Bank Competition and Capital Allocation

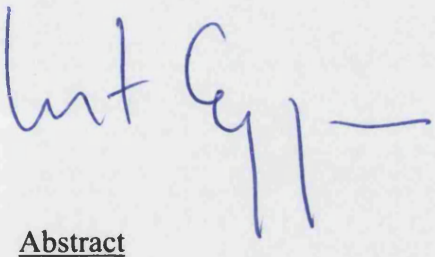
Knut Eggenberger
London School of Economics and Political Science
Submitted for the degree of Doctor of Philosophy (Ph.D.) in Economics

2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration by the Candidate

I hereby confirm that the work presented in the thesis is my own.



Abstract

This thesis consists of eight chapters investigating the relationship between bank competition and capital allocation. Following the introduction (chapter 1), the second chapter provides a review of the literature. The third chapter extends a seminal contribution in the bank competition literature, the model developed by Broecker (1990). In chapter 4 I show that an auction framework may be an inappropriate way of modeling bank competition and I explore the implications of modeling bank competition in a more robust fashion for the allocation of capital. The fifth chapter aims to resolve a long-standing discrepancy between the empirical and theoretical literatures on bank competition and capital allocation. While theoretical work tends to see few benefits from more intense competition, the evidence suggests that the allocation of capital improves as bank competition becomes more intense. The theoretical model developed in chapter 5 reconciles these results by modeling banks' objective function in a way consistent with empirical evidence on X-inefficiency in banks. Chapters 6 and 7 investigate the transmission mechanism through which a greater intensity of competition is transmitted within banks and study lending-related incentive structures through interview-based fieldwork. Chapter 6 provides motivation and outlines the scope of the study whereas the actual findings are presented in chapter 7. That chapter also analyzes the implications for the lending and monitoring decisions that co-determine the allocation of capital. Chapter 8 concludes.

Contents

Acknowledgements

Working towards a Ph.D. may at times be a solitary endeavor, but thankfully it is not one that has to be tackled in isolation. Throughout my graduate studies, I have benefited from the help and support of others in the form of advice, feedback, probing questions and simple (but vital!) encouragement. Teachers, fellow students, researchers elsewhere working on the same issues as well as (a bit unusually for a Ph.D. thesis in economics, I suspect) a large number of banking practitioners have all contributed to making this thesis possible. With one exception I dispense with the usual practice of mentioning people by name: I thank everyone collectively, thus avoiding a diplomatic incident by unfairly forgetting somebody (as, in view of the long gestation period of this project, I am afraid I probably would). Someone to whom I owe more than anyone else and who must be mentioned by name is Margaret Bray who supervised my work with exceptional dedication and patience.

Chapter 1

Introduction and Overview

Throughout the present thesis, the phrase 'bank competition' is shorthand for 'intensity of bank competition'. More specifically, I shall think of the intensity of competition as a variable controlled by public policy (e.g., via the number of banking licenses granted) and I shall be interested in whether there is a degree of competitive intensity that could be said to be optimum in some well-defined sense. The intensity of competition matters in a number of respects; I concentrate on its impact on the allocation of capital. 'Capital allocation', in turn, refers to the extent to which the potential surplus from investing is actually realized. While most of the literature discussed below takes a microtheoretic and partial-equilibrium approach, the deeper questions underlying this research are macroeconomic. A better capital allocation (i.e., more surplus from investing) means, metaphorically speaking, a larger cake to share. Such a measure is a reasonable proxy for economic growth and welfare.

I concentrate on the effect of bank competition on banks' role in overcoming information problems in lending; this implies a focus on small and medium-sized enterprises which do not have access to alternative sources of external finance (except trade credit). There is a small theoretical literature that questions whether banks' lending can be considered independently of their funding [see Hellwig (1991), p. 48, for a discussion; I briefly discuss the main contributions in chapter 2]. Most of the literature considers competition for loans in isolation. I follow the literature and study the effect of competition on capital allocation for a given supply of capital (typically modeled as infinitely elastic at some exogenous interest rate).

In the present section I provide an outline of the thesis and relate my work to the literature reviewed in chapter 2. A more detailed discussion of how this work fits into the broader financial intermediation and Finance & Growth literature is provided in Eggenberger (2006b).

The contribution by Broecker (1990) holds a special place in the bank competition literature. In chapter 3 I extend Broecker's model and show that contrary to Broecker's conjecture, his bank competition game does not have an equilibrium in pure strategies even when banks are asymmetric. In chapter 4 I move away from the auction framework used by Broecker by allowing applicants to renegotiate their offers. I show that the modified game has an equilibrium in pure strategies where for a given number of banks an applicant's interest rate decreases with the number of offers obtained. Surplus unambiguously increases with entry as long as additional banks have an incentive to enter, so the allocation of capital improves as competition gets more intense. The intuition is that when testing is costless each additional bank will test all applicants and aggregate information improves with entry. I also argue that all test results will effectively become public. [In principle, information spillovers lead to free-riding on others' information acquisition. However, this is not a concern when information is obtained costlessly.] These observations are relevant for the subject of this thesis in that they demonstrate that results obtained by modeling bank competition as an auction are critically non-robust to a more realistic specification (more realistic at least as far as lending to small- and medium-sized businesses is concerned). The implication for banking theory is that modeling creditworthiness tests and inter-bank competition should be robust to the empirically highly relevant phenomenon of test results becoming available to competitors.

In the literature review, I show that the theoretical literature on bank competition is pessimistic with respect to the effect of more intense bank competition on the allocation of capital whereas the empirical literature finds substantial gains. I argue that while the two literatures are not asking exactly the same questions, there is enough overlap to claim that, painting with a broad brush, these results contradict each other. While in chapter 4 I question the modeling of bank competition as an auction, in chapter 5 I offer a possible resolution for the discrepancy between empirical and theoretical results. I develop the argument that the main role of more intense competition may be that of a discipline device; a point hitherto overlooked in theoretical work. When other corporate governance mechanisms fail for whatever reason, a major benefit of more intense competition may be that it forces managers to work harder and thus leads to improvements in the allocation of capital. The existing theoretical work cannot capture this effect, because it assumes that banks maximize profit. Evidence in support of my conjecture is provided in section 2.7 on X-inefficiency in banks and section 2.2 which reviews the

empirical work on competition and managerial effort. In chapter 5 I formalize this idea and demonstrate that in the presence of an effort-minimizing bank, the allocation of capital unambiguously improves as competition becomes more intense. In building the model, I also develop a novel way of modeling bank competition.

Chapters 6 and 7 deal with the transmission mechanism that links the intensity of bank competition and the allocation of capital by investigating the subject of lending-related incentive arrangements. To understand the effects of more intense bank competition, one needs to understand how competitive pressures are transmitted within real-world banks. The existing theoretical literature abstracts from this problem by treating banks as black boxes represented by a single manager and only looks at incentive problems at the institution level, so the work reported in chapters 6 and 7 can be read as a complement to the theoretical work in the preceding chapters. Given the amount of theoretical work on, for example, incentives and bank regulation, it is astonishing how little effort has been devoted to studying actual lending-related incentive arrangements. Therefore, as far as the real-world applicability of theoretical work is concerned, one crucial question is how the pressures banks are facing at the institution level are transmitted throughout the organization. As an economist, one is trained to analyze incentives as a key element of this transmission mechanism.

The interview-based fieldwork I present in chapters 6 and 7 addresses the question 'What scope is there for devising effective lending-related incentive arrangements in practice?' Any such schemes or restrictions on such schemes ought to be of interest to a banking regulator, so I adopt the perspective of a regulator looking for sensible performance measures on which such incentive arrangements could be based. In chapter 6 I motivate my fieldwork and outline the scope of the study whereas the actual findings are described in chapter 7. I discuss a number of candidate objectives in detail and present obstacles and potential problems discovered in the interviews. My findings indicate that it may be possible to construct sensible schemes, although serious doubts remain in particular with respect to the supporting risk management infrastructure.

Chapter 2

Bank Competition and Capital Allocation: A Survey of the Literature

2.1 Introduction

In the present chapter, I shall attempt to summarize what the existing literature has to say about the optimum intensity of bank competition and I shall also point out gaps in the literature which I attempt to fill with my own work. In the remainder of this introduction, I give an overview.

