents. The hold up cost reduces entrepreneurial ex ante incentives to exert effort and leads to inefficient investment choices. Multiple relationships are a way for firms to avoid being locked in with a sole provider of funds and paying high rates. Sharing of proprietary information and competition among banks compete monopoly rents away and restore incentives for firm managers to exert effort (Padilla and Pagano (1997)).[3]

The soft budget constraint problem refers to the possibility that a relationship bank is unable to commit not to refinance unprofitable projects ex post. Firms may need additional credit when problems arise. A relationship bank may extend further credit even to unprofitable projects in the hope of recovering its initial loan. The possibility of renegotiating the loan reduces entrepreneurial effort in preventing default. Dewatripont and Maskin (1995) argue that multiple banking may represent a solution for the soft budget constraint problem, as it offers a way for banks to commit not to refinance unprofitable projects. This is because multiple creditors with limited funds find it more difficult to communicate and coordinate their actions, thus reducing the profitability of refinancing. In a similar spirit, Bolton and Scharfstein (1996) show that multiple banking complicates debt renegotiation but the ex post inefficiency may be beneficial ex ante, since it reduces entrepreneurial incentives to default strategically.[4] [5]

As a consequence, theories of multiple banking, building on the hold up and the soft budget constraint problems, suggest that firms borrowing from multiple banks should represent better risks and therefore pay lower loan rates than firms borrowing from a single bank. Evidence on

[2]Cross-country evidence on the number of bank relationships is provided in Ongena and Smith (1998) and Ongena and Smith (2000).

[3]A similar conclusion is reached by Von Thadden (1992).

[4]A similar conclusion is found by Povel (1995).

[5]Detragiache, Garella and Guiso (1999) offer an alternative rationale for why multiple banking may be beneficial. They argue that, when banks face a high risk of default, firms may have an incentive to establish multiple relationships to insure themselves against the loss of valuable information.

the impact of multiple banking on credit pricing and firm quality is mixed however. Petersen and Rajan (1994) find that multiple banking increases lending rates for small US firms and is not associated with an improvement of loan quality. Studies on Germany show that multiple banking does not increase lending rates but that it rather has negative effects on firm quality.[6] Foglia, Laviola and Marullo Reedtz (1998) find that multiple banking in Italy reduces lending rates and has a *negative* impact on firm quality. This result is justified by the parcellization of loans, weakening banks' disciplinary role and thus making borrowers more fragile.

How to explain the puzzles emerging when comparing the theory on multiple banking with these empirical results? One potential explanation is that the theoretical literature has not considered explicitly banks' incentives to monitor. In most of the contributions on multiple banking (such as Sharpe (1990), Rajan (1992) and Padilla and Pagano (1997)), banks acquire information on their borrowers simply as a by-product of their lending activity. In others (Von Thadden (1992)), banks' monitoring is costly but its level is exogenous. Banks can commit to monitor firms and, in equilibrium, they provide the right monitoring level, irrespective of whether they are single lenders or finance firms jointly with other banks. This chapter adds to the existing literature on multiple banking by addressing how the structure of credit relationships affects banks' monitoring decisions and how, in turn, these decisions affect loan rates and firms' choices between single and multiple lending relationships.[7]

Multiple banking creates duplications of costs in processing loan applications and reduce banks' willingness to exert an intensity of monitoring consistent with the total volume of credit granted. Without intensive monitoring, firms may find it easier to improperly use the funds obtained from banks. To address these issues, I develop a simple one-period model characterised by overlapping moral hazard problems between firms and banks.

[6]See Elsas and Krahnen (1998) and Harhoff and Koerting (1998). Ongena and Smith (1998) offer a more general review of the European evidence about the effects of relationship banking on credit pricing and firm quality.

[7]Dewatripont and Maskin (1995) also analyse banks' incentives to monitor firms and show that decentralisation of credit reduces the equilibrium level of monitoring. However, their analysis is quite different from the one developed in this chapter. In Dewatripont and Maskin, multiple banking refers to the situation in which the bank that has initially started a credit relationship with a firm does not have enough funds to refinance it in the interim period. Thus, the firm is forced to borrow from another bank. The sharing of benefits from the project with another lender reduces the incentives of the relationship bank to monitor, which however remains the sole monitor of the firm. In contrast, in this chapter multiple banking refers to the situation in which two banks finance and monitor the firm at the same time.

An entrepreneur needs bank financing in order to undertake a project. After obtaining the loan, he decides whether to exert effort and increase the success probability of the project. Effort is costly and the entrepreneur may not have incentives to exert it (firm moral hazard). Monitoring is a way for banks to control the entrepreneur's behaviour, by forcing him to exert effort and to increase the success probability of the project. However, monitoring is not contractible and is costly, exhibiting diseconomies of scale (bank moral hazard). This reduces its use as a control device, because banks choose the monitoring intensity that maximises their profits, given the terms of the loan contract, rather than the financial return of the project.

Banks' profits and, therefore, banks' incentives to monitor, depend on the size of monitoring costs, on the entrepreneur's behaviour and on the number of lenders. To keep the analysis simple, I assume that the firm has the possibility of borrowing either from one bank or from two banks. The equilibrium levels of monitoring and loan rates depend on which structure of bank relationship the firm chooses. Borrowing from two banks always implies a lower level of discipline. Multiple banking entails duplication of effort and sharing of the benefits of monitoring, which reduce banks' incentives to control the firm. However, two banks do not necessarily require a higher loan rate than a single bank as a premium for the lower probability of being repaid. The loan rate with multiple lenders depends on how the duplication of effort and the sharing of benefits interact with the convexity of monitoring costs. Under some parameter values, the relative importance of these factors is such that two lenders become cheaper than a single lender.

The firm chooses ex ante whether to borrow from one bank or from two banks. Its choice depends on the level of bank monitoring and the expected loan obligation in the two scenarios and, more precisely, on the relative *difference* between these two variables in the two cases. The structure of the credit relationships depends on the parameterisations of the two moral hazard problems in the model. In general, the firm tends to prefer a single relationship when bank moral hazard is weak and multiple relationships when bank moral hazard is strong. For intermediate values of bank moral hazard, the firm's choice depends on its own inclination for moral hazard. The greater the private return for the entrepreneur when he does not exert effort, the more likely the firm will find it optimal to borrow from two banks, although it may require a higher loan rate than with a single bank.

The present chapter makes two main contributions. The first contribution is to derive endogenously banks' monitoring decisions and loan rates as a function of the structure of bank relationships. The second contribution is to show that firm's choice between single and multiple relationships depends on both the level of discipline that banks exert through monitoring and the loan rate they require. This is in contrast to most of the literature on multiple banking, which assumes banks' monitoring activity to be exogenous.

Of course this is not the first model to address banks' incentives to monitor. Besanko and Kanatas (1993), Covitz and Heitfield (1999) and Cerasi and Daltung (1999) use a similar modelling approach of bank monitoring but in different contexts. Besanko and Kanatas (1993) rely on the non-contractibility of bank monitoring to explain the coexistence of banks and capital markets. Covitz and Heitfield (1999) focus on banks' incentives to monitor in order to analyse the relation between market power, loan rates and bank risk. Starting from the idea that delegated monitoring implies an incentive problem between banks and debtholders, Cerasi and Daltung (1999) show that diversification can improve bankers' incentives to monitor, if banks are debt financed. All these models restrict the analysis to the case where firms borrow from a single bank, which is also the sole monitor. I focus here on the relation between banks' monitoring activity and the structure of bank relationships.

The remaining of the chapter is organised in three sections. Section 2 describes the model. Section 3 derives the competitive equilibrium with bank monitoring. I proceed in three steps: I first analyse the equilibrium monitoring activity and loan rate with one bank, then with two banks, and, finally, I examine the firm's choice between the two scenarios. Section 4 concludes.

## 4.2 A Simple Model of Firm and Bank Moral Hazard

Consider a two-date economy ($T = 0, 1$) with two classes of risk neutral agents: a single firm and a perfectly competitive banking sector. The firm has access to a risky investment project but, since it has no capital, needs to raise external funds. Only bank lending is available and the firm can choose whether to borrow from one bank or from two banks simultaneously.[8] A

[8] The assumption that only bank lending is available to the firm is restrictive but the analysis of the interaction of bank lending and arm-length financing goes beyond the scope of this paper.

credit contract with a bank specifies the loan amount and the repayment obligation.[9] The firm is protected by limited liability, so that banks can be repaid only if the project succeeds. Project returns are observable. Banks raise capital at a cost equal to the riskless interest rate and lend to the firm only if they expect non-negative profits. The riskless interest rate and the discount rate are normalised to zero and one, respectively.

*The firm and the moral hazard problem*

The project available to the firm requires an outlay of one unit of capital at date 0 and yields $x = R$ per unit invested if it succeeds and $x = 0$ if it fails, at date 1. The success probability of the project is determined by the effort exerted by the entrepreneur, who is the manager and also the sole owner of the firm. The project succeeds with probability $p_H$ if the entrepreneur behaves well and with probability $p_L$ if he misbehaves, with $p_H > p_L$. The subscripts $H$ and $L$ denote high and low effort on the part of the entrepreneur. The project is creditworthy only in the case of good behaviour, that is $p_H R > 1 > p_L R$. Exerting effort is however costly for the entrepreneur and he may decide to misbehave in order to enjoy a non-transferable private benefit $B$. There is a moral hazard problem because the behaviour of the entrepreneur is neither observable nor contractible.

*Banks' financing and monitoring activities*

Since the project is economically viable only if the entrepreneur behaves well, banks are willing to finance the firm only if they anticipate that he will indeed do so. Denoting the (gross) loan rate as $r$, the entrepreneur prefers to exert a high level of effort only if:

$$p_H(R - r) > p_L(R - r) + B. \tag{1}$$

The entrepreneur's incentives depend upon the size of the private benefit $B$ and the cost of external financing $r$. To provide a benchmark against which to analyse the role of bank monitoring, I assume that the private benefit is sufficiently large to induce the entrepreneur to misbehave for all loan rates that would give banks an expected return equal to one. Substituting the

[9] As it will be clear in the following, the loan amount will be relevant only in the case of two lenders.

competitive loan rate $r = \frac{1}{p_H}$ in (1) and rearranging it gives:[10]

**Assumption 1** $p_H R - 1 - \frac{p_H}{(p_H - p_L)} B < 0$.

Assumption 1 implies that the project is not undertaken if banks lend to the firm without monitoring: The entrepreneur will not be able to raise funds because lenders anticipate that he will misbehave. One way to solve the problem is allow banks to observe and control the entrepreneur's behaviour through monitoring.

The monitoring technology is as follows: Each bank chooses the intensity of its monitoring activity $M \epsilon [0, 1]$, which corresponds to the probability of observing the entrepreneur's behaviour. If the bank monitors and finds that the entrepreneur is not behaving well, it forces him to do so. The idea is that bankers can get involved in the decision making of the firm and make sure that funds are appropriately used. Monitoring is costly and implies diseconomies of scale: An intensity $M$ costs $c(M) = \frac{m}{2} M^2$.

The first best monitoring level maximises the financial return of the project when the entrepreneur misbehaves, that is:

$$M p_H R + (1 - M) p_L R - 1 - \frac{m}{2} M^2$$

and is then given by $M^{FB} = \min\{\frac{(p_H - p_L)R}{m}, 1\}$.

I assume that the maximised financial return of the project is non-negative, that is:

$$M^{FB} p_H R + (1 - M^{FB}) p_L R - 1 - \frac{m}{2} (M^{FB})^2 \geq 0. \tag{2}$$

Substituting $M^{FB}$ in (2) implies:

**Assumption 2** $m \leq \begin{cases} 2(p_H R - 1) & \text{if } m \leq (p_H - p_L)R \\ \frac{(p_H - p_L)^2 R^2}{2(1 - p_L R)} & \text{if } m > (p_H - p_L)R. \end{cases}$

Assumption 2 puts an upper bound on the monitoring costs for the first best monitoring level to be feasible. Note that $\frac{(p_H - p_L)^2 R^2}{2(1 - p_L R)} > 2(p_H R - 1)$. This implies that the project is worth

[10] Note that, in absence of monitoring, the loan rate $r$ does not depend on the number of lenders. Indeed, $r$ is simply the total repayment the firm has to pay on one unit of external financing.

undertaking in a broader set of circumstances when $M^{FB}$ is an interior solution than when it is a corner solution.

There is a moral hazard problem in banks' monitoring activity because the intensity is neither observable nor contractible. This assumption reduces reliance on monitoring as control device for the entrepreneur's moral hazard problem: Given the inability to commit to a specific monitoring level, banks choose the intensity that maximises their expected profits and not the financial return of the project. Their expected profits, and therefore their monitoring choice, depend on the entrepreneur's behaviour and on the number of lenders, as specified below.

*Timing*

At the beginning of date 0, the firm chooses whether to borrow from one bank or from two banks. The choice between these two scenarios is observable to the market. Then the firm contacts lenders and a two-stage game starts. In the one bank game, the firm offers in the first stage a contract to one bank, which specifies the loan rate $r$. The bank decides whether to accept or reject the contract. If the bank rejects the offer, the firm does not raise funds and the game ends. If the contract is accepted, the project is financed and the second stage of the game (the continuation game) takes place. The entrepreneur chooses whether to behave well or to misbehave and the bank chooses its monitoring intensity. At date 1 project returns are realised and claims are settled. Conditional on the firm being solvent, the lender receives the loan rate $r$ and the firm retains the surplus. The two bank case has the same time structure. Figure 1 summarises the time structure of the model when the loan contract is signed.

| T=0 | | | | T=1 |
|---|---|---|---|---|
| the firm chooses whether to borrow from one or two banks | loan contract offered to the bank(s) | the entrepreneur chooses $p_H(p_L)$ | the bank(s) chooses $M$ | returns realised and claims settled |

Figure 1: Timing

## 4.3 The Competitive Equilibrium with Bank Monitoring

The equilibrium of the model consists of four elements: The firm's decision to borrow from one bank or from two banks at the beginning of date 0, the loan rate $r$ at date 0, the entrepreneur's behavioural decision and the banks' monitoring choice after financing has been secured.

I solve the model as follows: I first take the firm's choice on the number of lenders as given and I solve the two-stage continuation games with one bank and with two banks; then I analyse the firm's choice of borrowing from one bank or from two banks, taking into account the equilibrium of the continuation game in each scenario.

### 4.3.1 One Bank Financing

I start with the simple case in which only one bank finances and monitors the entrepreneur. I model the interaction between the firm and the bank as a two-stage game with imperfect information and I look for subgame perfect equilibria.

The assumption of a perfectly competitive banking system implies that the firm sets the loan rate $r_1$ at the lowest level which satisfies the zero-profit condition for the bank. In order for the project to be undertaken, the contract has to be feasible, that is $r_1$ must not exceed the total available cash flow $R$ on a successful project.