While a belief that banks are 'special' is widely held among banking theorists, it seems reasonable to ask what can be learnt about the impact of more intense product market competition in general before looking at the specifics of bank competition. The theoretical and empirical literature linking competition and managerial effort is central to my argument indeed and is reviewed in section 2.2. In addition, I have surveyed contributions which are not of critical importance in the context of this thesis in a separate note [Eggenberger (2006a)]. This note can be summarized as follows: I review the literature on product market competition outside the banking industry and examine the causal chain linking policy measures (e.g., entry liberalization) with certain outcomes of interest (e.g., productivity growth) via changes in the intensity of competition (measured, for example, by concentration indicators) as well as the underlying mechanisms (e.g., competition affecting incentives for innovation). I discuss each of these elements in turn, paying particular attention to measurement problems. I then look at empirical work. What one makes of the evidence inevitably is a matter of interpretation and judgement; given this qualification, the evidence suggests that more intense competition

is beneficial. [In section 2.2 below I show that, empirically, competition in particular has a favorable effect on managerial effort.] The note concludes that the findings of the general product market competition literature are consistent with those for the banking industry specifically (see below).

Section 2.3 provides an overview of surveys similar to the present one. Section 2.4 is at the heart of the review and surveys the theoretical literature on bank competition and capital allocation proper, considering issues such as risk assessment and monitoring. Section 2.5 summarizes and critically reviews 'lessons learnt' from the theoretical literature. The bottom line is that there are few clear-cut answers and that, if anything, the theoretical literature tends to be suspicious of competition.

In section 2.6 I look at the evidence on the effects of more intense bank competition, the findings of which stand in contrast to those of the theoretical literature: Empirical work by and large tends to find more competition to be rather beneficial. One possible way of reconciling these results (which I have pursued further in my own work) is to note that agency problems are quite possibly particularly severe in the banking industry and that more intense competition might reduce these problems and the resulting inefficiencies. However, agency problems have not been paid much attention by the existing theoretical literature. By way of motivation, in section 2.7 I present compelling evidence of X-inefficiency in the banking industry.

2.2 Product Market Competition and Managerial Effort

One important transmission mechanism linking more intense competition (brought about, for example, by entry liberalization) with outcomes of interest is thought to operate via managerial effort. Starting point for the discussion of this mechanism are the uncontroversial observations that 1) in many firms the managers are not the owners and that 2) managerial effort may not be directly observable. The original contributions reviewed here view the degree of managerial effort as corresponding to 'X-efficiency' [Leibenstein (1966); roughly, producing in a way that is cost-minimizing]: Higher effort leads to lower cost. The 'manager' represents the entire organization. (In particular, this means that incentive and information problems within the organization are assumed not to matter.)

Vickers (1995) conjectures that

'competitive pressure makes organizations internally more efficient by sharpening incentives to avoid sloth and slack' (p. 1).

The principal (and, at least for the layman, highly intuitive) idea is that product market competition disciplines the manager in a principal-agent-setup by, for example, preventing him or her from shirking and empire-building at the expense of shareholders and by forcing him or her to exercise effort[1]. Hicks (1935) described managerial slack with the memorable sentence

'The best of all monopoly profits is a quiet life' (p. 8),

although he did not talk about a principal-agent problem. [Hicks conjectured that

'It seems not at all unlikely that people in monopolistic positions will very often be people with sharply rising subjective costs; (...)'

and that monopolists would therefore not bother to find out by how much exactly they need to restrict output to achieve maximum profit.] In a well-known paper, Hayek (1945) noted:

'How easy it is for an inefficient manager to dissipate the differentials on which profitability rests, and that it is possible, with the same technical facilities, to produce with a great variety of costs, are among the commonplaces of business experience which do not seem to be equally familiar in the study of the economist.' (p. 523)

Three decades after Hicks had coined the 'quiet life' catchphrase, Leibenstein (1966) introduced the term 'X-inefficiency' which, roughly speaking, corresponds to the concept of 'slack':

'The essence of the argument is that microeconomic theory focuses on allocative efficiency to the exclusion of other types of efficiencies that, in fact, are much more significant in many instances.' (p. 392)

Leibenstein explains:

'The simple fact is that neither individuals nor firms work as hard, nor do they search for information as effectively, as they could.' (p. 407)

I review the theoretical literature on competition and effort in greater detail below.

How can more intense product market competition help in resolving the agency problem? On one level, it should not matter. This is the stance taken by Jensen and Meckling (1976) who note that

'It is frequently argued that the existence of competition in product (and factor) markets will constrain the behavior of managers to idealized

[1] Allen and Gale (2000b) point out that the exact nature of management failure might not be irrelevant (see p. 60 – 61). In other words, 'effort' may be a poor proxy for other kinds of failure that one might care about: Risk shifting, excessive conservatism, overconfidence, innate differences in ability independent of effort, private benefits of control. I will follow the literature in concentrating on effort.

value maximization, (...). Our analysis does not support this hypothesis. The owners of a firm with monopoly power have the same incentives to limit divergences of the manager from value maximization (i.e., the ability to increase their wealth) as do the owners of competitive firms.' (p. 329)[2] Drawing on the original insight in Holmstrom (1979) that

'essentially any imperfect information about actions or states of nature [FN 28 omitted] can be used to improve contracts' (p. 89),

Holmstrom and Tirole (1989) take issue with the view of Jensen and Meckling and observe that

'(...) competitive markets provide a richer information base on which to write contracts. The value of competition is obvious if one imagines explicit incentive schemes in which the manager is compared with other firms in his market. We know that relative evaluations will allow some reduction in the uncontrollable risk that the manager has to bear and this will reduce agency costs. [FN 30 omitted]' (p. 96)

Prominent contributions building on this mechanism are Holmstrom (1982); Lazear and Rosen (1981) who pioneered the idea of tournaments in the context of optimum compensation schemes; Mookherjee (1984); Nalebuff and Stiglitz (1983a) who provide a discussion of a number of related issues; Nalebuff and Stiglitz (1983b); and Shleifer (1985) who develops the idea of yardstick competition that has been influential in utility regulation. Vickers (1995) provides a discussion of competition facilitating performance comparisons in general (pp. 7-12) as well as an illustration of the value of observing the performance of a second agent in a simple model (pp. 8-10).

Would the 'competition facilitates performance comparisons' argument carry over into a context in which explicit incentive schemes are not used? Holmstrom and Tirole (1989) argue informally that information about competitors' performance would also be valuable in a model à la Holmstrom (1999) which does not feature explicit incentive contracts (the incentive to exert effort is provided by a concern for reputation):

[2]The authors also assert that

'(...) competition in the market for managers will generally make it unnecessary for the owners to share rents with the manager. The owners of a monopoly firm need only pay the supply price for a manager.' (p. 329)

One might, somewhat facetiously, ask whether that is how Harvard Business School, where Michael Jensen teaches, hires professors – by setting up an auction and awarding the job to the applicant offering to work for the lowest salary. (The analogy with a monopoly may not be perfect, but I would think that HBS has considerable market power allowing it to generate substantial rents.) More seriously, what Jensen and Meckling seem to assert is 1) that the labor market for managers is competitive and 2) that the degree of competitiveness of the labor market is independent of that of a firm's output market. These are two separate and interesting ideas that one should be able to test empirically. However, I will not try to pick holes into these claims here.

'(...) a sharper signal about performance will automatically lead to an increased level of effort in equilibrium (since effort responds positively to the signal-to-noise ratio). Observing competitors' performance is one way in which signal strength is increased.' (p. 96)

Vickers (1995) formalizes this intuition (see p. 10 –11) and adds precision as to the exact conditions under which competition leads to higher effort.

The contributions discussed next do not dispute the claim that competition may increase the informativeness of performance signals. Instead, taking that channel for granted, they ask whether there are additional effects through which the intensity of competition affects effort. Note that the contributions reviewed in the following continue to assume that some kind of incentive scheme is feasible. The effect of competition is then mediated by the incentive contract. The first model in this line of research is Hart (1983). Hart restricts permissible contractual forms in such a way as to rule out the effect of competition outlined above: Payments to the manager may only be conditioned on that manager's own profits (so the profits of other firms cannot be used and neither can the input or output price, knowledge of which, in my understanding, would be sufficient to attain first-best here). Competition does indeed increase effort in Hart's model. The mechanism is the following: Hart assumes the existence of two types of firms, entrepreneurial (implicitly owner-operated) ones that maximize profit and managerial ones operated by a manager paid according to an incentive contract (with the information base restricted as outlined above). A greater share of entrepreneurial firms could plausibly be interpreted as an increase in the intensity of competition. All firms need a common input, the price of which is stochastic (and unobservable to principals), affects both total and marginal costs and could in principle differ across firms, although Hart assumes perfect correlation of firms' input prices for simplicity. Managerial effort reduces costs and increases profit; indeed, the input (price) and effort are perfect substitutes by assumption. The product market is competitive: The good produced is homogeneous and there is a single market price that no firm can influence. Crucially, Hart assumes that the managers are infinitely risk-averse and that all they care about is a subsistence level of income: Any less would be catastrophic, any more provides no additional utility. The optimal incentive contract then boils down to the owners stipulating a (low enough) profit target such that

'there is no excuse for the firm's profit ever to be below [that profit level]' (p. 371)

The manager's decision rule upon observing the realization of the input price is simple: Exert just enough effort to attain the profit target. When the input price is high, the manager has to work hard, but when it is low, the manager can afford to slack off. However, since the entrepreneurial firms expand their output when the input price is low, the output price falls and the manager has to work harder to meet the target than if only that particular firm's input price were low. The link between effort and the intensity of competition thus comes about because this slack-limiting effect is the more pronounced the greater the share of entrepreneurial firms. Hart himself wonders whether this result would be robust to a more general specification of the manager's preferences:

'We have studied the extreme case where salary incentive schemes are minimally effective in controlling managers. It will be interesting to see whether our main result – that competition reduces managerial slack – continues to hold under more general conditions.' (p. 381)

As Scharfstein (1988) demonstrated, it does not. In fact, in Scharfstein's model competition worsens slack. Interestingly, Scharfstein does not need to relax the assumption of infinite risk aversion (although it is not required either for the results in Scharfstein's model to go through; all that is needed is risk aversion), but merely allows the manager's utility from income to linearly increase in income. (Utility from income and effort are both arguments of the overall utility function.) The result here – in a reversal of the mechanism in Hart's model – is that when the input price is low the manager works hard (i.e., there is no inefficiency), but slacks off when managerial productivity is low. Scharfstein does not mention these terms, but one way to think about the manager's behavior is in terms of income and substitution effects: Both leisure (non-effort) and consumption are desirable goods and it is not, in general, possible to predict how consumption of each of these goods will change as their relative price changes. Hermalin (1992) makes this idea quite explicit (labeling non-effort 'agency goods') and, apart from the income effect, considers three supplementary channels linking competition and effort[3]. None of the four can be signed unambiguously, although Hermalin describes sufficient conditions such that more intense competition makes a

[3]In addition to competition allowing the principal to make more accurate inferences about the actions chosen by the agent (as in Hart's model), Hermalin distinguishes a 'risk-adjustment effect' and a 'change-in-the-relative-value-of-actions-effect'. Hermalin defines these effects as relating 'to the possibility that competition changes the executive's choice of action – consumption of agency goods – by changing the riskiness inherent in implementing different actions' (p. 351) and 'the possibility that competition can change the difference in expected profit between two actions, so the relative profitability of the better action (e.g., consuming fewer agency goods) can be "enhanced" or "diminished" by an increase in competition' (p. 351), respectively.

manager slack less (consume fewer agency goods). Hermalin also shows that under certain circumstances, the effect of more intense competition can be critically dependent on whether the owner or the manager has all the bargaining power, i.e., gets to make a take-it-or-leave-it offer (intermediate degrees of bargaining power are not considered). Schmidt (1997) explores an additional channel which he terms "threat-of-liquidation" effect: As competition becomes more intense, all else equal profits are lower so the manager has to exert more effort to avoid liquidation (which would lead to the loss of private benefits). At the same time, a "value-of-a-cost-reduction" effect may work in the other direction. The latter effect only arises, however, if the manager obtains a rent. The basic idea underlying this second channel is that with more intense competition, it may not be worthwhile for the principal to induce high effort (even though the cost of doing so has gone down). Schmidt describes what he calls 'natural circumstances' in which the "value-of-a-cost-reduction" effect is positive so that effort increases even when the participation constraint is not binding. However, I believe that these should be considered illustrations and special cases; as Schmidt himself points out (and illustrates with an example), the effect of competition on effort need not be monotonic and is, in general, ambiguous. An effect similar to the "value-of-a-cost-reduction" effect is investigated by Willig (1987). He studies the case of more intense competition translating into greater sensitivity of profit to managerial effort. His main result is ambiguous: If the effect of competition is to reduce profit, X-inefficiency actually increases whereas the converse is true if profit is not affected by more intense competition (e.g., because the principal effect is to increase the elasticity of demand for the firm's output).

Allen and Gale (2000b) criticize reducing the manager's role to that of cost reduction. They also have misgivings about focusing on effort, noting that deviations from profit maximization may stem from different sources (see footnote above) and that competition may affect these differently. Output markets in all the models discussed here are competitive; Allen and Gale note that in oligopolistic markets competition may operate in different ways (p. 60-61). Bolton and Dewatripont (2005) also provide a brief discussion of the link between competition and managerial effort which, however, does not go beyond the work reviewed above (see pp. 636-641, 643-644).

Note that, as observed above, all the contributions reviewed here[4]still

[4]I do not provide an exhaustive review of the literature; additional contributions of interest are Martin (1993); Bertoletti and Poletti (1996) who clarify some fine points on Martin's paper; Graziano and Parigi (1998) who extend Martin's homogenous

assume that some kind of explicit incentive scheme can be used [although it may not be essential that incentives be explicit; implicit incentives (via reputation) might work]. A number of papers consider the case where incentive arrangements are not feasible at all. For example, Hart and Moore (1995) explain that

'To simplify, we consider the (admittedly) extreme case in which the empire-building motive is so strong that no feasible financial incentive payment can persuade the manager not to invest (...).' (p. 571)

However, Hart and Moore are not interested in the disciplinary role of product market competition, but in that of different types of debt contracts. Aghion, Dewatripont and Rey (1997) explore the importance of agency considerations for the role of product market competition in an endogenous growth model with innovation. Their 'satisficing' managers (which they label 'conservative') incur private costs from innovating and are motivated not by profit-maximization, but purely by a desire to hang on to private benefits of control at the least cost of innovation. With this preference structure, the effect of competition on growth is positive. A later, more nuanced model investigates the simultaneous operation of financial market discipline (via debt) as a discipline device [Aghion, Dewatripont and Rey (1999)][5].

I end the discussion of the theoretical literature here. Clearly, this work does not exactly constitute a ringing endorsement of the common-sense view that competition is an effective discipline device. All in all, from a theorist's perspective the claim that more intense product market competition reduces managerial slack is not terribly robust. Only the evidence can tell whether common sense is leading us astray here or not.

Despite the importance of the subject from both an academic and a policy point of view, attempts to investigate the link between the intensity of competition and 'slack' rigorously are fairly recent. One line of work uses frontier production function techniques to estimate technical efficiency. Prominent book-length contributions in this literature are Caves and Barton (1990), looking at U.S. evidence, as well as Caves and Associates (1992) which extends the scope of the analysis to an additional five countries (Australia, Canada, Japan, Korea, UK). Broadly speak-

Cournot model to differentiated Cournot competition; Sjöström and Weitzman (1996) which is more of a statistics paper; Krishna (2001); Horn, Lang and Lundgren (1994); Horn, Lang and Lundgren (1995), Stennek (2000), and Raith (2003).