If financing has been secured at the loan rate $r_1$, the entrepreneur and the bank choose their actions in the behaviour/monitoring stage simultaneously so to maximise their expected profits.[11]

Given $r_1$ and the bank's monitoring intensity $M_1$, the firm's expected profits are given by:

$$\Pi_{F_1}^{H} = p_H(R - r_1) \tag{3}$$

when the entrepreneur behaves well and by:

$$\Pi_{F_1}^{L} = M_1 p_H(R - r_1) + (1 - M_1)[p_L(R - r_1) + B] \tag{4}$$

[11] The choice of a simultaneous second stage game between the firm and the bank does not affect the results, as long as agents' decisions are not observable. Results are affected only if the bank chooses its monitoring intensity first and its choice is observable by the entrepreneur before he moves. This order of moves would correspond to the situation in which banks can commit to a specific monitoring intensity.

when the entrepreneur misbehaves.

If the bank monitors with intensity $M_1$, its expected profits are given by:

$$\Pi_{B_1}^H = p_H r_1 - 1 - \frac{m}{2}(M_1)^2 \quad (5)$$

when the entrepreneur behaves well and by:

$$\Pi_{B_1}^L = M_1 p_H r_1 + (1 - M_1) p_L r_1 - 1 - \frac{m}{2}(M_1)^2 \quad (6)$$

when the entrepreneur misbehaves.

Formally, a competitive equilibrium in which the project is undertaken with one bank is a subgame perfect equilibrium of the continuation game that satisfies the feasibility condition, that is a vector $\{r_1^*, \{H, L\}, M_1^*\}$ such that:

1) The firm sets $r_1$ so as to maximise its expected profits subject to the bank's zero profit condition and the contract feasibility condition, correctly anticipating its behavioural choice and the bank's monitoring intensity $M_1^*$;

2) The firm and the bank maximise their expected profits given $r_1$, so that the firm's behavioural choice $\{H, L\}$ and the bank's monitoring intensity $M_1^*$ constitute a Nash equilibrium of the behaviour/monitoring subgame.

The comparison of the profit functions of the entrepreneur given in (3) and (4) shows that the entrepreneur's incentives to behave well are decreasing in the loan rate $r_1$, for given monitoring intensity $M_1$. Further, assumption 1 implies that in equilibrium the entrepreneur always misbehaves for any $M_1 \in [0, 1)$. Indeed, he finds it optimal to do so if $\Pi_{F_1}^L > \Pi_{F_1}^H$, that is if:

$$(1 - M_1)[(p_H - p_L)(R - r_1) - B] < 0 \quad (7)$$

which is satisfied given assumption 1 for any $M_1 \in [0, 1)$. When $M = 1$ the entrepreneur is indifferent between behaving well and misbehaving but the bank always forces him to behave well.

Assumption 1 also implies that, after financing has been secured, the bank always exerts a positive intensity of monitoring. Indeed, monitoring is the only way to eliminate entrepreneurial moral hazard and to have a sufficient probability of success of the project to guarantee the bank zero-profit. The equilibrium monitoring intensity maximises the bank's expected profits. The maximisation problem has a corner solution if $m \leq \overline{m}$, and an interior solution if $m > \overline{m}$. From equation (6), $\overline{m} = \frac{2(p_H - p_L)}{p_H + p_L}$ (see below for details). This means that although bank monitoring is not contractible, it can still eliminate the entrepreneurial moral hazard if it is not too costly.

Proposition 1 states conditions for the existence of a competitive equilibrium of the one bank game in which the project is undertaken and characterises the equilibrium. I restrict attention to pure strategy equilibria and I show in the appendix that this is without loss of generality. The following definition is useful for characterising the equilibrium.

**Definition 1** Define the function:

$$g_1(m) = \begin{cases} \frac{2+m}{2p_H} & \text{if } m \leq \overline{m} = \frac{2(p_H - p_L)}{p_H + p_L} \\ \frac{-mp_L + \sqrt{m^2 p_L^2 + 2m(p_H - p_L)^2}}{(p_H - p_L)^2} & \text{if } m > \overline{m} = \frac{2(p_H - p_L)}{p_H + p_L}. \end{cases}$$

The function $g_1(m)$ is continuous in $m$, with continuous derivatives. Further, it is increasing and concave in $m$, that is $g_1'(m) > 0$ and $g_1''(m) \leq 0$.

**Proposition 1** *There exists a competitive equilibrium of the one bank game in which the firm's project is financed if and only if $R \geq g_1(m)$. If it exists, the equilibrium is unique and has the following features:*

*1) $r_1^* = g_1(m)$;*

*2) the entrepreneur always chooses $L$ and the bank monitors with intensity:*

$$M_1^* = \begin{cases} 1 & \text{if } m \leq \overline{m} \\ \frac{-mp_L + \sqrt{m^2 p_L^2 + 2m(p_H - p_L)^2}}{m(p_H - p_L)} < 1 & \text{if } m > \overline{m}. \end{cases} \quad (8)$$

**Proof:**

I solve the game backwards: I find the Nash equilibrium of the behaviour/monitoring subgame given $r_1$; then I solve the loan rate setting in the first stage.

*i) The behaviour/monitoring subgame:* Define $s_{E_1} = \{H, L\}$ the pure-strategy space for the entrepreneur and $M_1 \in [0,1]$ that of the bank. The (pure) strategy Nash equilibrium of the behaviour/monitoring game is a profile $(s^*_{E_1}, M^*_1)$ such that:

$$\Pi_{F_1}(s^*_{E_1}, M^*_1) \geq \Pi_{F_1}(s_{E_1}, M^*_1) \qquad \forall s_{E_1}$$

$$M^*_1 = \arg\max_{M_1} \Pi_{B_1}(s^*_{E_1}, M_1).$$

If the entrepreneur chooses $H$, then the bank maximises (5) and chooses $M^*_1 = 0$. This is not an equilibrium: If $M^*_1 = 0$, the entrepreneur would prefer to deviate to $L$ since, as shown in (7), $\Pi^L_{F_1} > \Pi^H_{F_1}$ for any $M_1 < 0$.

If the entrepreneur chooses $L$, then the bank chooses $M_1$ so to maximise (6), which is differentiable and strictly concave in $M_1$. The first condition gives:

$$\frac{\partial \Pi^L_{B_1}}{\partial M_1} = (p_H - p_L)r_1 - mM_1 = 0$$

and, as $M_1$ must be in $[0,1]$, $M^*_1(r_1) = \min\{\frac{(p_H - p_L)r_1}{m}, 1\}$. If $M^*_1(r_1) = \frac{(p_H - p_L)r_1}{m} < 1$, the entrepreneur still prefers $L$; if $M^*_1 = 1$, he is indifferent between $H$ and $L$. Therefore, the profile $(L, M^*_1(r_1))$ is the unique pure strategy Nash equilibrium of the behaviour/monitoring subgame.

*ii) The loan setting stage:* Given $(L, M^*_1(r_1))$, the firm chooses $r_1$ to solve the following problem:

$$\max_{r_1} \Pi^L_{F_1} = M^*_1(r_1)p_H(R - r_1) + (1 - M^*_1(r_1))[p_L(R - r_1) + B]$$

subject to:

$$\Pi^L_{B_1} = M^*_1(r_1)p_H r_1 + (1 - M^*_1(r_1))p_L r_1 - 1 - \frac{m}{2}(M^*_1(r_1))^2 = 0.$$

Solving $\Pi^L_{B_1} = 0$ after substituting $M^*_1(r_1)$ gives the solution $r^*_1$. As $\Pi^L_{B_1}$ is continuous and increasing in $r_1$, there is only one value of $r^*_1$ which satisfies $\Pi^L_{B_{11}} = 0$. This is a continuous, differentiable and increasing function of $m$ and coincides with $g_1(m)$.

Substituting $r^*_1$ in $M^*_1(r^*_1)$ gives the equilibrium monitoring intensity $M^*_1$. This is a corner solution if and only if $\frac{(p_H - p_L)r_1}{m} \geq 1$ that holds, given $r^*_1$, if and only if $m \leq \overline{m}$. $M^*_1$ is an interior solution if and only if $\frac{(p_H-p_L)r_1}{m} < 1$ that holds, given $r_1^*$, if and only if $m > \overline{m} = \frac{2(p_H-p_L)}{p_H+p_L}$.

The vector $\{r_1^*, L, M_1^*\}$ is therefore the unique subgame perfect equilibrium of the one bank game. If $r_1^* \leq R$, the contract is feasible and the project is undertaken. □

Proposition 1 is noteworthy for two reasons. First, it shows that in this model monitoring does not change entrepreneurial incentives to behave well. Rather, it worsens them because it increases the loan rate, which has to cover both the amount of the loan and the monitoring costs.[12]

Second, proposition 1 implies that, given assumption 1 on the size of entrepreneur's private benefit from misbehaving, monitoring is a necessary condition for the existence of an active credit market. In the absence of monitoring, the project is never financed because of the entrepreneurial moral hazard. In its presence, the project can be financed. Monitoring forces the entrepreneur to behave well, thus increasing the probability that the project succeeds.

The parameter $m$ measures the severity of the bank moral hazard problem and determines the equilibrium monitoring level. When $m \leq \overline{m}$, the bank monitors with intensity $M_1^* = 1$. The loan rate $r_1^*$ is increasing in $m$ as monitoring costs increase with it. When $m > \overline{m}$, the bank monitors with intensity $M_1^* < 1$, which is decreasing in $m$. The loan rate $r_1^*$ is still increasing in $m$: When $M_1^*$ is an interior solution, both the monitoring intensity and the monitoring costs are decreasing in $m$ but the first effect dominates, thus increasing the loan rate.

In both circumstances, the contract between the firm and the bank is feasible and the project is always financed in equilibrium. Substituting $r_1^*$ in $r_1^* \leq R$ requires $m \leq 2(p_H R - 1)$ for $m \leq \overline{m}$ and $m \leq \frac{(p_H-p_L)^2 R^2}{2(1-p_L R)}$ for $m > \overline{m}$, which is always satisfied given assumption 2. When $r_1^* = R$, the equilibrium monitoring intensity $M_1^*$ equals the first best monitoring level $M^{FB}$; when $r_1^* < R$, $M_1^* = M^{FB}$ if $M_1^* = 1$ and $M_1^* < M^{FB}$ if $M_1^* < 1$. This is because the bank chooses its monitoring intensity as a function of its return $r_1^*$ instead of as a function of the project return $R$.

[12]Note that this result derives from assumption 1 but it is a more general consequence of the 'autorithy' aspect of monitoring. Indeed, the negative relation between entrepreneurial incentives and bank's monitoring would emerge even if, relaxing assumption 1, the entrepreneurial moral hazard problem would be less severe and monitoring would affect the entrepreneur's incentives. See for example the trade-off between loss of control and initiative in Aghion and Tirole (1997) and also in Bukart, Gromb and Panunzi (1997).

### 4.3.2 Two Bank Financing

I turn now to the case in which two (identical) banks ($i = A, B$) finance and monitor the entrepreneur. As before, the interaction between the firm and the two banks is a two-stage game with imperfect information and I look for subgame perfect equilibria. I look in particular for symmetric equilibria, in which the two banks lend a half unit each to the firm against the same loan rate and exert the same monitoring intensity.

The firm sets the loan rate $r_2$ at the lowest level which satisfies the zero-profit condition for each bank. The project is financed only if the contract is feasible, that is if $r_2$ does not exceed the return $R$ of the project. Once financing has been secured at the loan rate $r_2$, the game enters the behaviour/monitoring stage in which simultaneously the two banks and the entrepreneur choose their respective actions.

The difference between single and multiple lending crucially depends on how the two banks interact in their monitoring decisions. I assume the following decision process, which implies a trade-off in the optimality of multiple lenders (see below for details): The two banks choose their individual monitoring intensities simultaneously and independently of each other. Their monitoring efforts are however interrelated in the effect they have on the entrepreneur's behaviour. In particular, when one bank discovers the entrepreneur misbehaving, it forces him to behave well on the *whole* project. The idea is that the decision to undertake the monitoring activity is private information for each bank but the outcome of such an activity is publicly known. The *aggregate* monitoring intensity (also referred as total impact of monitoring) is then given by:

$$\overline{M}_2 = M_A + M_B - M_A M_B \tag{9}$$

where the first two terms reflects the 'positive externality' between the two banks' monitoring activities and the last term reflects the duplication of effort due to the independence of banks' decision process.

Given $r_2$ and the total monitoring impact $\overline{M}_2$, the firm's expected profits are given by:[13]

$$\Pi^H_{F_2} = p_H(R - r_2) \tag{10}$$

[13]Note that $r_2$ corresponds to the loan rate the firm pays to each bank but also to the repayment that the firm pays on the total loan of one unit.

when the entrepreneur behaves well and by:

$$\Pi^L_{F_2} = \overline{M}_2 p_H(R - r_2) + (1 - \overline{M}_2)[p_L(R - r_2) + B] \tag{11}$$

when the entrepreneur misbehaves.

If banks $A$ and $B$ monitor with intensities $M_A$ and $M_B$, respectively, then for each bank $i$ $(i = A, B)$ the expected profits are given by:

$$\Pi^H_{B_i} = p_H \frac{r_2}{2} - \frac{1}{2} - \frac{m}{2}(M_i)^2 \tag{12}$$

when the entrepreneurs behaves well and by:

$$\Pi^L_{B_i} = \overline{M}_2 p_H \frac{r_2}{2} + (1 - \overline{M}_2) p_L \frac{r_2}{2} - \frac{1}{2} - \frac{m}{2}(M_i)^2 \tag{13}$$

when the entrepreneur misbehaves.

Comparing expressions (5) and (6) with (12) and (13) highlights the differences between one and two banks: On the benefit side, two banks face a duplication of effort in the aggregate impact of monitoring $\overline{M}_2$ and each of them gets only $\frac{r_2}{2}$ while paying the full cost of its monitoring activity ("sharing of monitoring benefit"); on the cost side, they gain from the diseconomies of scale implied by the convexity of monitoring costs. The duplication of effort and the sharing of benefit affect banks' monitoring choice and, together with the convexity of monitoring costs, determine the equilibrium loan rate.

Formally, a competitive equilibrium in which the project is undertaken with two banks is a subgame perfect equilibrium of the continuation game that satisfies the feasibility condition, that is a vector $\{r^*_2, \{H, L\}, M^*_i, \forall i\}$ such that:

1) The firm sets $r_2$ so as to maximise its expected profits subject to the banks' zero profit condition and the contract feasibility condition, correctly anticipating its behavioural choice and the bank's monitoring intensity $M_i$ for $i = A, B$;

2) The firm and the banks maximise their expected profits given $r_2$, so that the firm's behavioural choice $\{H, L\}$ and the bank's monitoring intensity $M_i$ for $i = A, B$ constitute a Nash equilibrium of the behaviour/monitoring subgame.