[5]These papers complement work by Aghion and Howitt (1996) as well as Aghion et al. (2001) showing that different assumptions with respect to the innovation technology can establish a beneficial effect of competition on innovation, thus reversing the "Schumpeterian" effect that makes innovation dependent upon market power.

ing, these authors find competition to be good for efficiency, although it should be noted that their work suffers from some serious methodological shortcomings [see Tybout (1992) and Comanor (1994)]. Green and Mayes (1991) has grown out of the same research effort and discusses the methodology and some of the problems. Hay and Liu (1997) also make use of frontier production function techniques and investigate a dataset of 19 UK manufacturing industries at the industry level. They cautiously summarize their results as showing that

'There is evidence of response by firms to short run declines in market share. Short run improvements in efficiency resulted, probably the result of greater managerial effort (...)' (p. 615)

Another line of research considers not econometric estimates of (X-in) efficiency but the impact of competition on the level and growth rate of productivity. The first paper in this literature is Nickell (1996) which uses a panel data set on UK firms. Nickell finds strong evidence that more intense competition [measured by (lagged) market share] increases the level of productivity and that an increase in the number of competitors or lower level of rents lead to stronger productivity growth. Nickell acknowledges that there are problems in the way that competition is measured, but presents plausible counter-arguments in his defense (pp. 732–734). He also draws attention to the possible endogeneity problem, but argues that his approach manages to defuse this potential difficulty (p. 734, p. 739). Griffith and Harrison (2004) are not convinced and – with respect to his article – note that

'potential endogeneity in market share is not dealt with properly, so imputing a causal relationship between competition and TFP growth is problematic.' (p. 15)

Such quibbles aside, the paper is widely cited and seems to have been the most solid evidence available on the reduction in agency costs brought about by greater competition at the time of its publication. A number of subsequent articles have bolstered the case for competition as a remedy to agency problems. Nickell, Nicolitsas and Dryden (1997) use a dataset almost identical to the one used by Nickell (1996) to investigate the link between 1. the intensity of product market competition (as measured by the average level of ex post rents during the sample period normalized on value added), 2. financial market pressure (as measured by interest payments normalized on cashflow), as well as 3. shareholder control (as measured by a fairly complicated but not implausible indicator explained in the paper) and productivity growth. They find that all three measures of managerial discipline devices raise productivity growth (although the

finding that a dominant external shareholder leads to higher productivity growth only if it is from the financial sector, but leads to lower productivity growth if it is not a financial institution is not really explained, see p. 793). There is some evidence in the data that financial pressure and a dominant external shareholder can substitute for product market competition as a discipline device:

'our measure of competition has no positive impact on productivity performance in the presence of a dominant outside shareholder' (p. 793).

Somewhat contrasting results are provided by Januszewski, Köke and Winter (2002) who investigate a panel of German firm-level data for close to 500 firms from the manufacturing sector. The authors claim that they manage to overcome the endogeneity problem by using lagged values as instrumental variables (although they acknowledge that these could be weak instruments). They primarily measure market power with firm-level ex post rents and find that productivity growth is higher in markets with more intense competition. In an interesting contrast to the results in Nickell, Nicolitsas and Dryden (1997), they find that the presence of an ultimate controlling owner (see definition on p. 310) is associated with higher productivity growth only if that ultimate owner is not a financial institution and their results suggest that product market competition and the presence of a controlling owner are complements rather than substitutes. These differences with the earlier work by Nickell et al. are acknowledged. The increase in the intensity of competition in Pavcnik (2002) comes about via trade liberalization. Her analysis is so careful that it supports a causal link running from the intensity of competition to productivity; productivity improvements in her plant-level panel data set of Chilean manufacturing plants are large (in addition to selection effects). Griffith (2001) has panel data on UK establishments and uses the phasing in of the EU's Single Market Program as an instrument to sidestep the endogeneity of measures of the intensity of competition. She manages to show that the impact of this exogenous increase in competition was very different for two groups of firms: Those which could reasonably be expected to be suffering from agency problems (where the increase in competition led to improvements) and those that could reasonably be expected to be entrepreneurial without a major conflict of interest between owners and managers (where the increase in competition did not have that effect). This allows her to argue that the channel through which an increase in the intensity of product market competition leads to an increase in productivity is via a reduction in agency costs. Jagannathan and Srinivasan (1999) and Jagannathan

and Srinivasan (2000) do not measure productivity directly, but rather choose a somewhat roundabout route to argue that product market competition disciplines managers. They claim that firms can be categorized into 'specialists' that operate in rather uncompetitive markets as well as 'generalists' who operate in markets where product market competition is more intense. Since in the former type of firm, managers are assumed to have leeway to use 'free cash-flow', the relationship between current leverage and future profitability is postulated to be positive. The reverse is said to be the case for 'generalist' firms operating in supposedly more competitive markets. The empirical results on a sample of 165 US firms support the authors' contentions.

In summary, these findings tie in directly with the theoretical literature on slack and competition and suggest that, empirically, there is support for the view that more intense competition leads to increased productivity via a reduction in agency costs. In the remainder of this section, I discuss alternatives to product market competition as managerial discipline devices and I discuss some empirical evidence that arguably compares the effectiveness of different devices.

In their survey of corporate governance, Shleifer and Vishny (1997) are sceptical with respect to the scope for incentive contracts to overcome the agency problem between owners and management:

'While it is a mistake to jump (...) to the conclusion that managers do not care about performance at all, it is equally problematic to argue that incentive contracts completely solve the agency problem.' (p. 745)

There are, of course, other mechanisms that mitigate agency problems. Jensen (1993) distinguishes 1. capital markets, 2. the legal, political, and regulatory systems, 3. product and factor markets, and 4. the internal control system (headed by the board of directors) (p. 850). Holmstrom and Tirole (1989) discuss labor market discipline explicitly (p. 94-95). Allen and Gale (2000b) single out the following governance mechanisms as 'most important':

'(i) The board of directors; (ii) executive compensation; (iii) the market for corporate control; (iv) concentrated holdings and monitoring by financial institutions; (v) debt.' (p. 37)

To give one well-known concrete example, Jensen (1986) advocates the use of capital structure (specifically, high leverage) to constrain management. In the later paper, Jensen (1993) argues against relying on product market discipline by pointing out that

'Unfortunately, when product and factor market disciplines take effect it can often be too late to save much of the enterprise.' (p. 850)

Shleifer and Vishny (1997) concede that

'product market competition is probably the most powerful force toward economic efficiency in the world' (p. 738),

but argue that one cannot rely on product market competition alone. Allen and Gale (2000b) effectively take the diametrically opposed position and – in view of divergent governance mechanisms that seem to have very little bearing on the relative success of an organization and the continuing success of non-profit organizations that have little in the way of explicit governance arrangements – argue that the crucial disciplinary instrument is product market competition.

There is some empirical work that attempts to run 'horse-races' between the different mechanisms; the idea being that, in principle, the evidence should allow researchers to determine which devices to mitigate agency problems are the most important. To discuss this literature in depth would go beyond the scope of this review, but I will briefly discuss three articles of interest in addition to the papers by Nickell, Nicolitsas and Dryden (1997) and Januszewski, Köke and Winter (2002).

The first article does not look at the banking industry in particular. Dyck and Zingales (2004) use data on almost 400 control transactions in 39 different jurisdictions to directly measure the 'private benefits of control' and to analyze which factors are most important in reducing these:

'Besides the law, we also consider extra-legal institutions, which have been mentioned in the literature as possible curbs for private benefits: competition, labor pressures, and moral norms.' (p. 539)

In addition, the authors also investigate the influence of the press and the government in its role as corporate tax collector. While their focus is on expropriation by owners with a controlling stake rather than managerial slack, their results are clearly of relevance for the issue of reducing the latter:

'We find that a high level of diffusion of the press, a high rate of tax compliance, and a high degree of product market competition are associated with lower private benefits of control.' (p. 539)

Thus these results provide some additional support for the disciplinary role of product market competition.

In Schranz (1993) the market for corporate control (i.e., the disciplinary role of the threat of takeover) takes center stage. She takes advantage of takeover regulations in the banking industry that differ across U.S. states and finds that

'banks in active takeover markets are more profitable than those in

markets restricting takeovers.' (p. 323).

Firms that operate in markets where takeovers are restricted make more use of alternative mechanisms (such as concentrated ownership):

'However, use of these alternative methods does not completely compensate for the absence of an active takeover market in improving firm performance.' (p. 323)

Schranz controls for a number of alternative discipline devices among which is the competitiveness of the output market. That variable comes out as unimportant but she acknowledges that there may be problems with the way she measures competition using the state-wide Herfindahl index (see p. 307).

Finally, Crespi, Garcia-Cestona and Salas (2004) investigate the Spanish banking industry. Just as in the banking industries of many other countries, various ownership forms co-exist in Spain and their differing governance mechanisms are hypothesized by the authors to differ in the effectiveness with which they constrain management. They find this to be the case, but note that

'[P]roduct-market competition compensates for those weaker internal governance mechanisms [of the non-profit Savings banks]' (p. 2311).

What is there to take away from this section? The first insight, I believe, is that on balance the evidence suggests that product market competition is effective in curbing managerial slack. Such a blanket assessment does not do justice to the numerous qualifications that are called for and must be considered tentative. However, my reading of the literature suggests that the empirical work does indeed lend itself to such an optimistic assessment.

In addition, a second observation can be made: That theoretical work should on the whole (and again simplifying a complex debate somewhat) be much less in favor of product market competition than the evidence (as well as common sense, admittedly a suspicious arbiter) is a phenomenon that appears rather consistently across a number of literatures that this thesis touches on, namely work on the effects of more intense competition on

–financial stability and the failure probability of banks [see Eggenberger (2006b)]

–innovation and productivity [Aghion and Griffith (2005) explicitly point this out; I review the relevant literature in Eggenberger (2006a)]

–managerial effort (see above)

In the remainder of this literature review, I shall demonstrate that the same discrepancy also arises in the differing conclusions of theoretical

work on the one hand and empirical evidence on the other as to the effects of more intense bank competition with respect to the allocation of capital. First, however, I shall briefly provide a 'meta-survey' of other literature reviews on the subject.

2.3 Surveys of the Bank Competition Literature

The purpose of this section is to give an overview of surveys similar to the present one (though sometimes with a different emphasis – e.g., a focus on the competition-stability nexus rather than the link between bank competition and capital allocation). I shall comment briefly on each of them as I go along. Allen and Gale (2004) concentrate on the links between competition and stability and combine a survey of the theoretical literature with some original work. The bottom line of their review is that the presumed tradeoff between competition and stability may, even on theoretical grounds, be far more tenuous than is currently suggested by much of the literature. Allen et al. (2001) is an overview of papers presented at a conference promisingly titled 'Competition Among Banks: Good or Bad?' and not only summarizes the conference contributions, but also puts these into context. Berger et al. (2004) is likewise partly a conference overview, but gives far greater weight to a detailed review of the prior literature and the issues being investigated. As the focus of that survey is almost exclusively on empirical issues, I shall discuss it in more detail in section 2.6. The stated theme of Canoy et al. (2001) is the supposed tradeoff between competition and stability, although the paper surveys a lot of other, only loosely related literature along the way. Section 2.3. on empirical evidence provides a helpful introduction on the effects of market power in banking with respect to efficiency. Overall, the paper is clearly written from a public policy practitioner's perspective and does not focus on theory. (For example, it contains an applied section on the Dutch retail banking industry.) Section 6 on the impact of banks' corporate governance on the relationship between competition and stability is, on the face of it, very interesting from the point of view of my own interests. I believe that this perspective is highly promising, but the discussion in the paper is largely limited to a review of the general corporate governance literature. Surprisingly the authors do not discuss the argument that (bank) managers might actually be more risk-averse than owners so that agency conflicts in banking could be thought to make banks safer than they otherwise would be. Carletti

and Hartmann (2002) also focus on competition and stability rather than the links between competition and some measure of efficiency. In its first part the paper takes an institutional perspective in that it concentrates on

'an analysis of the relative roles of competition and supervisory authorities in the review of bank mergers for the G-7 industrialized countries and the European Union.' (p. 5)

It provides a crisp outline of what competition policy is all about and sketches attitudes toward the desirability of bank competition (or lack thereof) in a number of industrialized countries before giving a description of the relevant institutional arrangements in the G-7 countries (plus the EU institutions which constitute separate actors). In the second part the survey moves on to a discussion of the theoretical literature on competition and stability as well as the relevant empirical evidence. Plausible dividing lines are proposed for the theoretical literature (e.g., effects of competition on the liability side of banks' balance sheets vs. effects on the asset side) and a number of contributions are reviewed in detail. In the assessment of the authors,

'whereas most papers find some trade-off between bank competition and stability, the claim that they are generally negatively related is not necessarily robust.' (p. 24)

In the discussion of the available empirical work, four sub-strands of literature are identified and discussed. The bottom line of that part of the survey is that

'there does not appear to be a single ever-valid relationship between competition and stability in the banking system.' (p. 29)

Cetorelli (2001) begins his survey by setting up a strawman and claiming that

'[t]he common wisdom would hold that restraining competitive forces should unequivocally produce welfare losses.' (p. 38)

That is actually not the common wisdom at all as I shall show in section 2.4. [Vives (2001b), for example, begins his survey with the claim that

'Competition has always been contentious in banking.' (p. 535)]

However, this is not a real point of criticism with respect to this selective, but insightful survey. Cetorelli looks at theoretical work first and outlines a number of important contributions in some detail. In both the theoretical and empirical work considered he focuses on efficiency considerations and ignores the competition-stability nexus. He also investigates a number of (what are from the point of view of the present

survey) relative side issues, such as how government ownership of banks interacts with the effects of competition (Cetorelli outlines some of his empirical work suggesting that a purported positive effect of bank concentration vanishes when the degree of government ownership is high), the interaction of competition with regulatory restrictions on banks, and the links between market structure in the banking industry and the market structure in borrower industries. In summary,

'[t]he main conclusion that seems to emerge from the review of the current literature is that the market structure of the banking industry and the related conduct of banking firms affect the economy in a much more complicated way than through the simple association: more market power equals higher lending rates and lower credit quantities [i.e., via an improvement or a deterioration with respect to allocative efficiency]' (p. 39)

I will discuss Claessens and Klingebiel (2001) in section 2.6 since they concentrate on empirical work. They do not provide an academic literature review but a very broad-brush policy-practitioner discussion which, however, gives a good overview of the issues. Dell'Ariccia (2003) is not a survey so much as a collection of punch lines. He mentions about 30 papers in the space of three pages, but for this very reason is helpful in that the article provides a highly condensed overview of some recent banking literature (though with a broader scope than just 'competition and capital allocation'). Not to be left out of this discussion of surveys is Freixas and Rochet (1997), the standard textbook on banking theory and related topics. The book is very helpful in discussing many different ingredients of bank competition models (monitoring to deal with moral hazard on the part of the borrower, for example), but the literature on bank competition and capital allocation (or bank competition and financial stability for that matter) is too recent to be given systematic coverage in the book. A number of models of bank competition are reviewed therein, but these are not investigating issues such as information acquisition incentives or monitoring intensity specifically. Guzman (2000) discusses a number of papers looking at competition and efficiency in detail. His survey is particularly helpful in that he goes beyond summaries and draws out some of the more general mechanisms at work. His dividing the literature into partial and general equilibrium models and his comments on the implications, for example, are very insightful. Hellwig (1991) looks at financial intermediation by banks (as opposed to markets) more broadly as well as relationship lending in particular and only touches on competition issues (in the treatment of exclusivity in the context of relationship lending).

In the coverage of the material which that paper discusses it is, of course, somewhat superseded by now. However, it is still insightful in the way it sets out the distinguishing differences between various strands of literature (e.g., in a brief discussion of static vs. dynamic models of moral hazard or in a section that critically assesses to what extent it is appropriate to leave out the refinancing side in models of financial intermediation) and in its willingness to really probe whether the theoretical literature (at least as it stood at the time of his writing) is on to something:

'How does our theoretical analysis relate to the stylized facts presented by the economic historians (...)? What clues do we have that theory is in fact telling us the right story?' (p. 37)

For example, Hellwig confronts the predictions of the theoretical models of relationship lending with empirical work and points out that these are not supported by the evidence. Neuberger (1998) concentrates on surveying empirical work on a range of industrial organization issues in banking, a number of which are relevant from the point of view of this survey. Her article is mostly a summary of results, not a discussion of methodological issues, and provides a helpful collection of stylized facts (e.g., on the extent of scale economies in banking). The article by Padoa-Schioppa (2001) was originally delivered as the keynote dinner speech at the above-mentioned conference titled 'Competition among banks: good or bad?' and deals with the subject of bank competition from a senior policy-practitioner perspective. Interestingly, Padoa-Schioppa asserts that technological progress, by making it easier to circumvent regulatory restrictions, is what triggered a more relaxed attitude toward competition on the part of regulators:

'Even though I regret saying this to a predominantly academic audience, ideas tended to follow, rather than anticipate, a change which was largely due to technological factors.' (p. 17)

Parigi (1998) concentrates on theoretical contributions related to competition in the banking industry (including a section on competition and stability). He also covers areas that I do not mention at all such as, for example, network considerations in the operation of ATM networks. Van Damme (1994) surveys micro-theoretical work on banking and intermediation more broadly and is still a useful guide to the earlier literature. The section on competition, however, is brief and outdated. Vives (2001b) provides a survey on bank competition that blends a discussion of theoretical work with that of the evidence. In terms of the literature discussed, Vives' article is very independent in the sense that it is not a collection of article summaries and, indeed, discusses few articles in

detail. I think that this should be seen as a strength of the paper as it helps to ensure readability and gives the reader an excellent overview. A minor drawback is that the broad-brush nature entails confident assertions with respect to issues on which both the empirical and theoretical literature are still somewhat divided. For example, Vives claims that

'Even accepting that competition will be necessarily imperfect, vigorous rivalry seems to be necessary both for static and dynamic efficiency of an industry.' (p. 538)

I believe that Vives is right, but as I argue elsewhere that issue might be thought to deserve a more nuanced assessment [see Eggenberger (2006a)].

I now turn to a discussion of the primary theoretical contributions.

2.4 Bank Competition and Capital Allocation - Theory

2.4.1 Common assumptions and overview

In the present section, I survey the theoretical literature on bank competition and capital allocation. The models reviewed tend to have certain features in common which I outline in the following. They take the existence of financial intermediaries and the use of debt contracts for granted and do not feature aggregate uncertainty (individual bank failures or systemic crises; see above for surveys). The literature has evolved such that stability issues are largely ignored in models that focus on capital allocation and conversely. Competition is typically driven by more intense competition between banks (not between banks and capital markets).

I concentrate on competition for loans, not deposits, although I do briefly review models in which bank competition operates via savings mobilization. More intense competition could conceivably lead to higher deposit interest rates and thus (even if the effect is ambiguous in theory) a higher savings rate. However, the evidence indicates that savings are rather interest-inelastic in practice [Loayza, Schmidt-Hebbel and Servén (2000), Bandiera et al. (2000)]. At any rate, most of the models simply assume that banks can obtain any amount of funds at some exogenous interest rate.

Agents are typically risk-neutral, benefit from limited-liability and an absence of bankruptcy or application costs and, with few exceptions [e.g., Caminal and Matutes (2002)], entrepreneurs' funding requirements are fixed on a per-project basis and identical across projects. The cashflow

characteristics and distribution of entrepreneurs' projects are assumed to be exogenous.

Collateral, equity and other self-selection devices are typically not available in order to keep the action firmly on the side of the banks. In fact, quite a few of the models reviewed here do not even allow loan applicants to know their own type (thereby obtaining an informational advantage relative to lenders)[6]. I should note that it is a matter of deep conviction for some theorists that entrepreneurs must have an informational advantage with respect to their financiers. However, the evidence suggests an informational advantage in the opposite direction: The financier may actually know more than the entrepreneur[7].

Entrepreneurs are assumed to borrow from a single bank and, in most models, are assumed to borrow only once. My discussion of dynamic ('relationship lending') models is limited to contributions explicitly examining the effects of competition. Empirically, a dynamic setup is maybe more appropriate as far as small- and medium-sized business borrowers are concerned (see below). However, the one-shot setup is by far the more prevalent modeling approach.

The literature largely avoids discussing the effects of competition on allocative efficiency. As far as the real-world problem of maximizing surplus from investing is concerned, the evidence suggests that such pure market power effects are of considerable importance. However, they are conceptually uninteresting and much of the literature has implicitly assumed them away to focus on informational issues instead. One way of removing the pure market power effect from models of bank competition is to keep demand for loans completely interest-inelastic and the supply

[6]Park, Brandt and Giles (2003) and Villas-Boas and Schmidt-Mohr (1999) – reviewed below – are exceptions; in the first model applicants know their success probability and, in the presence of application costs, have to trade off between borrowing from a lower-interest rate bank with some probability or obtaining a loan from a higher-price bank with certainty; in the second model, sorting can be induced via the use of collateral. Since I follow much of the literature in keeping all the action on the side of the banks, I should note that there are a number of general banking theory contributions that feature sorting, in particular in connection with the use of collateral [e.g., Bester (1985), Besanko and Thakor (1987)]. Freixas and Rochet (1997) also discuss this material in their textbook. A general discussion of collateral is provided by Coco (2000).

[7]Manove, Padilla and Pagano (2001) give references to the literature and memorably note that 'Economists of a certain stripe seem to have difficulty accepting this obvious point. For example, an anonymous referee of a previous version of this article wrote: "The idea that a bank has a better idea of the success of a project than the entrepreneur who dreamed up the project seems so off that I find the rest of the paper uninteresting." An anonymous colleague of ours has a similar point of view: "The idea that a referee is more able to evaluate the quality of a paper than the economist who wrote it", he says, "seems so off that I find the contents of their reports completely uninteresting."' [see FN3, p. 727]

of deposits perfectly interest-elastic[8]. With few exceptions [e.g., Gehrig (1998a) and Cetorelli (1997) where loan demand is elastic] that is the modeling approach pursued by the contributions reviewed below.

Not all models have all the features outlined here, but the preceding description strikes me as a valid general outline of common assumptions. If the papers in this strand of research really have as much in common as I claim, where do the differences stem from? I see three main sources of differentiation. Assumptions with respect to cashflow distributions on both the project and aggregate levels are the first element. These assumptions give rise to different kinds of information problems, they determine which (screening/monitoring) technologies make sense, and they underlie the variations in surplus that different market structures generate. Second, technologies for reducing information problems vary widely and entail a number of additional assumptions (e.g., whether signals are privately observed or become public knowledge and how accuracy and cost are parameterized). Third, there is some variation in how banks are assumed to compete (e.g., Bertrand competition vs horizontal differentiation) and how the intensity of competition is parameterized.

From the standpoint of this thesis, the most interesting feature of lenders is their role in overcoming information problems. Roughly speaking, these information problems can be grouped into three categories according to the stage of the lending process at which they arise [see Freixas and Rochet (1997), p. 29]. Ex ante informational asymmetries between borrower and lender lead to adverse selection and can be remedied by the bank's screening of projects. Banks assess applicants' creditworthiness and refuse to give entrepreneurs assessed as bad risks credit. At the interim stage, i.e., after a loan has been disbursed, informational asymmetries may give rise to moral hazard on the part of the entrepreneur. This moral hazard is typically modeled as a non-verifiable choice of effort level or production technology. The bank can help to mitigate the moral hazard problem by monitoring the borrower. Third, at the ex post stage, the borrower's cashflows might not be verifiable. In this so-called 'costly state verification' framework, the bank may be able to alleviate the problem by 'auditing' or 'punishing' a borrower who does not repay[9]. A (loose) real-world interpretation of this might point to

[8] Note that the effects of inefficient cross-subsidization between different risk classes of borrowers (good risks being charged 'too high' a rate and bad risks being charged 'too low' a rate) differ from those of the conscious exercise of market power. Cross-subsidization was arguably a first-order real-world concern until more risk-sensitive credit risk measurement and pricing tools were introduced by banks during the last ten years or so.

[9] Note that there is no agreement on the exact meaning of the term 'monitoring'. For example, Hellwig (1991) uses the term 'monitoring' to denote the bank's actions

banks' often central role in bankruptcy proceedings and in recycling the capital of failed firms. Fourth, in addition to these three information problems, banks may also have an important role in assisting borrowers who experience a negative liquidity shock, e.g. by renegotiating the loan contract or providing emergency liquidity assistance.
Although all four aspects ought to matter in real-world terms, they do not have easily captured equivalents available in datasets for empirical work – a point I shall return to when discussing the empirical evidence. My emphasis on the first aspect, screening, is justified by the relative lack of literature on the others: For example, I am aware of only one paper dealing with monitoring and competition [Caminal and Matutes (2002)]. With the exception of Guzman (2000b) and Smith (1998), to the best of my knowledge there are no publications specifically investigating competition in the context of 'auditing'. As for emergency liquidity provision and renegotiation, one would expect this to be given some attention in the literature on competition and relationship lending; however, actual formal models of this aspect are few indeed [I am only aware of Dinç (2000)].