As in the one bank case, assumption 1 implies that the entrepreneur prefers misbehaving for any $\overline{M}_2 \in [0,1)$ and is indifferent between behaving well and misbehaving for $\overline{M}_2 = 1$. This is because $\Pi^L_{F_2} \geq \Pi^H_{F_2}$ simplifies to:

$$(1 - \overline{M}_2)[(p_H - p_L)(R - r_2) - B] \leq 0 \tag{14}$$

which holds with strict inequality for any $\overline{M}_2 \in [0,1)$ by assumption 1 and with weak inequality for $\overline{M}_2 = 1$. This also implies that banks always monitor after financing the firm so as to increase the probability of success of the project.

Proposition 2 states conditions for the existence of a competitive equilibrium of the two bank game in which the project is undertaken and characterises the equilibrium. Again, I restrict attention to pure strategy equilibria and I show in the appendix that there are no mixed strategy equilibria for the two bank game. The following definition is useful for characterising the equilibrium.

**Definition 2** Let:
$f(g_2) = p_H(p_H-p_L)^2 g_2^3 + (p_H-p_L)(3mp_H+mp_L-(p_H-p_L))g_2^2 - 4m((p_H-p_L)-mp_L)g_2 - 4m^2 = 0.$

$f(g_2)$ is a cubic in $g_2$, with a positive coefficient on $g_2^3$, $f(0) = -4m^2 < 0$ and $f'(0) < 0$. Therefore, the equation $f(g_2) = 0$ has a unique positive real solution. Denote this solution by $g_2 = g_2(m)$.

**Proposition 2** *There exists a symmetric competitive equilibrium of the two bank game in which the firm's project is financed if and only if $R \geq g_2(m)$. If it exists, the equilibrium is unique and has the following features:*[14]

*1) $r_2^* = g_2(m)$;*

*2) the entrepreneur always chooses $L$ and each bank monitors with intensity:*

$$M_i^* = M_2^* = \frac{(p_H - p_L)r_2^*}{(p_H - p_L)r_2^* + 2m} < 1 \qquad \textit{for } i = A, B. \tag{15}$$

[14] I restrict attention to symmetric equilibria since I am interested in situations of multiple lending where both banks monitor the firm actively. Note however that for $(p_H - p_L)r_2 > 2m$ there exists also an asymmetric equilibrium in which only one bank monitors the firm, that is $M_i = 1$ and $M_j = 0$ for $i \neq j$.

**Proof:**

I solve the game backwards: I find the Nash equilibrium of the behaviour/monitoring subgame given $r_2$; then I analyse the loan rate setting in the first stage.

*i) The behaviour/monitoring subgame:* Define $s_{E_2} = \{H, L\}$ the pure-strategy space for the entrepreneur and $M_i \in [0,1]$ that for each bank $i$. The (pure) strategy Nash equilibrium of the behaviour/monitoring game is a profile $(s^*_{E_1}, M^*_i, \forall i)$ such that:

$$\Pi_{F_1}(s^*_{E_1}, M^*_i, \forall i) \geq \Pi_{F_1}(s_{E_1}, M^*_i, \forall i) \qquad \forall s_{E_1}$$

$$M^*_i = \arg\max_{M_i} \Pi_{B_i}(s^*_{E_1}, M_i) \text{ for all } i.$$

If the entrepreneur chooses $H$ and bank $B$ chooses $M_B$, then bank $A$ maximises (12) and chooses $M^*_A = 0$. By symmetry, $M^*_B = 0$. This is not an equilibrium: If $M^*_A = M^*_B = M^*_i = 0$ and $\overline{M}^*_2 = 2M^*_i - (M^*_i)^2 = 0$, the entrepreneur would prefer to deviate to $L$, since, as shown in (14), $\Pi^L_{F_2} > \Pi^H_{F_2}$ for any $\overline{M}_2 < 1$.

If the entrepreneur chooses $L$ and bank $B$ chooses $M_B$, bank $A$ chooses $M_A$ so to maximise (13), which is continuous and strictly concave in $M_A$. The first-order condition gives:

$$\frac{\partial \Pi^L_{B_A}}{\partial M_A} = (1 - M_B)(p_H - p_L)\frac{r_2}{2} - mM_A = 0$$

which implies:

$$M_A = \frac{(p_H - p_L)r_2}{2m}(1 - M_B). \tag{16}$$

Substituting $M_A = M_B$ in a symmetric equilibrium gives $M^*_A = M^*_B = M^*_i(r_2)$. Note that $M^*_i(r_2) < 1$, so that $\overline{M}^*_2 = (2M^*_i(r_2) - (M^*_i(r_2))^2) < 1$. Thus, the entrepreneur still prefers $L$ and the profile $(L, M^*_i(r_2)\ \forall i)$ is the unique pure strategy symmetric Nash equilibrium of the behaviour/monitoring game.

*ii) The loan setting stage:* Given $(L, M^*_i(r_2)\ \forall i)$, the firm chooses $r_2$ to solve the following problem:

$$\max_{r2} \Pi^L_{F_2} = \overline{M}^*_2(r_2)p_H(R - r_2) + (1 - \overline{M}^*_2(r_2))[p_L(R - r_2) + B]$$

subject to:

$$\Pi^L_{B_i} = \overline{M}^*_2(r_2) p_H \frac{r_2}{2} + (1 - \overline{M}^*_2(r_2)) p_L \frac{r_2}{2} - \frac{1}{2} - \frac{m}{2}(M^*_i(r_2))^2 \text{ for all } i.$$

Solving $\Pi^L_{B_i} = 0$ after substituting $M^*_i(r_2)$ gives the solution $r^*_2$. The expression $\Pi^L_{B_i} = 0$ coincides with $f(g_2(m)) = 0$ and $r^*_2 = g_2(m)$.

Substituting $r^*_2$ in $M^*_i(r^*_2)$ gives the equilibrium monitoring intensity $M^*_i = M^*_2$.

The vector $\{r^*_2, L, M^*_i\ \forall i\}$ is therefore the unique subgame perfect equilibrium of the two bank game. If $r^*_2 \leq R$, the contract is feasible and the project is undertaken. □

Proposition 2 states that the entrepreneur always chooses to misbehave and both the two banks monitor the firm in equilibrium. The expression $M^*_2 = \frac{(p_H - p_L)r^*_2}{(p_H - p_L)r^*_2 + 2m}$ reflects the drawbacks of multiple lending: The term $2m$ in the denominator results from the sharing of monitoring benefit; the term $(p_H - p_L)r^*_2$ in the denominator stems from banks' interaction in the monitoring decision process and reflects the duplication of monitoring effort. For given loan rate, the reaction function of each bank is downward sloping: In expression (16), the higher the monitoring intensity of bank $i$, $M^*_i$, the lower that of bank $j$, $M^*_j$, for each $i = A, B$ and $i \neq j$. This is because the negative effect of the duplication of monitoring effort on each bank's decision becomes more important as banks monitor more. The final result is that in equilibrium $M^*_2 < 1$ for any $p_H, p_L$, $r^*_2$, and $m$. This implies that two banks do not fully monitor even in the aggregate, that is $\overline{M}^*_2 = (2M^*_2 - (M^*_2)^2) < 1$.

The equilibrium loan rate $r^*_2$ depends on how the duplication of effort and the sharing of benefit interact, through the level of $M^*_2$, with the convexity of monitoring costs. Given the expression of the function $f(g_2)$, the analytical expression for $r^*_2$ is quite complicated and economically uninteresting. I proceed as follows: Firstly, for given loan rate, I compare the monitoring levels and the total monitoring costs in the one bank game and in the two bank game; secondly, I analyse numerically the equilibrium loan rates $r^*_1$ and $r^*_2$.

Lemma 1 compares the monitoring levels with one bank and two banks for given loan rate.

**Lemma 1** *For given $r_1 = r_2 = r$, the monitoring intensity in the one bank game is greater than the aggregate monitoring intensity in the two bank game, which is greater than each indi-*

*vidual monitoring intensity, that is:*

$$M_1^*(r) > \overline{M}_2^*(r) > M_2^*(r).$$

**Proof:**

Comparing the expressions $M_1^*(r_1)$, $M_2^*(r_2)$ and $\overline{M}_2^*(r_2)$ for $r_1 = r_2 = r$ and manipulating them gives the result:
*(i)* for $m \leq (p_H - p_L)r$, $M_1^*(r) = 1$, $\overline{M}_2^*(r) = (2M_2^*(r) - (M_2^*(r))^2) = \frac{(p_H-p_L)r(4m+(p_H-p_L)r)}{((p_H-p_L)r+2m)^2} < 1$ and $\overline{M}_2^*(r) > M_2^*(r) = \frac{(p_H-p_L)r}{(p_H-p_L)r+2m} < 1$.
*(ii)* for $m > (p_H - p_L)r$, $M_1^*(r) - \overline{M}_2^*(r) = \frac{((p_H-p_L)r)^2(3m+(p_H-p_L)r)}{m((p_H-p_L)r+2m)^2} > 0$ and $\overline{M}_2^*(r) > M_2^*(r)$.□

Lemma 1 states that, for given loan rate, banks monitor more when they are single lenders than when they finance the firm jointly. This is because the duplication of effort and the sharing of benefit in multiple banking affect negatively individual banks' incentives, thus worsening their moral hazard problem. Such effects are so strong that two lenders monitor less than a single lender even in the *aggregate.*

Figure 2 illustrates the dependence of $M_1^*(r)$, $M_2^*(r)$ and $\overline{M}_2^*(r)$ on $m \in [0,1]$ and shows that the differences $M_1^*(r) - \overline{M}_2^*(r)$ and $M_1^*(r) - M_2^*(r)$ are not monotonic in $m$: They increase for $m \leq (p_H - p_L)r$ and decrease for $m > (p_H - p_L)r$. This can be explained as follows. For $m \leq (p_H - p_L)r$, $M_2^*(r)$ and $\overline{M}_2^*(r)$ decrease as $m$ increases while $M_1^*(r)$ remains constant and equal to 1; for $m > (p_H - p_L)r$, a higher $m$ reduces all $M_1^*(r)$, $M_2^*(r)$ and $\overline{M}_2^*(r)$. However, $M_2^*(r)$ and $\overline{M}_2^*(r)$ decrease less than $M_1^*(r)$. This is because the higher $m$ the lower $M_2^*(r)$ and therefore the lower the importance of the duplication of effort in affecting the two banks' monitoring incentives.

Insert figure 2

Lemma 2 compares the total monitoring costs in the one bank case with the total monitoring costs in the two bank case for given loan rate.

**Lemma 2** *For given $r_1 = r_2 = r$, the difference between the total monitoring costs with one and two banks is:*

$$c_1(M_1^*(r)) - 2c_2(M_2^*(r)) = \begin{cases} \leq 0 & for\ m \leq \frac{(p_H-p_L)r}{2(1+\sqrt{2})} \\ \\ > 0 & for\ m > \frac{(p_H-p_L)r}{2(1+\sqrt{2})}. \end{cases}$$

**Proof:**

Comparing the relevant expressions for $c_1^*(M_1^*(r_1))$ and $2c_2^*(M_2^*(r_2))$ with $r_1 = r_2 = r$ gives the result:

(*i*) for $m \leq (p_H - p_L)r$, $c_1(M_1^*(r)) - 2c_2(M_2^*(r)) = \frac{m}{2} - m(\frac{(p_H-p_L)r}{(p_H-p_L)r+2m})^2$ which is positive only if $m > \frac{(p_H-p_L)r}{2(1+\sqrt{2})}$.

(*ii*) for $m > (p_H - p_L)r$, $c_1(M_1^*(r)) - 2c_2(M_2^*(r)) = \frac{(p_H-p_L)^2r^2}{2m} - m(\frac{(p_H-p_L)r}{(p_H-p_L)r+2m})^2 > 0$. □

Lemma 2 states that, for given loan rate, the difference between the total monitoring costs with one bank and with two banks depends on the value of the parameter $m$: It is negative when $m$ is low relative to the marginal benefit of monitoring $(p_H - p_L)r$ and positive otherwise. Figure 3 illustrates the dependence of $c_1(M_1^*(r))$ and $2c_2(M_2^*(r))$ on $m \in [0, 1]$. The intuition is as follows: For $m$ close to 0, $M_2^*(r)$ tends to $M_1^*(r) = 1$ and $2c_2(M_2^*(r))$ tends to $m$, which is greater than $c_1(M_1^*(r)) = \frac{m}{2}$. That is, when monitoring costs are very low, each of the two banks monitors approximately as much as a single bank but together they face nearly double costs. This is due to the high duplication of effort and the low importance of diseconomies of scale when $m$ is low. As $m$ becomes larger, the relative importance of the duplication of effort and of diseconomies of scale change. The total monitoring costs $2c_2(M_2^*(r))$ in the two bank case increase for $m < \frac{(p_H-p_L)r}{2}$ and decrease thereafter. The monitoring costs $c_1(M_1^*(r))$ in the one bank case increase $m < (p_H - p_L)r$ and decrease thereafter. Therefore, $2c_2(M_2^*(r))$ crosses $c_1(M_1^*(r))$ and remains below it.

Insert figure 3

Given lemmas 1 and 2, I now illustrate in figures 4-6 the equilibrium loan rates $r_1^*$ and $r_2^*$ as a function of $m \in [0, 1]$. Letting $p_H = 1$ for simplicity, the figures are drawn for different values of the success probability $p_L$ of the project when the entrepreneur misbehaves. In figure

4 $p_L = 0.4$, in figure 5 $p_L = 0.5$ and in figure 6 $p_L = 0.6$.

Insert figures 4-6

These figures are significant for two reasons. Firstly, they show that, as with $r_1^*$, $r_2^*$ is increasing in $m$ for any parameter configuration. The graphs of the monitoring intensity $M_2^*(r)$ and of the monitoring costs $2c_2(M_2^*(r))$ illustrated in figures 2 and 3 explain this result. When $m \leq \frac{(p_H - p_L)r}{2}$, the equilibrium loan rate $r_2^*$ increases with $m$ as $M_2^*(r)$ (and thus the probability of success of the project) decreases and $2c_2(M_2^*(r))$ increases. When $m > \frac{(p_H - p_L)r}{2}$, both $M_2^*(r)$ and $2c_2(M_2^*(r))$ decrease with $m$ but the first effect dominates, thus increasing $r_2^*$.

Secondly, figures 4-6 show that two lenders do not always require a higher loan rate than a single lender. The relation between $r_2^*$ and $r_1^*$ depends on the parameters of the model. When $p_L$ is low ($p_L < 0.5$) $r_2^* > r_1^*$ for any $m \in [0, 1]$ –see figure 4–; when $p_L$ is high ($p_L > 0.5$), $r_2^* > r_1^*$ for low values of $m$ but $r_1^* > r_2^*$ for $m$ sufficiently high. The two loan rates intersect at the value $\widehat{m} \in [\overline{m}, 1]$ –see figure 6–. The result is due to the joint effects of $p_L$ and $m$ on banks' expected profits and can be explained as follows. Banks require an expected repayment sufficient to cover the amount of their loan and their monitoring costs, so that they break even. Formally, the equilibrium loan rate satisfies:

$$M_1^*(r_1)p_H r_1 + (1 - M_1^*(r_1))p_L r_1 = 1 + \frac{m}{2}(M_1^*(r_1))^2 \tag{17}$$

in the case of one bank and:

$$\overline{M_2^*}(r_2)p_H r_2 + (1 - \overline{M_2^*}(r_2))p_L r_2 = 1 + m(M_2^*(r_2))^2 \tag{18}$$

in the case of two banks. As $M_1^*(r_1) \neq M_2^*(r_2)$, expressions (17) and (18) differ in the expected repayment (terms on the left hand side) and in the monitoring costs (last term on the right hand side). Lemmas 1 and 2 imply that, for given loan rate, $p_H$ and $p_L$, the expected repayment is higher with one bank for any $m \in [0, 1]$, whilst the monitoring costs are higher with two banks for low values of $m$ and become smaller as $m$ reaches $\frac{(p_H - p_L)r}{2(1+\sqrt{2})}$. Fixing $p_H = 1$ and given the loan rate, the relative differences in expressions (17) and (18) change with $m$ and $p_L$ as follows.

An increase in $p_L$ reduces the severity of the firm moral hazard, thus making bank monitoring less important. The higher $p_L$, the lower the impact of monitoring on the success probability of the project and, thus, the lower the equilibrium monitoring intensities.

In the one bank case, a higher $p_L$ shrinks the range of parameters for which the bank monitors fully, as it reduces the threshold level $\overline{m} = \frac{2(p_H - p_L)}{(p_H + p_L)}$ below which $M_1^*(r) = 1$, and reduces the monitoring intensity $M_1^*(r)$ when this is an interior solution. For $m \leq \overline{m}$, the expected repayment and the total monitoring costs in (17) are unaffected by $p_L$ as $M_1^*(r) = 1$. For $m > \overline{m}$, the expected repayment decreases as $p_L$ increases for $\overline{m} < m < 2(p_H - p_L)r$ and increases thereafter (when the direct effect of $p_L$ on the expected repayment starts to dominate that of a lower $M_1^*(r)$), whilst the monitoring costs are decreasing in $p_L$ for all $m > \overline{m}$. The loan rate $r_1^*$ is decreasing in $p_L$ for any given $m > \overline{m}$, as the effect of a higher $p_L$ on the monitoring costs (through a lower $M_1^*(r)$) dominates that on the expected repayment for $\overline{m} < m \leq 2(p_H - p_L)r$, whilst reinforcing it for $m > 2(p_H - p_L)r > \overline{m}$.

In the two bank case, a higher $p_L$ reduces $M_2^*(r)$ and $\overline{M}_2^*(r)$ for all $m$. In (18) the expected repayment decreases as $p_L$ increases for $m < \frac{(p_H - p_L)r}{2}$ and increases thereafter (when the direct effect of $p_L$ on the expected repayment starts to dominate that of a lower $\overline{M}_2^*(r)$) and the monitoring costs decrease for all $m$. An increase in $p_L$ lowers the importance of the duplication of effort through a lower $M_2^*(r)$; thus, $M_2^*(r)$ and $\overline{M}_2^*(r)$ decrease less than $M_1^*(r)$ when this is an interior solution. The loan rate $r_2^*$ is decreasing in $p_L$ for any given $m \in [0,1]$. This is because, as with $r_1^*$, the effect of a higher $p_L$ on the monitoring costs dominates that on the expected repayment for $m \leq \frac{(p_H - p_L)r}{2}$, whilst reinforcing it for $m > \frac{(p_H - p_L)r}{2}$.

These considerations show that the effects of a change in $p_L$ on the relative differences in (17) and (18) depend on the value of $m$. When $m \leq \overline{m}$, an increase in $p_L$ leaves (17) unchanged whilst it modifies both the expected repayment and the total monitoring costs in (18). The expected repayment in (18) increases for $\frac{(p_H - p_L)r}{2} < m < \overline{m}$ whilst the total monitoring costs decrease on the whole range $m \leq \overline{m}$. The net effect in (18) is not sufficiently important, however, and $r_2^*$ remain bigger than $r_1^*$ for all values of $p_L$.

When $m > \overline{m}$, an increase in $p_L$ affects both expressions (17) and (18). As $M_1^*(r)$ decreases more than $M_2^*(r)$ and $\overline{M}_2^*(r)$ but $M_1^*(r) > \overline{M}_2^*(r) > M_2^*(r)$, the expected repayment increases more in (18) than in (17), whilst monitoring costs decrease more in (17) than (18). For relatively high values of $m$ and $p_L$, the effect on the expected repayment dominates that on monitoring costs and $r_2^*$ becomes smaller than $r_1^*$.

In summary, when $m$ and $p_L$ are sufficiently high, the firm pays a lower loan rate when it borrows from two lenders than when it borrows from one. This is because a higher $p_L$ reduces the level of bank monitoring and high values of $m$ reduce the importance of the duplication of effort with two lenders whilst increasing that of diseconomies of scale. These effects reinforce each other, making it cheaper for the firm to borrow from two banks.

The parameters $m$ and $p_L$ capture the severity of the bank and the firm moral hazard problems, respectively. Regarding the firm, in the following I distinguish between the *financial* and the *private* moral hazard problem. The former, measured by $p_L$, concerns the financial return of the project when the entrepreneur misbehaves; the latter, measured by the private benefit $B$, relates to the private return that the entrepreneur enjoys from misbehaving. The two problems have different effects in the model. The financial moral hazard influences banks' monitoring incentives and the equilibrium loan rates, while the private moral hazard affects only the firm's choice between one and two banks, as I will show in the next section.

Proposition 3 summarises the equilibrium relationship between the loan rates with one and two banks.

**Proposition 3** *When the firm has weak financial moral hazard ($p_L > 0.5$) and banks have strong moral hazard ($m > \overline{m}$), two banks require a lower loan rate than a single bank for $m > \widehat{m} \in [\overline{m}, 1]$. In all other circumstances, two banks require a higher loan rate than a single bank.*

### 4.3.3 The Firm's Choice between One and Two Banks

I now turn to the firm's choice between one and two lenders. The firm makes its decision at the beginning of date 0, taking into account the equilibrium of the two-stage game with one and two banks. Bank loans are zero NPV and the entrepreneur would always prefer to misbehave but is forced to behave well by bank monitoring. Thus, after substituting (17) and (18) in (4) and (11), the firm's expected profits are given by:

$$\Pi_{F_1}^L = [M_1^* p_H R + (1 - M_1^*) p_L R] + (1 - M_1^*) B - [M_1^* p_H r_1^* + (1 - M_1^*) p_L r_1^*] =$$

$$= [M_1^* p_H R + (1 - M_1^*) p_L R] + (1 - M_1^*) B - [1 + \frac{m}{2}(M_1^*)^2] \qquad (19)$$

when the firm borrows from one bank and as:

$$\Pi_{F_2}^L = [\overline{M}_2^* p_H R + (1 - \overline{M}_2^*) p_L R] + (1 - \overline{M}_2^*) B - [\overline{M}_2^* p_H r_2^* + (1 - \overline{M}_2^*) p_L r_2^*] =$$

$$= [\overline{M}_2^* p_H R + (1 - \overline{M}_2^*) p_L R] + [(1 - \overline{M}_2^*) B] - [1 + m(M_2^*)^2] \qquad (20)$$

when the firm borrows from two banks. In expression (19), the first term is the total expected financial return of the project, the second term is the expected private return for the entrepreneur and the last term is the bank's total expected repayment, which has to cover both the amount of the loan and the total monitoring costs. Terms in expression (20) have exactly the same meaning.

The firm prefers to have a single lender if (19)>(20) and two lenders otherwise. Firm's expected profits are a function of the parameters $m$, $p_H$, $p_L$, $B$ and $R$. The parameter $m$ captures the size of the monitoring costs; $p_H$ and $p_L$ are the success probability of the project when the entrepreneur behaves well and misbehaves, respectively; $B$ is the private benefit that the entrepreneur enjoys from misbehaving while $R$ is the financial return of the project in case of success. In the following, I fix $p_H$ and $R$ and analyse the firm's expected profits as a function of $m$, $B$ and $p_L$. These parameters measure, respectively, the bank moral hazard, the firm private moral hazard and the firm financial moral hazard. Thus, I can interpret the firm's choice between one lender and two lenders in terms of the relative severity of bank moral hazard as compared to firm moral hazard, as I will discuss below.

Letting $p_H = 1$ and $R = 1.6$, figures 7-9 show the firm's expected profits given in (19) and (20) as a function of $m$, $B$ and $p_L$. In each figure, which is drawn for given $p_L$, $m \in [0, 1]$ is depicted on the x-axis while the different curves for $\Pi_{F_1}^L$ and $\Pi_{F_2}^L$ are associated with different values of $B \in [0.4, 0.8]$, with a jump of 0.1 on each curve. The three figures differ in the value of $p_L$. In particular, $p_L = 0.4$ in figure 7, $p_L = 0.5$ in figure 8, and $p_L = 0.6$ in figure 9. These configurations of parameters satisfy assumptions 1 and 2 and guarantee that the loan contract is feasible with both one bank and two banks for any $m \in [0.1]$. The graphs do not change appreciably with other parameter configurations.

In the one bank case, the monitoring intensity $M_1^*$ is a corner solution for $m \leq \overline{m}$ and an interior solution for $m > \overline{m}$. This implies that $\Pi_{F_1}^L$ depends on $B$ only for $m > \overline{m}$. Thus, the graph of $\Pi_{F_1}^L$ in figures 7-9 is a single line for $m \leq \overline{m}$ and a bundle of curves for $m > \overline{m}$, each of them corresponding to a different value of $B$. The corner solution $M_1^* = 1$ implies also that $\Pi_{F_1}^L$ does not depend on $p_L$ for $m < \overline{m}$. Thus, the graph of $\Pi_{F_1}^L$ for $m < \overline{m}$ is the same in all figures. However, as the threshold level $\overline{m} = \frac{2(p_H - p_L)}{p_H + p_L}$ is decreasing in $p_L$, it reduces in figures 7-9, as $p_L$ increases. In particular, $\overline{m} = 0.85$ in figure 7, $\overline{m} = 0.6\overline{6}$ in figure 8 and $\overline{m} = 0.5$ in figure 9. This implies that the range of values of $m$ in which $M_1^* = 1$ and the graph of $\Pi_{F_1}^L$ is a single line reduces from figure 7 to 9 as $p_L$ increases.

In the two bank case, the monitoring intensity $M_2^*$ is always an interior solution, which implies that the aggregate monitoring intensity $\overline{M_2^*}$ is also less than 1. Thus, $\Pi_{F_2}^L$ depends on $B$ for all $m \in [0,1]$ and its graph in figures 7-9 is always a bundle of curves, each of them corresponding to a different value of $B$.

insert figures 7-9

These figures show that the structure of bank relationships crucially depends on the values of $m$, $B$ and $p_L$.

For $m \leq \overline{m}$, in each figure there exists $\widetilde{m}(B, p_L) \in [0, \overline{m}]$ such that, for given $p_L$ and $B$, the firm prefers to borrow from one bank if $m < \widetilde{m}(B, p_L)$ and from two banks if $m > \widetilde{m}(B, p_L)$. For given $p_L$, the threshold level $\widetilde{m}(B, p_L)$ is decreasing in $B$. In each figure, the higher $B$, the smaller the value of $\widetilde{m}(B, p_L)$ which determines the firm's choice. In figure 7, for example, for given $p_L = 0.4$, $\widetilde{m}(0.4, p_L) = 0.58$, $\widetilde{m}(0.5, p_L) = 0.46$, $\widetilde{m}(0.6, p_L) = 0.36$, $\widetilde{m}(0.7, p_L) = 0.27$ and $\widetilde{m}(0.8, p_L) = 0.21$. For given $B$, the threshold level $\widetilde{m}(B, p_L)$ is decreasing in $p_L$. Comparing figures 7-9 shows that the higher $p_L$, the lower is the threshold $\widetilde{m}(B, p_L)$ for a given level of $B$. For example, fixed $B = 0.4$, $\widetilde{m}(B, 0.4) = 0.58$ in figure 7, $\widetilde{m}(B, 0.5) = 0.41$ in figure 8 and $\widetilde{m}(B, 0.6) = 0.24$ in figure 9.

For $m > \overline{m}$, the firm prefers two lenders for all $B \in [0.4, 0.8]$ and for $p_L \in [0.4, 0.6]$.

This result can be explained as follows. The firm decides between one and two lenders on the basis of two variables: loan rate and bank monitoring activity. These two elements affect the firm's expected profits in the two scenarios differently, depending on the parameters of the model.

In the one bank case, $\Pi^L_{F_1}$ is decreasing in $m$ for $m \leq \overline{m}$, independently of $B$ and $p_L$, as $M_1^* = 1$ and $r_1^*$ increases with $m$. For $m > \overline{m}$ and given $p_L$, $\Pi^L_{F_1}$ is decreasing in $m$ for low values of $B$ but is increasing in $m$ for higher values of $B$. This is because a higher $m$ reduces $M_1^*$, thus increasing $r_1^*$. For given $p_L$, a lower $M_1^*$ reduces the total expected financial return of the project, $M_1^* p_H R + (1 - M_1^*) p_L R$, decreases the total expected repayment, $M_1^* p_H r_1^* + (1 - M_1^*) p_L r_1^*$, and increases the expected private return, $(1 - M_1^*)B$. For high $B$, these two last effects dominate and $\Pi^L_{F_1}$ increases with $m$. The level of $B$ above which $\Pi^L_{F_1}$ depends positively on $m$ decreases with $p_L$. This is because, as illustrated in figures 4-6, for given $m$, $r_1^*$ decreases with $p_L$ when $m > \overline{m}$, although $M_1^*$ decreases too.

In the two bank case, $\Pi^L_{F_2}$ is decreasing in $m$ for low values of $m$ and $B$, whilst it increases with $m$ for higher values of $m$ and $B$. The higher $p_L$ the more significant the positive relation between $\Pi^L_{F_2}$ and $m$ is. The explanation is similar to that in the one bank case for $m > \overline{m}$.

Given a level of $p_L$, the firm prefers to borrow from a single bank for very low values of $m$, independently if $B$. This is because when $m$ is very low, $r_2^* > r_1^*$ whilst $M_2^*$ tends to $M_1^*$. Thus, the firm pays a lower loan rate with one bank, while facing a similar monitoring discipline.