In the following, three sub-sections deal with models investigating screening, monitoring, and liquidity assistance, respectively. Interestingly, to the best of my knowledge there is no contribution that investigates these functions simultaneously in a bank competition setting. For example, most screening papers keep the cashflow distribution invariant and thus do not allow for a moral hazard problem on the part of the borrower in the first place [Schnitzer (1999a) and Schnitzer (1999b) are exceptions]; in the absence of this feature there clearly is no point in modeling banks' monitoring[10].
Both of the 'auditing' models mentioned are built around an overlapping-generations setup; a class of models I discuss in a separate sub-section. Related contributions (which do not directly address the issue of bank competition and capital allocation) are discussed next. Finally, I will briefly give some real-world illustrations.

2.4.2 Screening

I first consider screening models in which the quality of banks' information is exogenously given and screening is typically nothing more sophis-

in all three of the settings outlined above (p. 46). Freixas and Rochet (1997) follow Hellwig, but also use the terms 'screening' and 'punishing'/'auditing' in the way in which they are used here (p. 29).

[10]There are a number of articles featuring, for example, both screening and monitoring in the broader theoretical banking literature; see, e.g., Thakor (1996).

ticated than a costless test with fixed error probabilities. I then discuss models with endogenous information acquisition.

In Broecker (1990), banks receive costless imperfect binary signals on applicants' type (which can be either good or bad). With banks engaging in pure price competition, the key result in the context of this literature review[11] is the non-existence of an equilibrium in pure strategies. Broecker shows the existence of a unique and symmetric mixed-strategy equilibrium, derives the distribution function according to which banks set interest rates and obtains an expression for equilibrium profits. Banks have an incentive to underbid their rivals in order to achieve a better composition of their borrower pool – alternatively, i.e., when they cannot profitably underbid, they may find it optimal to ask for the highest possible rate. The intuition behind this result is that an applicant accepting to pay a high interest rate must have been rejected by all the banks setting lower interest rates and must therefore on average be a very bad risk. In other words, banks' rejection decisions create externalities for other banks. Broecker shows that the average creditworthiness (or success probability) must be decreasing in the number of banks that are active; conversely, the expected interest rate is increasing in the number of banks. It is presumably this result that leads to Broecker's paper being cited as showing that more intense competition between banks is bad for the allocation of capital. In fact, whether this highly stylized model warrants such an interpretation with respect to real-world banks is open to question (see below). Broecker's model has a more than formal similarity with a first-price, sealed-bid, common value auction model and it is therefore not surprising that the externalities generated by banks' rejection decisions are similar to the winner's curse. I provide a slightly more detailed discussion of this important paper in the third chapter.

The contribution by Riordan (1993) makes the auction setting very explicit and, unlike Broecker (1990), directly addresses the question

'Is more competition in loan markets a good thing?' (p. 329)

Riordan's model is basically a simplified version of the general auction model developed by Milgrom and Weber (1982). Once again, projects come in 'good' and 'bad' varieties. Given the auction setup, there is a single borrower and n banks who bid for the right to make a loan with a principal of one unit of capital. The bids are the repayments that the borrower would have to promise the bank. Unlike in Broecker's model where both types and signals are binary, the signals here are continuous (and assumed to satisfy the monotone likelihood ratio property). Conditional

[11]The paper has a second part that is largely of interest to game-theoreticians.

on type, the signals observed by the banks are independent. Riordan obtains a condition that guarantees a unique interior cut-off level of the signal such that any bank observing a worse signal will not make a bid. It is that cut-off level that determines how likely it is that a loan will be offered. Riordan derives the unique, symmetric bidding function and shows that the equilibrium cut-off is increasing in the number of banks – essentially due to the winner's curse. He also obtains the probabilities that type I errors (a good type does not obtain a loan) and type II errors (a bad type obtains a loan) are made, but the effect of an increase in the number of banks on these error probabilities is ambiguous. This implies that there is no clear-cut answer to Riordan's question. In sum, there is no definite result even within the context of the model as to whether more intense bank competition (proxied by the number of banks that are potentially bidding) would improve or worsen the allocation of capital. Still, Riordan is sceptical as to the benefits of more intense competition and notes:

'The model suggests two reasons why more competition in loan markets might damage market performance. First, the statistic the market uses to select loans may become less informative about the quality of the loan. Second, the loan approval practices of individual banks may be too conservative, and more competition may make them even more so.' (p. 330)

Nakamura (1996) provides another variation on the auction theme. This model features the standard 'worsening pool' feature; as rejected applicants approach more and more banks, the average quality of the remaining applicants (i.e., those who have not yet managed to obtain a loan) worsens. Signals are assumed to follow the Normal and any lending that happens takes place at an exogenously fixed interest rate. (Nakamura also looks at the case where the prospective borrower gets to set the loan terms; a case that I do not consider here.) Three market structures capturing the extent of competition are compared: 1. Each borrower only approaches a single lender (captured by assuming that the lender knows that she is the only lender approached). 2. Borrowers can approach multiple lenders, but all the test results (rejections) that they have obtained are common knowledge (this is called a 'hierarchy'). 3. A market labeled 'anonymous' in which borrowers can approach multiple lenders without their previous test results being known. The signal that the potential lender observes is costless and the accuracy (distribution) of the signal exogenously given. The relationship between the degree of competition and surplus is ambiguous.

The whole point of an auction is a commitment on the part of all players not to renegotiate. Also, the three models sketched out here assume that price is all that matters. So how appropriate is this as a description of bank competition? It depends. A segment of the loan market, namely that of lending to large, established firms, really does function along similar lines (including the one-shot, arm's length, auction setup). However, the screening role of banks in this market segment is of marginal importance – borrowers in this segment are relatively transparent with a credit rating from the big agencies and probably followed by dedicated equity and credit analysts. For the kinds of borrowers that the banking literature narrowly defined is concerned with, there is solid evidence suggesting that price is not all that matters (see section 2.6). In particular, rivals' (price) offers should be thought of as common knowledge and, as I show in the companion chapter titled 'On Broecker's "Credit-worthiness tests and interbank competition"', the results on the worsening of the applicant pool as the number of banks increases are critically non-robust to this modification. Why is the modified setup more plausible than the standard arrangement? Tests are assumed costless and good applicants at least would find it worthwhile to get tested by all banks. They then have an incentive to show all their favorable test results (showing them to be, on average, of better quality) to negotiate better loan terms; then by implication it is clear that any applicant unable to show good test results must be of lesser quality (in expectation). So this modification should emerge endogenously, unless loan offers were non-verifiable (which seems to be at odds with contemporary banking practice in industrialized countries, where loan offers by and large are made in writing). Shaffer (1998) notes a similar effect and points out that a greater number of banks does not worsen the applicant pool if banks' signals are perfectly correlated.

I next review three papers that do not use an auction setup. In the first two information is arguably endogenous; however, the testing decision is so coarse that it seemed more appropriate to review these articles along with contributions in which accuracy is exogenous. Chiesa (1998) tackles the central topic of this survey head-on:

'Does the removal of intra-state entry barriers increase welfare?' (p. 409) 'Does a concentrated banking industry dominate a fragmented one? This is the question this paper seeks to address.' (p. 410)

Her model 'almost' features aggregate uncertainty: In the bad state of the world, the bank decumulates capital because of losses on bad projects that fail in that state of the world (whereas good projects are always profitable), but in equilibrium the bank will not become insolvent. Outside

investors – the bank's depositors – will see to that in that they limit the amount of funds they provide the bank with. (The article assumes that banks' capital is not sufficient to finance all applicants with a favorable test result.) Only if the bank has enough at stake can outside investors be certain that the bank will in fact screen the applicants and only lend to those assessed as good (which will include some bad types falsely assigned to this category by the imperfect test) which in turn will ensure that the bank is solvent even in the bad state of the world. If the bank cannot fail, then it has every incentive to incur the cost of screening borrowers. Screening accuracy in that model is not exogenous; in fact, banks decide whether to apply their test but will always screen applicants in equilibrium and only lend to entrepreneurs with a favorable test result. However, it is assumed that the type I/II error probabilities are exogenously given. Chiesa compares a monopoly with a two-bank scenario. The result: Less lending and a worse composition of the borrower pool under duopoly relative to monopoly. I have reservations as to the interpretability of this result, as the model setup is such that the banking market comes rather close to being a natural monopoly: The efficient outcome is to test an applicant exactly once. In particular, since there is no moral-hazard problem on the part of the borrower, even a complete expropriation of the entrepreneur does not affect surplus.