As $m$ increases for $m \leq \overline{m}$, the difference $(r_2^* - r_1^*)$ increases even further but $M_2^*$ decreases and $M_1^*$ remains constant. This implies a lower expected financial return of the project $\overline{M_2^*} p_H R + (1 - \overline{M_2^*}) p_L R$, a lower total expected repayment $\overline{M_2^*} p_H r_2^* + (1 - \overline{M_2^*}) p_L r_2^*$ but a higher expected private return $(1 - \overline{M_2^*})B$ in the two bank case relative to the one bank case. Thus, for given $B$ and $p_L$, there exists a level $\widetilde{m}(B, p_L) < \overline{m}$ beyond which the firm chooses to borrow from two banks as the higher expected private return and the lower expected repayment dominate the lower financial return of the project . The higher $B$, the more the firm benefits from not being monitored and, therefore, the lower the threshold level $\widetilde{m}(B, p_L)$ for given $p_L$.

As $m$ increases further and reaches $\overline{m}$, the difference $(r_2^* - r_1^*)$ decreases (and it becomes negative in the case $p_L > 0.5$ for $m > \widehat{m} > \overline{m}$, as illustrated in figure 6), as well as the difference $(M_1^* - \overline{M_2^*})$ (but $M_1^* > \overline{M_2^*}$ for all $m > \overline{m}$ and all $p_L \in [0.4, 0.6]$). The first effect dominates and the firm finds it convenient to borrow from two lenders for all $m > \overline{m}$, independently of $B$.

The value of $p_L$ affects the range of parameters in which the firm prefers two lenders. A higher $p_L$ reduces both the threshold level $\overline{m}$ and the critical value $\widetilde{m}(B, p_L)$, thus making the firm's choice of two banks more likely. This is due to the effects of a higher $p_L$ on the loan rates and the monitoring intensity in the one bank case and in the two bank case, which in turn affect the total expected financial return of the project, the expected private return for the entrepreneur and the total expect repayment in the two scenarios.

As before, the parameters $m$, $p_L$ and $B$ can be interpreted as a measure of the bank moral hazard, the firm financial moral hazard and the firm private moral hazard, respectively. Proposition 4 summarises the firm's choice between one and two lenders in terms of these moral hazard problems.

**Propositon 4** *When bank moral hazard is very weak, the firm prefers to borrow from one bank, independently of its own moral hazard.*
*As bank moral hazard becomes more severe for* $m \leq \overline{m}$*, there exists a level of* $\widetilde{m}(B, p_L) \in [0, \overline{m}]$*, such that, given the firm financial and private moral hazard, the firm borrows from one banks for* $m < \widetilde{m}(B, p_L)$ *and from two banks for* $m > \widetilde{m}(B, p_L)$*. The more severe the firm private moral hazard and the weaker the firm financial moral hazard, the lower the threshold* $\widetilde{m}(B, p_L)$ *and thus, the more likely the firm borrows from two banks.*
*When bank moral hazard is very strong* ($m > \overline{m}$)*, the firm prefers to borrow from two banks, independently of its own moral hazard.*

## 4.4 Conclusions

Monitoring of entrepreneurs is widely recognised as one of the main roles that banks play in the economy. However, how much monitoring they exert and the effects of this activity on credit pricing and the structure of bank relationships is not much explored in the existing literature of financial intermediation.

In this chapter I have developed a one-period model of overlapping moral hazard problems between borrowers and lenders, which focuses on banks' incentives to monitor and analyses why firms may find it optimal to establish multiple relationships, although they entail duplication of effort, free riding and potentially also a higher repayment obligation.

The idea is quite simple. An entrepreneur may prefer not to exert enough effort and enjoy a private benefit instead of being diligent and increasing the success probability of the project. Banks can control the entrepreneur's behaviour through monitoring but they cannot commit to a predetermined monitoring level. The number of lenders affects the monitoring intensity that banks exert in equilibrium as well as the loan rate they require. Multiple banking always implies a lower level of discipline than single banking, whereas -as a consequence of diseconomies of scale in monitoring- the relation between the number of lenders and the loan rate is not monotonic. The firm's choice between single and multiple relationships is not univocal but depends on the relative parameters of the two moral hazard problems in the model.

Future research could extend the analysis in several directions. First, by using a one-period model with a perfectly competitive banking system, I have abstracted from the potential hold-up problem that might arise in the case of a single lender. This could be considered by extending the model to two periods and by allowing banks to either extract a surplus ex post or operate in an imperfectly competitive sector. A two-period model would also allow for an analysis of renegotiation with single and multiple banking. Second, the entrepreneur's effort choice is modelled as a discrete variable and he always chooses not to exert effort, independently of bank monitoring. One could model the entrepreneur's choice as a continuous variable and analyse the impact of bank monitoring on entrepreneur's incentives.

**Appendix: Mixed Strategy Equilibria in the Behaviour/Monitoring Subgame**

*1. One Bank Financing*

I now look for mixed strategy equilibria in the behaviour/monitoring subgame with one bank and show that restricting attention to pure strategy Nash equilibria in Proposition 1 is without loss of generality.

Entrepreneur's indifference between $H$ and $L$ implies:

$$p_H(R-r_1) = M_1 p_H(R-r_1) + (1-M_1)[p_L(R-r_1)+B]$$

which has solution $\widehat{M_1} = 1$.
If the entrepreneur chooses $H$ with probability $x$, then the bank maximises:

$$x[p_H r_1 - 1 - \frac{m}{2}(M_1)^2] + (1-x)[M_1 p_H r_1 + (1-M_1)p_L r_1 - 1 - \frac{m}{2}(M_1)^2]$$

which simplifies to:

$$x p_H r_1 + M_1(1-x)p_H r_1 + (1-M_1)(1-x)p_L r_1 - 1 - \frac{m}{2}(M_1)^2.$$

The first-order condition with respect to $M_1$ gives:

$$(1-x)(p_H - p_L)r_1 = mM_1. \quad (21)$$

Substituting $\widehat{M_1} = 1$ in (21) implies:

$$\widehat{x} = 1 - \frac{m}{(p_H - p_L)r_1}$$

where $0 < \widehat{x} < 1$ if $\frac{m}{(p_H-p_L)r_1} < 1$. Thus, the strategy profile $(\widehat{x}, \widehat{M_1})$ is a mixed strategy Nash equilibrium of the game if $\frac{m}{(p_H-p_L)r_1} < 1$.

The outcome of the mixed strategy equilibrium is 'equivalent' to that of the pure strategy equilibrium $(L, \min\{\frac{(p_H-p_L)r_1}{m}, 1\})$ shown in Proposition 1. In both equilibria, if $(p_H - p_L)r_1 > m$, $M_1 = 1$ and the entrepreneur is always forced to behave well, independently of his strategy

($\widehat{x}$ or $L$). Therefore, restricting attention to pure strategy equilibria is without loss of generality. □

*2. Two Banks Financing*

I now show that the behaviour/monitoring subgame with two banks has no mixed strategy Nash equilibria.

Entrepreneur's indifference between $H$ and $L$ implies:

$$p_H(R - r_2) = \overline{M}_2 p_H(R - r_2) + (1 - \overline{M}_2)[p_L(R - r_2) + B]$$

where $\overline{M}_2 = 2M_i - (M_i)^2$ with $i = A, B$. The solution is $\widehat{M_i} = \widehat{M_A} = \widehat{M_B} = 1$.
If the entrepreneur chooses $H$ with probability $x$ and bank $B$ chooses $M_B$, then bank $A$ maximises:

$$x[p_H \frac{r_2}{2} - \frac{1}{2} - \frac{m}{2}(M_A)^2] + (1 - x)[\overline{M}_2 p_H \frac{r_2}{2} + (1 - \overline{M}_2) p_L \frac{r_2}{2} - \frac{1}{2} - \frac{m}{2}(M_A)^2]$$

which simplifies to:

$$x p_H \frac{r_2}{2} + \overline{M}_2(1 - x) p_H \frac{r_2}{2} + (1 - \overline{M}_2)(1 - x) p_L \frac{r_2}{2} - 1 - \frac{m}{2}(M_A)^2.$$

The first-order condition with respect to $M_A$ gives:

$$(1 - M_B)(1 - x)(p_H - p_L)\frac{r_2}{2} = m M_A. \tag{22}$$

Substituting $\widehat{M_i} = \widehat{M_A} = \widehat{M_B} = 1$ implies that the game has no (nondegenerate) mixed strategy Nash equilibrium since there does not exist any $\widehat{x} > 0$ which satisfies (22). □

Figure 2: Monitoring intensities with one bank and with two banks for given r

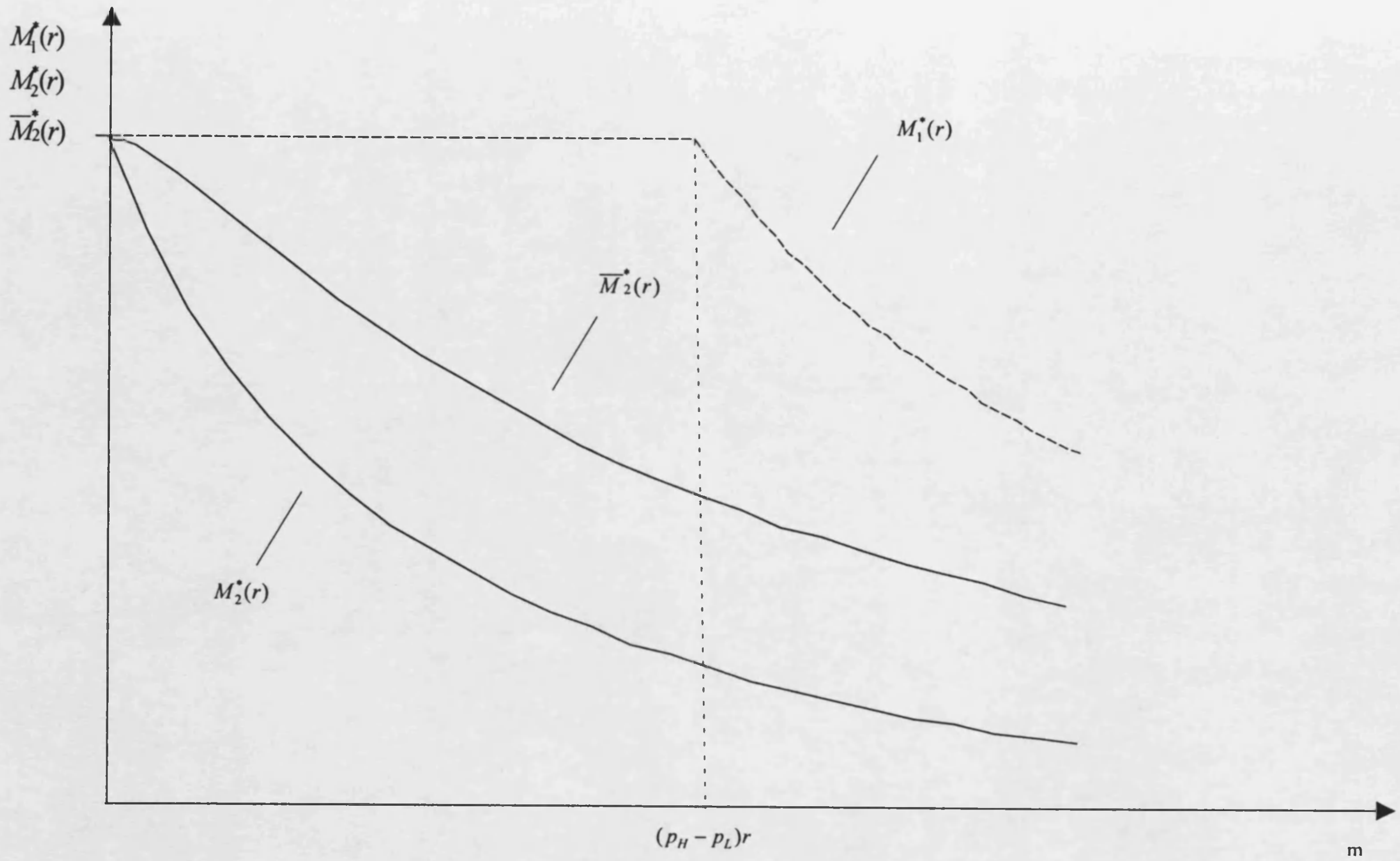


Figure 3: Total monitoring costs with one bank and with two banks for given r

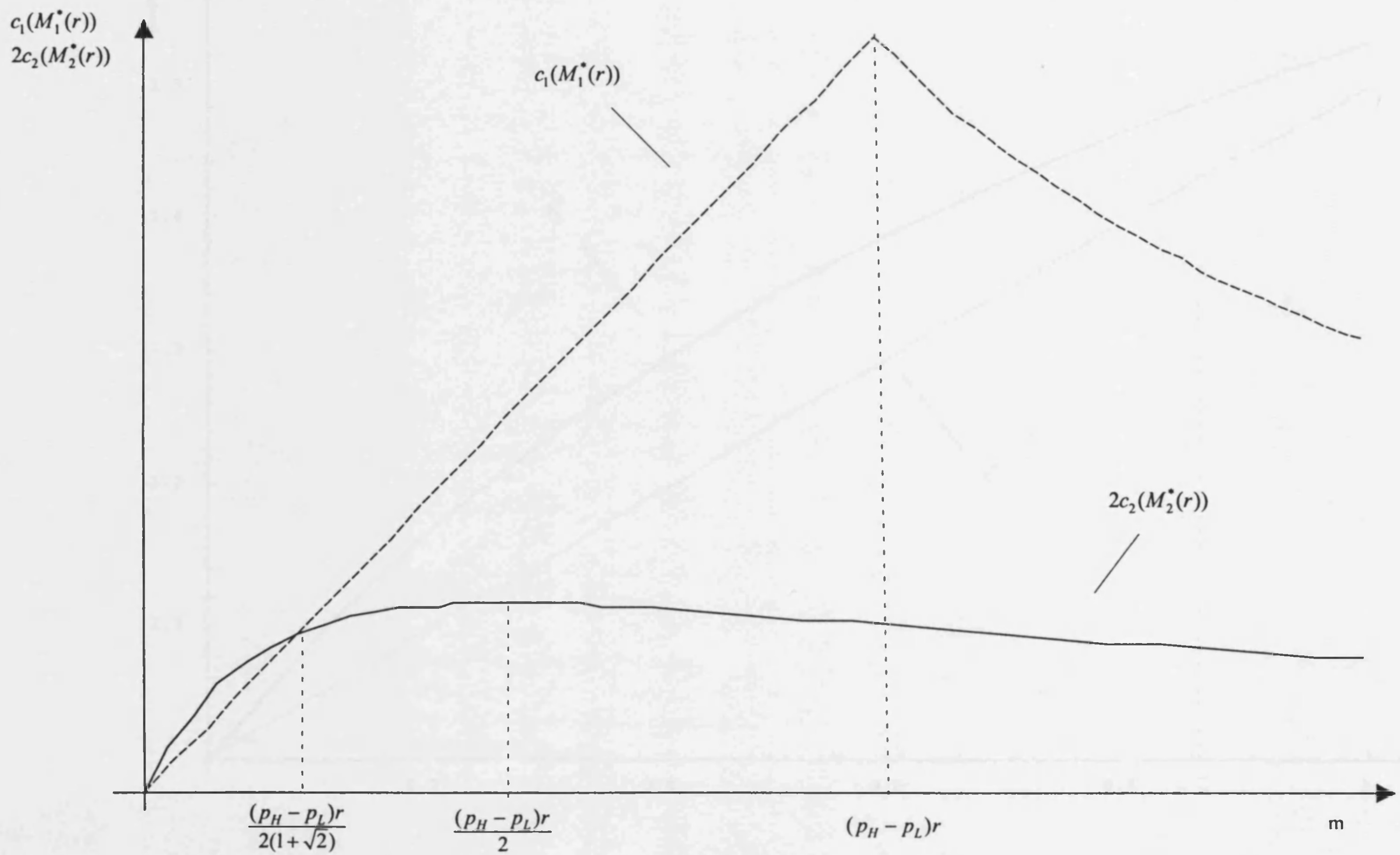


Figure 4: Loan rates with one bank and with two banks for pl=0.4

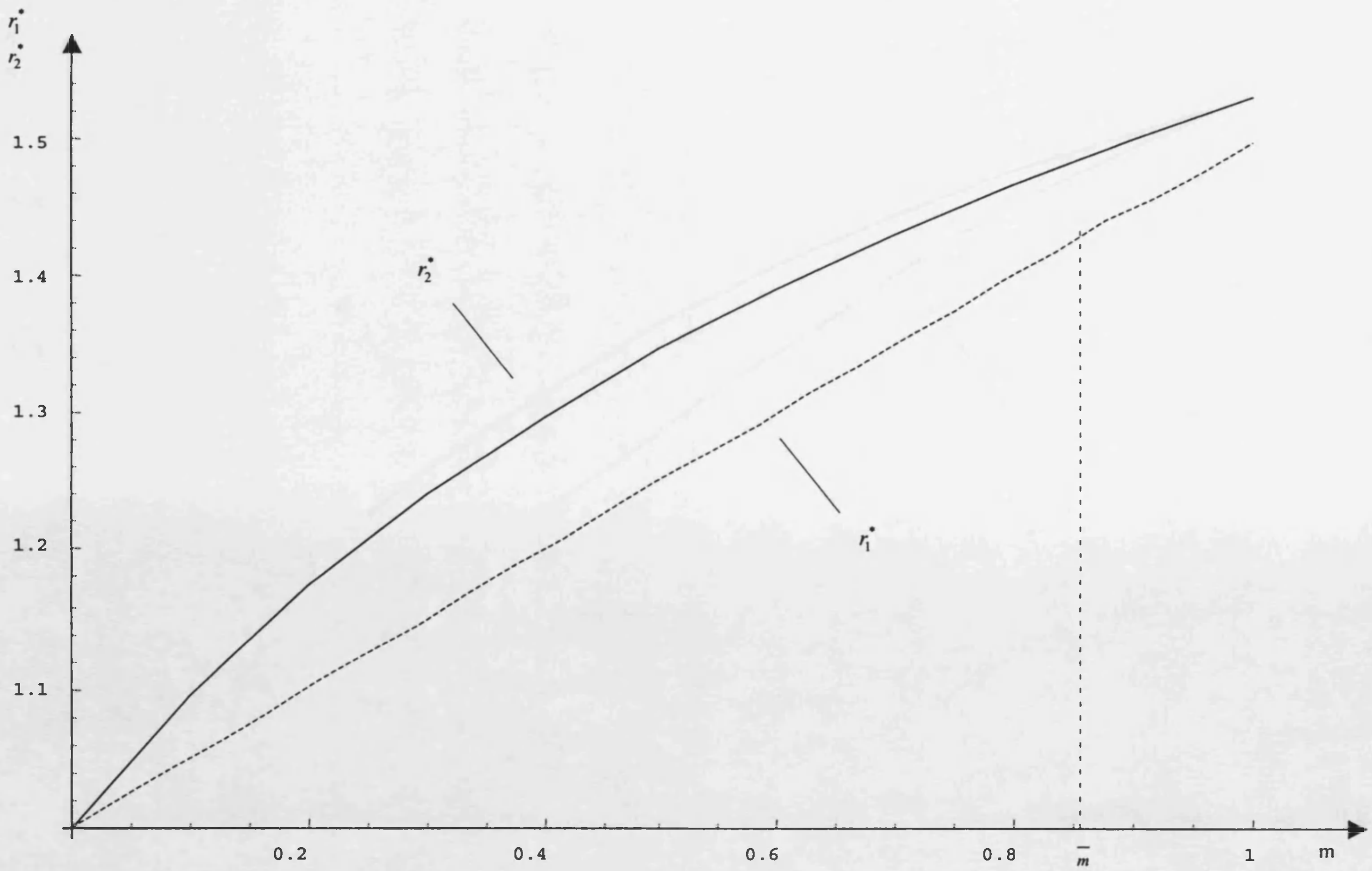


Figure 5: Loan rates with one bank and with two banks for pl=0.5

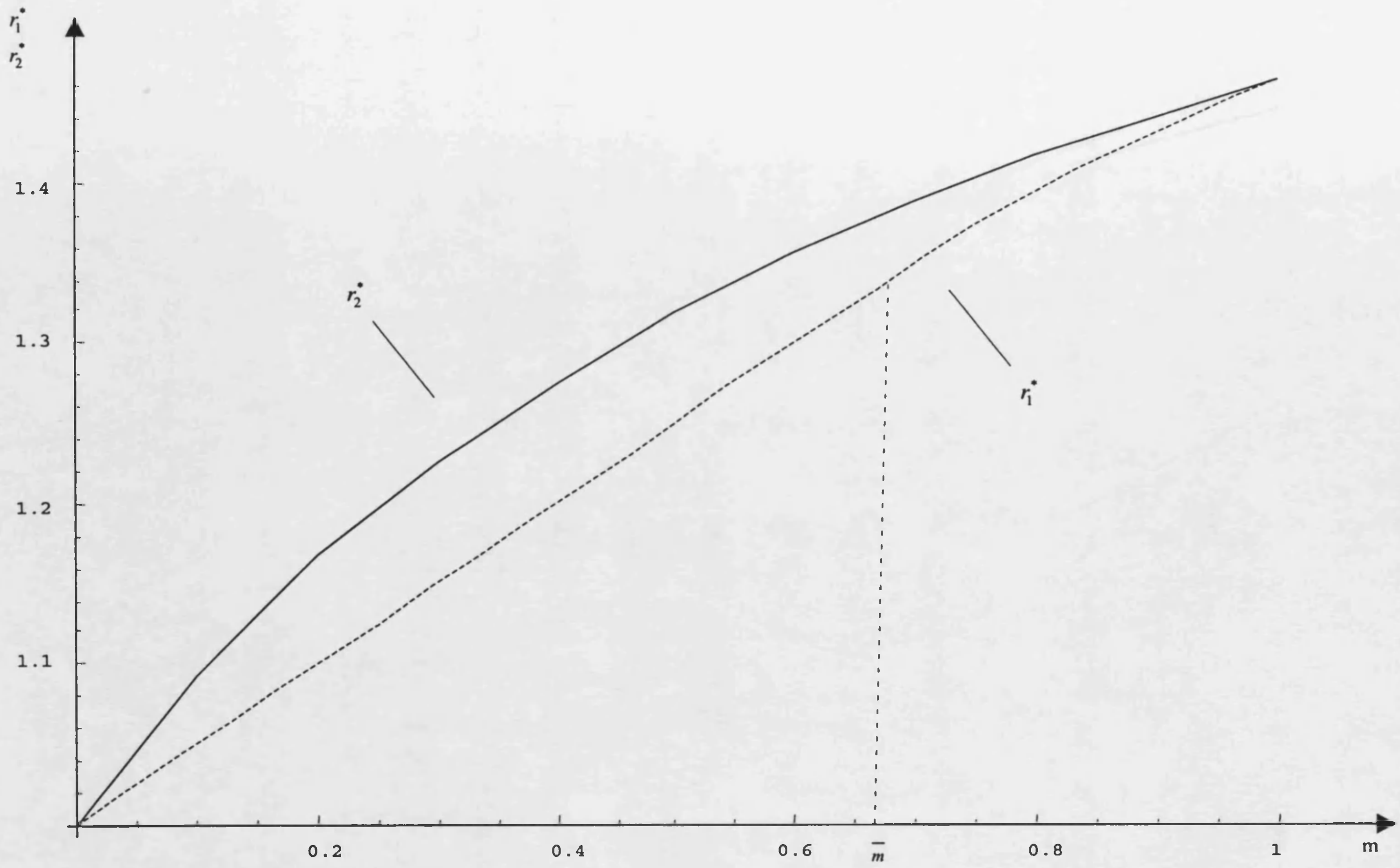


Figure 6: Loan rates with one bank and with two banks for pl=0.6

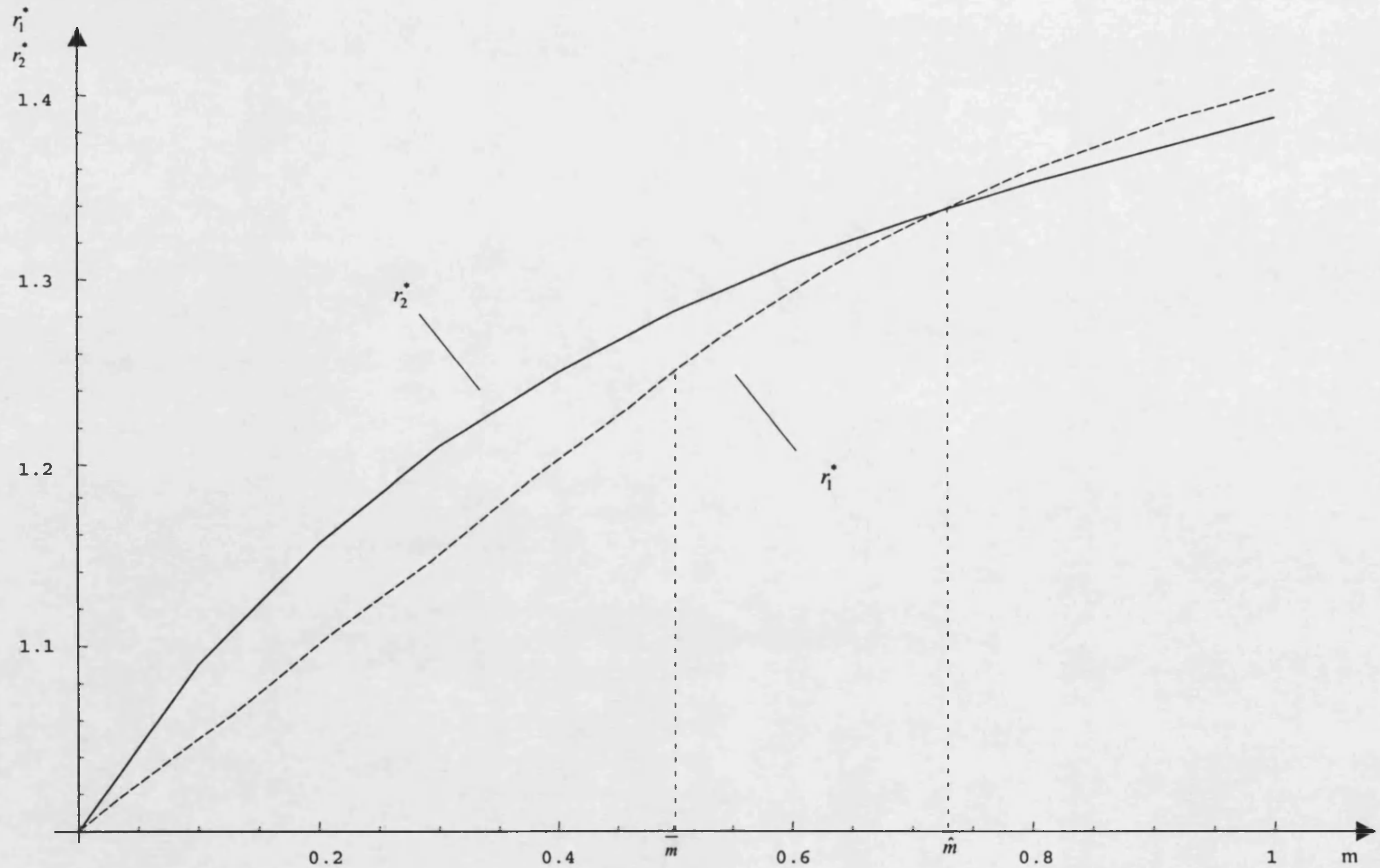


Figure 7: Firm's expected profits with one bank and with two banks as a function of m and B for pl=0.4

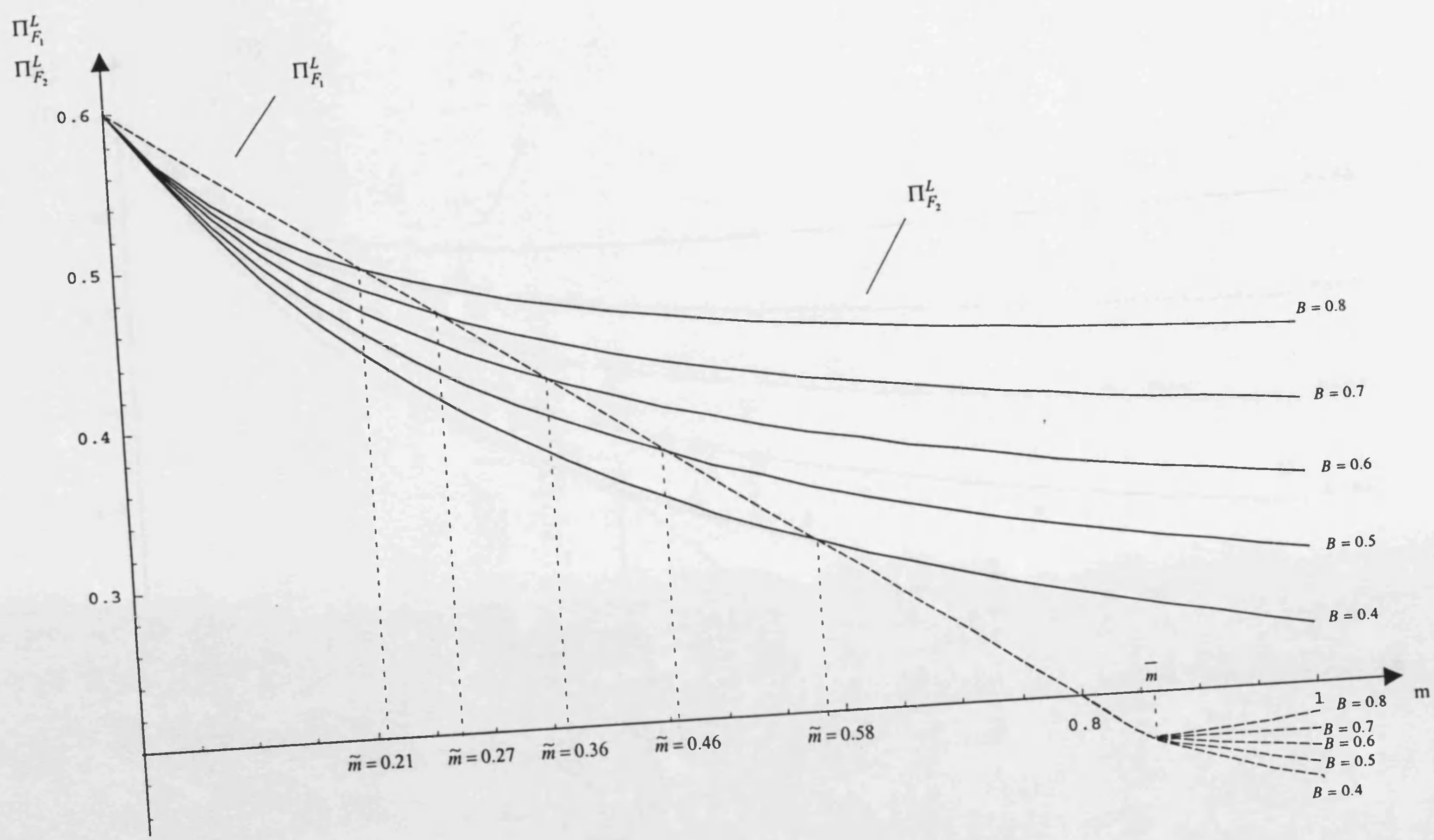


Figure 8: Firm's expected profits with one bank and with two banks as a function of m and B for pl=0.5

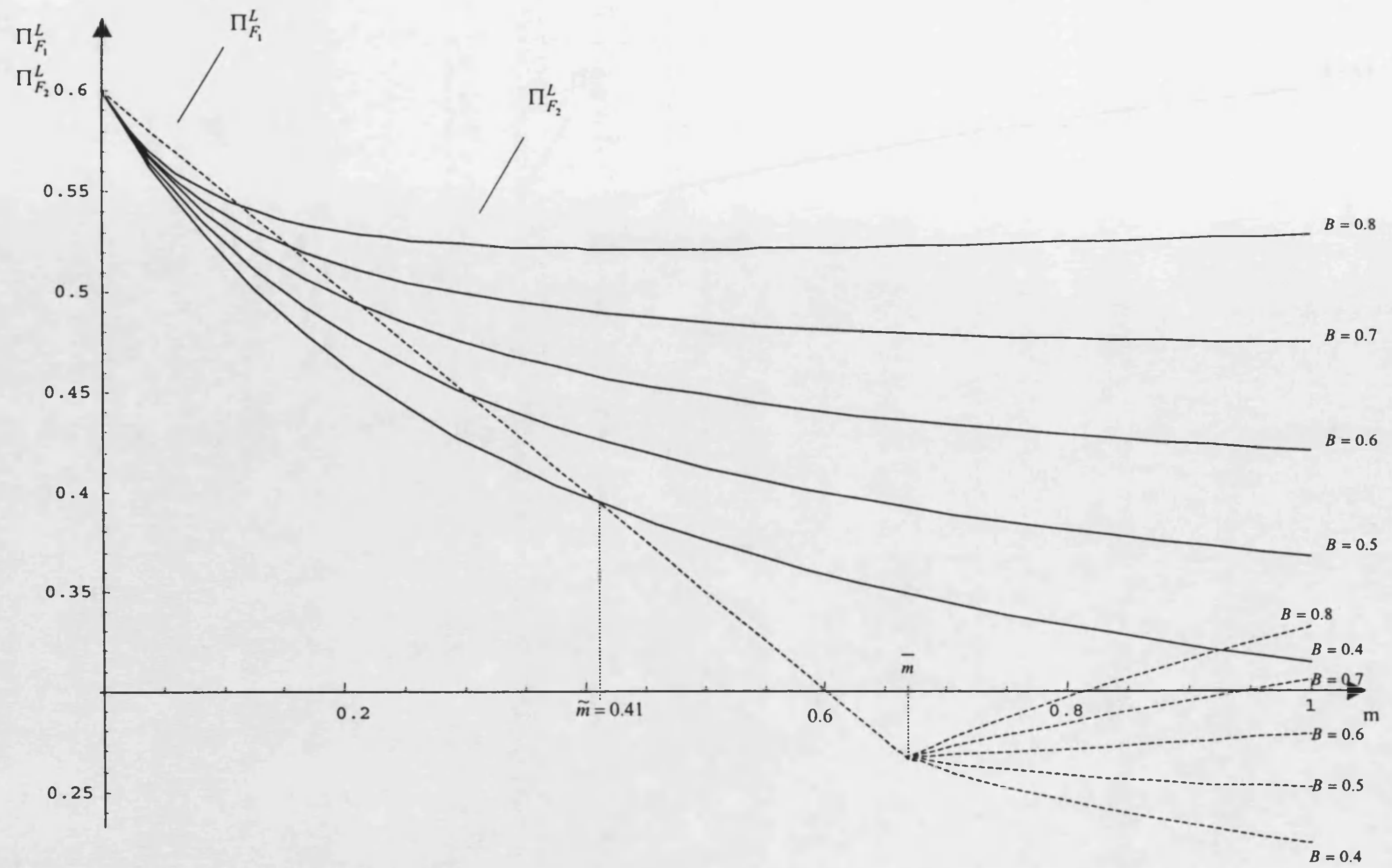
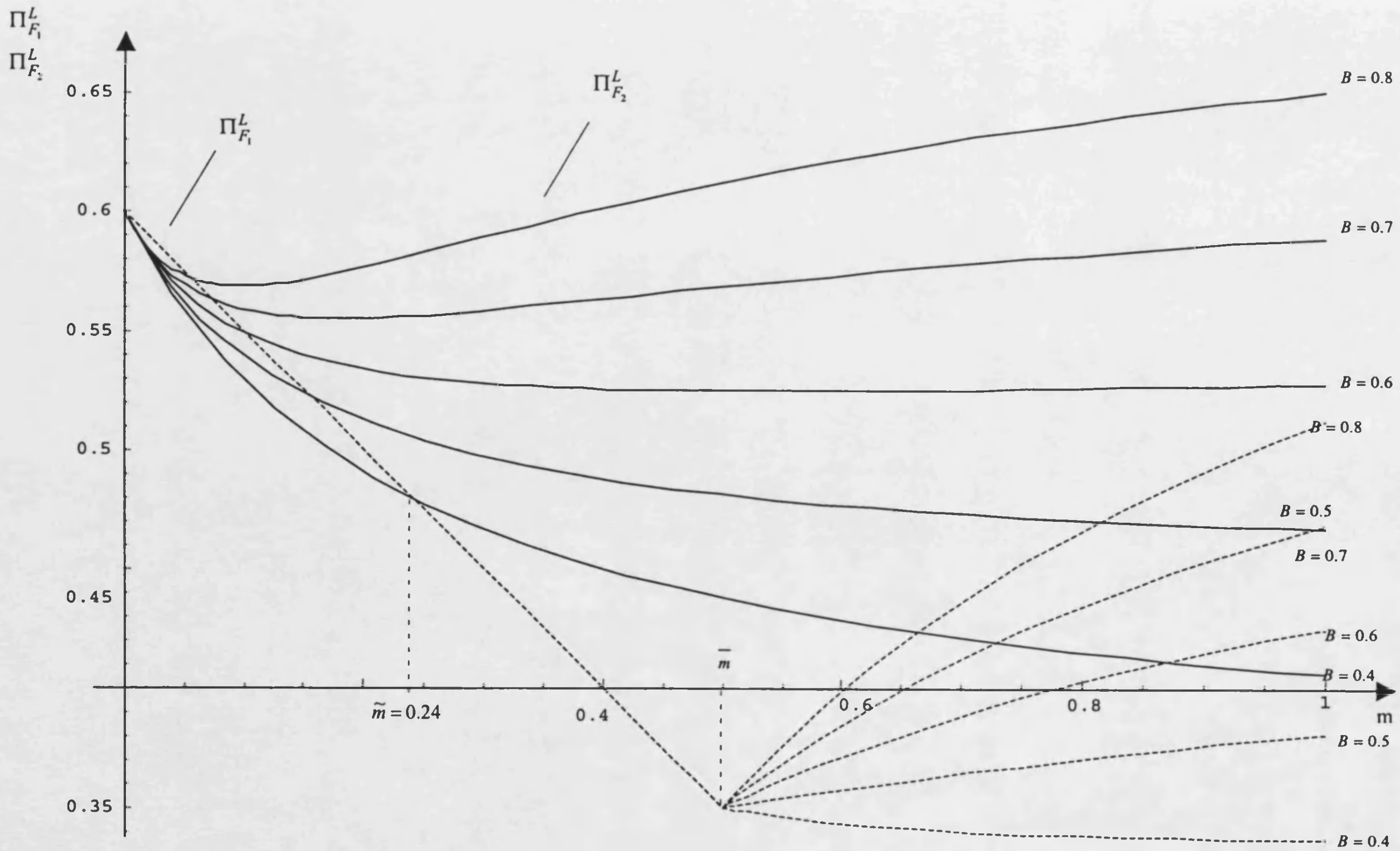


Figure 9: Firm's expected profits with one bank and with two banks as a function of m and B for pl=0.6



# Bibliography


Acharya S. and S. Dreyfus, 1989, "Optimal Bank Reorganisation Policies and the Pricing of Federal Deposit Insurance", *Journal of Finance*, 44, 1313-1333

Aghion P. and J. Tirole, 1997, "Formal and Real Autorithy in Organizations", *Journal of Political Economy*, 105, 1-29

Aghion P., P. Bolton and M. Dewatripont, 1999, "Contagious Bank Failures", forthcoming in *European Economic Review*

Akerlof G.A., 1970, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, 84, 488-500

Allen F. and D. Gale, 1998, "Optimal Financial Crises", *Journal of Finance*, 47, 1245-1284

Allen F. and D. Gale, 2000, "Financial Contagion", *Journal of Political Economy*, 108, 1-34

Alonso I., 1996, "On Avoiding Bank Runs", *Journal of Monetary Economics*, 37, 73-87

Alworth J.S. and S. Bhattacharya, 1995, "The Emerging Framework of Bank Regulation and Capital Control", LSE Financial Markets Group Special Paper Series, n. 78

Bagehot W., 1873, "*Lombard Street*", Kegan, Paul & Co, London

Baltensperger E., 1980, "Alternative Approaches to the Theory of Banking Firm", *Journal of Monetary Economics*, 6, 1-37

Baltensperger E. and E. Dermine, 1987, "The Role of Public Policy in Ensuring Financial Stability: A Cross-country, Comparative Perspective", in R. Portes and A. Swoboda (eds.) *Threats to International Stability*, Cambridge University Press, Cambridge

Baltensperger E. and E. Dermine, 1993, "Deregolamentazione Bancaria in Europa" traduzione italiana in Marotta G. and G.B. Pittaluga (eds.) *La Teoria degli Intermediari Bancari*, Il Mulino, Bologna

Bensaid B., H. Pages and J.C. Rochet, 1995, "Efficient Regulation of Banks' Solvency", Universite des Sciences Sociales, Centre National de la Recherche Scientifique, Ecole des Hautes Etudies en Sciences Sociale, Cahier n. 95.07.370, G.R.E.M.A.Q.

Benston G.J., R. Eisenbeis, P. Horwitz, E. Kane and G. Kaufman, 1986, "*Perspectives on Safe and Sound Banking: Past, Present and Future*", MIT Press, Cambridge MA

Benston G.J. and C.W. Smith, 1976, "A Transaction Costs Approach to the Theory of Financial Intermediation", *Journal of Finance*, 31, 215-31

Berlin M., A. Saunders and G.F. Udell, 1991, "Deposit Insurance Reform: What are the Issues and What Needs to be Fixed?", *Journal of Banking and Finance*, 15, 735-752

Besanko D. and G. Kanatas, 1993, "Credit Market Equilibrium with Bank Monitoring and Moral Hazard", *Review of Financial Studies*, 6, 213-232

Besanko D. and G. Kanatas, 1996, "The Regulation of Bank Capital Regulations: Do Capital Standards Promote Bank Safety?", *Journal of Financial Intermediation*, 5, 160-83

Besanko D. and A.V. Thakor, 1993, "Relationship Banking, Deposit Insurance and Bank Portfolio", in C. Mayer and X. Vives (eds.), *Capital Markets and Financial Intermediation*, Cambridge University Press, Cambridge

Bhattacharya S. and A. Thakor, 1993, "Contemporary Banking Theory", *Journal of Financial Intermediation*, 3, 2-50

Bhattacharya S., A.W.A. Boot and A. Thakor, 1998, "The Economics of Bank Regulation", *Journal of Credit, Money and Banking*, 30, 745-771

Bolton P. and D. Scharfstein, 1996, "Optimal Debt Structure and the Number of Creditors", *Journal of Political Economy*, 104, 1-25

Boot A.W.A., 2000, "Relationship Banking: What do We Know?", *Journal of Financial Intermediation*, 9, 7-25

Boot A. and S. Greenbaum, 1993, "Bank Regulation, Reputation and Rents: Theory and Policy Implications", in C. Mayer and X. Vives (eds.), *Capital Markets and Financial Intermediation*, Cambridge University Press, Cambridge

Boot A. and A. Thakor, 1993, "Self-interested Bank Regulation", *American Economic Review*, 83, 206-212

Bordo M., 1990, "The Lender of Last Resort: Alternative Views and Historical Experience", *Federal Reserve Bank of Richmond Economic Review*, 76, 18-29

Boyd J. and M. Gertler, 1993, "U.S. Commercial Banking: Trends, Cycles and Policy" , NBER working paper, n. 4404

Broaddus J.A., 1994, "Choices in Banking Policy", *Federal Reserve Bank of Richmond*, 80, 1-10

Bruni F., 1995, "Prudential Regulation in an Integrated Financial Market: Issues of Optimality and Credibility", in G. Ferrarini (ed.), *Prudential Regulation of Banks and Securities Firms: European and International Aspects*", Kluwer

Bryant J., 1980, "A Model of Reserves, Bank Runs and Deposit Insurance", *Journal of Banking and Finance*, 4, 335-344

Burkart M., D. Gromb and F. Panunzi, 1997, "Large Shareholders, Monitoring, and the Value of the Firm", *Quarterly Journal of Economics*, 112, 693-728

Calomiris C., 1993, "Financial Factors in the Great Depression", *Journal of Economic Perspectives*, 7, 61-86

Calomiris C., 1998, "Blueprint for a New Global Financial Architecture", mimeo, Columbia Business School, Columbia

Calomiris C. and G. Gorton, 1991, "The Origins of Banking Panics: Models, Facts, and Bank Regulation", in Hubbard R. G. (ed.), *Financial Markets and Financial Crises*, University of Chicago Press, Chicago

Calomiris C.W. and C.M. Kahn, 1991, "The Role of Demandable Debt in Structuring Optimal Banking Arrangements", *American Economic Review*, 81, 497-513

Calomiris C. and J.R. Mason, 1997, "Contagion and Bank Failures during the Great Depression: The June 1932 Chicago Banking Panic", *American Economic Review*, 87, 863-883

Caminal R. and C. Matutes, 1997a, "Bank Solvency, Market Structure and Monitoring Incentives", CEPR Discussion Paper, n. 1665

Caminal R. and C. Matutes , 1997b, "Can Competition in the Credit Market be Excessive?", CEPR Discussion Paper, n. 1725

Campbell T., Y. Chan and A. Marino, 1992, "An Incentive-Based Theory of Bank Regulation", *Journal of Financial Intermediation*, 2, 255-276

Caprio G. Jr. and D. Klingebiel, 1996, "Bank Insolvency: Bad Luck, Bad Policy or Bad Banking?", paper prepared for the World Bank's Annual Bank Conference on Development Economics

Carletti E., 1999, "Bank Moral Hazard and Market Discipline", LSE Financial Markets Group Discussion Paper, n. 326

Carletti E., 2000, "Competition, Regulation and Stability", in M. Polo (ed.), *Industria Bancaria e Concorrenza*, Il Mulino, Bologna

Cerasi V. and S. Daltung, 1999, "The Optimal Size of a Bank: Costs and Benefits of Diversification", forthcoming in *European Economic Review*

Chan Y., S. Greenbaum and A. Thakor, 1992, "Is Fairly Priced Deposit Insurance Possible?", *Journal of Finance*, 47, 227-246

Chari V.V. and R. Jagannathan, 1988, "Banking Panics, Information and Rational Expectations Equilibrium", *Journal of Finance*, 43, 749-763

Chen Y., 1999, "Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities", *Journal of Political Economy*, 107, 946-968

Cordella T. and L. Yeyati, 1998, "Financial Opening, Deposit Insurance and Risk in a Model of Banking Competition", CEPR Discussion Paper, n. 1939

Covitz D. and E. Heitfield, 1999, "Monitoring, Moral Hazard and Market Power: A Model of Bank Lending", FEDS Working Paper, n. 99-37

De Bandt O. and P. Hartmann, 1999, "Systemic Risk: A survey", mimeo, European Central Bank, Frankfurt

De Palma A. and R.J. Gary-Bobo, 1996, "Coordination Failures in the Cournot Approach to Deregulated Bank Competition", mimeo, THEMA

Deshmukh S.D., S. Greenbaum and G. Kanatas, 1983, "Lending Policies of Financial Intermediaries Facing Credit and Funding Risk", *Journal of Finance*, 38, 873-86

Detragiache E., P. Garella and L. Guiso, 1999, "Multiple vs. Single Banking Relationships: Theory and Evidence", forthcoming in *Journal of Finance*

Dewatripont M. and J. Tirole, 1994, *The Prudential Regulation of Banks*, MIT Press, Cambridge MA

Dewatripont M. and E. Maskin, 1995, "Credit and Efficiency in Centralized and Decentralized Economies", *Review of Economic Studies*, 62, 541-555

Diamond D.W., 1984, "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, 51, 393-414

Diamond D.W., 1991, "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt", *Journal of Political Economy*, 99, 689-721

Diamond D.V. and P.H. Dybvig, 1983, "Bank Runs, Deposit Insurance and Liquidity", *Journal of Political Economy*, 91, 401-419

Diamond D.V. and R. G. Rajan, 1999, "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking", mimeo, University of Chicago, Chicago

Dowd K., 1989, "Automatic Stabilizing Mechanism under Free Banking", *Cato – Journal*, 7, 643-59

Dowd K., 1992, "Models of Banking Instability: A Partial Review of the Literature", *Journal of Economic Surveys*, 6, 107-133

Edwards F. and F. Mishkin, 1995, "The Decline of Traditional Banking: Implications for Financial Stability and Regulatory Policy", *Federal Reserve Bank of New York Economic Policy Review*, 1, 27-45

Elsas R. and J.P. Krahnen, 1998, "Is Relationship Banking Special? Evidence from Credit-File Data in Germany", *Journal of Banking and Finance*, 22, 1283-1316

Engineer M., 1989, "Bank Runs and the Suspension of Deposit Convertibility", *Journal of Monetary Economics*, 24, 443-454

Federal Reserve of Chicago, 1990, "What we know about the Deposit Insurance Problem", *Chicago Fed Letter*, 35, 1-4

Foglia A., S. Laviola and P. Marullo Reedtz, 1998, "Multiple Banking Relationships and the Fragility of Corporate Borrowers", *Journal of Banking and Finance*, 22, 1441-1456

Freixas X., 1999, "Optimal Bail-out Policy, Conditionality and Creative Ambiguity", LSE Financial Markets Group Discussion Paper, n. 327

Freixas X. and B. Parigi, 1998, "Contagion and Efficiency in Gross and Net Interbank Payment Systems", *Journal of Financial Intermediation*, 7, 3-31

Freixas X., B. Parigi and J.C. Rochet, 1999, "Systemic Risk, Interbank Relations and Liquidity Provision by the Central Bank", mimeo, Universite des Sciences Sociales, Toulose

Freixas X. and J.C Rochet, 1997, *Microeconomics of Banking*, MIT Press, Cambridge MA

Freixas X. and J.C. Rochet, 1998, "Fair Pricing of Deposit Insurance. Is it possible? Yes. Is it Desiderable? No", *Research in Economics*, 52, 217-232

Gehrig T., 1995, "Excessive Risk Taking and Banking Regulation", mimeo, Universitat Basel, Basel

Gennotte G. and D. Pyle, 1991, "Capital Controls and Bank Risk", *Journal of Banking and Finance*, 15, 805-824

Gianmarino R.M., T.R. Lewis and D.E.M. Sappington, 1993, "An Incentive Approach to Banking Regulation", *Journal of Finance*, 48, 1523-1542

Goodfriend M. and R. King, 1988, "Financial Deregulation, Monetary Policy and Central Banking", *Federal Reserve Bank of Richmond Economic Review*, 74, 3-22

Goodhart C.A.E., 1987, "Why do Banks need a Central Bank?", *Oxford Economic Papers*, 39, 75-89

Goodhart C.A.E., 1995, "Some Regulatory Concerns", LSE Financial Market Group, Special Paper Series, n. 79

Goodhart C.A.E., P. Hartmann, D.T. Llewellyn, L. Rojas-Suarez and S.R. Weisbrod, 1998, *Financial Regulation: Why, How and Where now?*, Monograph prepared for the Bank of England Central Bank Governors' Symposium 1997, London, Routledge

Goodhart C.A.E. and H. Huang, 1999, "A Model of the Lender of Last Resort", International Monetary Fund Working Paper, n. 99/39

Gorton G., 1985, "Clearinghouses and the Origin of Central Banking in the United States", *Journal of Economic History*, 55, 277-283

Gorton G., 1988, "Banking Panics and Business Cycles", *Oxford Economic Papers*, 40, 751-781

Gorton G. and D.J. Mullineaux, 1987, "The Joint Production of Confidence: Endogenous Regulation and Nineteenth Century Commericial-Bank Clearinghouses", *Journal of Money, Credit and Banking*, 19, 457-468

Guttentag J.M. and R.J. Herring, 1986, "Disaster Myopia in International Banking", *Essays in International Finance*, n. 164

Harhoff D. and T. Koerting, 1998, "Lending Relationships in Germany - Empirical Evidence from Survey Data", *Journal of Banking and Finance*, 22, 1317-1353

Hellwig M., 1995, "Systemic Aspects of Risk Management in Banking and Finance", mimeo, Universitat Basel, Basel

Holmstrom B. and J. Tirole, 1997, "Financial Intermediation, Loanable Funds and the Real Sector", *Quarterly Journal of Economics*, 52, 663-691

Jacklin C.J. and S. Bhattacharya, 1988, "Distinguishing Panics and Information-based Bank Runs: Welfare and Policy", *Journal of Politycal Economy*, 96, 568-592

James C., 1991, "The Losses realized in Bank Failures", *Journal of Finance*, 46, 1223-1242

John K., T.A. John and Sembet, 1991, "Risk-shifting Incentives of Depository Institutions: A New Perspective on Federal Deposit Insurance Reform", *Journal of Banking and Finance*, 15, 898-916

Kahane Y., 1977, "Capital Adequacy and the Regulation of Financial Intermediaries", *Journal of Banking and Finance*, 2, 207-218

Kane E.J., 1985, *The Gathering Crisis in Federal Deposit Insurance*", MIT Press, Cambridge MA

Kane E.J.,1989, "Changing Incentives facing Financial-services Regulators", *Journal of Financial Services Research*, 265-274

Kane E.J., 1990, "Principal-Agent Problems in S&L Salvage", *Journal of Finance*, 45, 755-64

Kareken J.M. and N. Wallace, 1978, "Deposit Insurance and Bank Regulation: A Partial-Equilibrium Exposition", *Journal of Business*, 51, 413-38

Kaufman G.G., 1994, "Bank Contagion: A Review of the Theory and Evidence", *Journal of Financial Services Research*, 8, 135-160

Keeley M. C., 1990, "Deposit Insurance, Risk and Market Power in Banking", *American Economic Review*, 80, 1183-1200

Keely M.C. and F.T. Furlong, 1990, "A Reexamination of Mean-Variance Analysis of Bank Capital Regulation", *Journal of Banking and Finance*, 14, 69-84

Koehn M. and M. Santomero, 1980, "Regulation of Bank Capital and Portfolio Risk", *Journal of Finance*, 35, 1235-50

Leland H.E. and D. Pyle, 1977, "Informational Asymmetries, Financial Structure and Financial Intermediation", *Journal of Finance*, 32, 371-387

Mailath G.J. and L.J. Mester, 1994, "A Positive Analysis of Bank Closure", *Journal of Financial Intermediation*, 3, 272-299

Matutes C. and X. Vives, 1996, "Competition for Deposits, Fragility and Insurance", *Journal of Financial Intermediation*, 5, 184-216

Matutes C. and X. Vives, 2000, "Imperfect Competition, Risk Taking and Regulation in Banking", *European Economic Review*, 44, 1-35

Merton R., 1977, "An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees", *Journal of Banking and Finance*, 1, 3-11

Merton R., 1978, "On the Costs of Deposit Insurance when there are Surveillance Costs", *Journal of Business*, 51, 439-52

Miron J.A., 1986, "Financial Panics, the Seasonality of the Nominal Interest Rate and the Founding of Fed", *American Economic Review*, 76, 125-140

Mitchell J., 1998, "Strategic Creditor Passivity, Regulation and Bank Bailouts", CEPR Discussion Paper, n. 1780

Nagarajan S. and C.W. Sealey, 1995, "Forbearance, Deposit Insurance Pricing and Incentive Compatible Bank Regulation", *Journal of Banking and Finance*, 19, 1109-1130

Nagarajan S. and C.W. Sealey, 1997, "Can Delegating Bank Regulation to Market Forces Really Work?", working paper, New York University, New York

Ongena S. and D.C. Smith, 1998, "Bank Relationships: A Review", forthcoming in P. Harker and S.A. Zenios (eds.), *The Performance of Financial Institutions*, Cambridge University Press, Cambridge

Ongena S. and D.C. Smith, 2000, "What Determines the Number of Bank Relationships? Cross-Country Evidence", *Journal of Financial Intermediation*, 9, 26-56

Padilla A.J. and M. Pagano, 1997, "Endogenous Communication among Lenders and Entrepreneurial Incentives", *Review of Financial Studies*, 10, 205-236

Padoa-Schioppa T., 1999, "EMU and Banking Supervision", LSE Financial Markets Group Special Paper, n. 112

Parigi B., 2000, "Competition in Banking: A Survey of the Literature", in M. Polo (ed.), *Industria Bancaria e Concorrenza*, Il Mulino, Bologna

Park S. and S. Peristani, 1998, "Market Discipline by Thrift Depositors", *Journal of Money, Credit and Banking*, 30, 347-364

Peria M.S.M. and S. Schmukler, 1998, "Do Depositors Punish Banks for 'Bad' Behavior?: Examining Market Discipline in Argentina, Chile and Mexico", mimeo, World Bank, Washington

Petersen M. and R.G. Rajan, 1994, "The Benefit of Lending Relationships: Evidence from Small Business Data", *Journal of Finance*, 49, 1367-1400

Postlewaite A. and X. Vives, 1987, "Bank Runs as an Equilibrium Phenomenon", *Journal of Political Economy*, 95, 485-491

Povel P., 1995, "Multiple Banking as a Commitment Not to Rescue", LSE Financial Markets Group Discussion Paper, n. 206

Qi J., 1998, "Deposit Liquidity and Bank Monitoring", *Journal of Financial Intermediation*, 7, 198-218

Rajan R.G., 1992, "Insiders and Outsiders: The Choice between Informed and Arm's-Lenght Debt", *Journal of Finance*, 47, 1367-1400

Rochet J.C. and J. Tirole, 1996a, "Controlling Risk in Payment Systems", *Journal of Money, Credit and Banking*, 28, 832-862

Rochet J.C. and J. Tirole, 1996b, "Interbank Lending and Systemic Risk", *Journal of Money, Credit and Banking*, 28, 733-762

Saunders A. and B. Wilson, 1998, "Contagious Bank Runs: Evidence from the 1929-33 Period", *Journal of Financial Intermediation*, 5, 409-23

Schoenmaker D., 1996a, "Contagion Risk in Banking", LSE Financial Market Group Discussion Paper, n. 204

Schoenmaker D., 1996b, "The Lender of Last Resort and Contagion Risk in Banking", mimeo, LSE Financial Market Group, London

Sharpe S.A., 1990, "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships", *Journal of Finance*, 45, 1069-1087

Shull B. and G.A. Hanweck, 1998, "Banking Competition and Antitrust Policy: Introduction", *The Journal of Reprints for Antitrust Law and Economics*, 27, 3-15

Smith B.D., 1984, "Private Information, Deposit Interest Rates and the 'Stability' of the Banking System", *Journal of Monetary Economics*, 14, 293-317

Smith C.W. and J.B. Warner, 1979, "On Financial Contracting: An Analysis of Bond Covenant", *Journal of Financial Economics*, 7, 117-61

Thornton H., 1802, *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, Hatchard, London

Tirole J., 1994a, "The Internal Organisation of Government", *Oxford Economic Papers*, 46, 1-29

Tirole J., 1994b, "On Banking and Intermediation", *European Economic Review*, 38, 469-487

Von Thadden E.L., 1992, "The Commitment of Finance, Duplicated Monitoring and the Investment Horizon", CEPR Discussion Paper, n. 27

Wallace N., 1988, "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously", *Federal Reserve Bank of Minneapolis Quarterly Review*, 12, 3-16

White L., 1984, *Free Banking in Britain : Theory, Experience and Debate*, 1800 – 1845, Cambridge University Press, Cambridge