The article by Park, Brandt and Giles (2003) is very original in that it manages to relate its theoretical analysis to empirical work on bank competition in rural China. The theoretical setup considers a monopoly and a duopoly; an unusual feature of the model is that the incumbent (monopolist) is assumed to have to charge an administratively set interest rate below the market-clearing rate. This assumption accords well with the situation in China (and, as the authors point out, many developing countries in general). The resulting credit rationing is therefore not of the equilibrium [Stiglitz and Weiss (1981)] type. There is no moral hazard on the part of the bank nor on the part of the borrower. Unlike in many other papers in the literature, which assume that the bank can obtain any amount of funds at a given interest rate, the bank is here constrained in its lending volume by the amount of deposits at its disposal (assumed exogenous). Again, the motivation for this feature is the situation in rural China where access to an interbank market is strictly limited. The information problem centers on adverse selection where borrowers differ in the success probabilities of their projects. The success probabilities of the aggregate borrower pool are assumed to be distributed according to some general distribution function (no specific functional form, although

the Uniform is briefly investigated later in the paper). A monopolist that does not screen borrowers has to ration randomly until the available deposits have all been loaned out; the investment requirement of projects is assumed fixed, so rationing takes the form of a probability of obtaining a loan. The bank's problem is anticipated and solved by the applicants who are assumed to know their type; applicants with a success probability lower than some threshold do not bother to apply for a loan because of a (fixed) application cost. Given this cut-off success probability and the (exogenously fixed) interest rate, the bank is assumed to be at least breaking even (a fair assumption since otherwise the model would hold little interest). The bank can screen borrowers using the following technology: The effort decision is binary (in the sense that either all actual, not potential, applicants are screened or none at all; this assumption is required in light of the authors' stated intention of avoiding fixed costs of screening) and the outcome of the test is an accurate assessment of the true success probability. Whether a bank screens is common knowledge. This means that if a bank screens, all the entrepreneurs that bother to apply to it will also obtain a loan and therefore there will be no rationing. The quality of the borrower pool improves as the bank screens; the question is whether the increase in profitability outweighs the cost of screening. In the duopoly model, the entrant can screen or not screen and – in contrast to the incumbent – is free in the interest rate it sets. It is assumed that the entrant gains an exogenously determined share of the fixed-size deposit volume (which determines that bank's lending capacity and reduces the lending capacity of the incumbent relative to the monopoly situation). With two banks and two possible effort choices, there are four combinations of screening decisions to be investigated. This lends itself quite nicely to the standard normal form matrix representation (with mixed strategy equilibria discreetly being assumed away) and the authors argue that

'the two most likely outcomes under competition are that the incumbent or entrant exerts effort to capture the best borrowers while the competitor does not exert effort.' (p. 473)

It is shown that regardless of which of the four outcomes occurs, the threshold borrower quality (with applicants obtaining funding if and only if they have a success probability greater than this threshold) is the same as in the case of an informed monopolist. In other words, even when both duopolists do not screen their applicants, the allocation of capital improves relative to the case of monopoly without screening and is as good as that under monopoly with screening. The assumptions of

application costs together with that of applicants knowing their types are the key drivers of this result. When thinking about which bank to apply to (if any), an applicant needs to be concerned not just with the interest rate, but also with the probability of actually being given a loan. Since the entrant bank is free in its choice of interest rate, the certainty of getting funding at the higher interest rate gives rise to sorting. The model does not make unambiguous predictions with respect to the effects of competition on the incumbent's incentive to exert effort, although the authors point out

'that the effect of competition on the incumbent's effort decision depends positively on the entrant's market share and negatively on the extent of credit rationing when there is a monopolist lender, which increases the effort incentive under monopoly.' (p. 475)

It might be argued that this article ought to have been reviewed together with models in which screening accuracy is endogenous; in view of the technology employed and the fact that it is the possibility for self-selection rather than screening as such that drives the results, I found it more appropriate to discuss this paper here.

The article by Villas-Boas and Schmidt-Mohr (1999) is not so much a paper on bank competition as an exercise in contract theory. Two banks exogenously located at the extremes of a Hotelling line (where the travel cost as an indicator of the degree of horizontal differentiation is interpreted as the intensity of competition) face a pure adverse selection problem in lending to borrowers of two different types (success probabilities). Through the use of credit contracts which specify both an interest rate and collateral level, banks may be able to induce borrowers to reveal their types in a separating equilibrium. Initially the inefficiencies considered derive from the social cost of collateral (worth less to the bank than to the borrower because of liquidation costs) and the travel costs rather than from type I/II errors with respect to positive/negative NPV projects. (The case where the low-type applicants are negative NPV projects is considered later in the paper and it is shown that the parameter set for which surplus decreases with competition in that case is actually larger than in the case where lending to both types of projects is efficient.) The authors argue that in contrast to the standard Hotelling result higher travel costs in their setting do not necessarily reduce surplus ("welfare"), since more differentiation (less intense competition) may – at the same time as causing greater surplus losses from 'travel' – lead to lower surplus losses from the socially wasteful use of collateral. The set of parameter combinations for which this is the case is characterized:

The difference in success probabilities needs to be small and the absolute value of the probabilities relatively large. The preceding results are obtained for small enough travel cost, so that banks actually compete for all borrowers. The authors point out that as travel costs increase and banks become local monopolies (with respect to one or even both types) surplus may in fact increase further thanks to lower social costs of collateral.

The main shortcoming of the papers reviewed above is that banks' accuracy is exogenous or the information acquisition decision is extremely coarse. This feature would not only make practitioners frown; it is also dismissed by some of the world's foremost banking theorists. Freixas and Rochet (1997) argue that

'This appraisal of risk and correlative estimation of the risk-return on a bank loan is one of the main functions of modern banking.' (p. 6)

and they note that

'Banks may invest in an informational technology that allows them to screen the different demands for loans they are confronted with and to monitor the projects, (...)' (p. 7)

Greenbaum and Thakor (1995) write

'We should begin by noting that credit analysis, which is an integral part of the lending decision, is not a binary (0 or 1) process whereby the bank either conducts credit analysis or not. It should more appropriately be viewed as a continuum; the bank can perform credit analysis to varying degrees of detail. The more elaborate the analysis, the more costly it is for the bank. The point to note is that the degree of elaboration is a matter of choice for the bank and represents an important element of the spot lending decision-making process.' (pp. 276–277)

In the next few models, information acquisition is endogenous. In the preceding auction models, the accuracy of the signals was exogenously given. One might think that modeling a bank's creditworthiness test with the ensuing price competition as an (first-price, sealed-bid, common value) auction with endogenous information acquisition might be a promising route for further research. However, the technical difficulties of modeling this in a tractable way are considerable indeed. Well-known surveys of auction theory by McAfee and McMillan (1987) and Klemperer (1999) hardly touch on the subject despite its importance in a far wider context than that of bank competition. Two authors that have tackled the challenge (in a setup of restricted generality) are Matthews (1984) and, more recently, Persico (2000). Tröge (2000a) investigates information acquisition in ascending bid auctions with both common and private value components. In a companion paper [Tröge (2000b)],

Tröge studies information acquisition in a banking duopoly and finds that whether a bank knows the effort that the rival has expended on information acquisition is of critical importance for the extent of information acquisition relative to the surplus-maximizing amount. In a later contribution also using an auction setup, Tröge (2002) explores how the characteristics of an economy's portfolio of (potential) projects influences the surplus-maximizing number of banks. A first-price, sealed-bid auction with one better-informed bidder is studied by Hendricks, Porter and Wilson (1994).

As for bank competition being modeled as an auction, the paper by Cao and Shi (2001) goes some way toward endogenizing information acquisition: Type I/II error probabilities are assumed exogenous, but the probability of obtaining a signal at cost c can be chosen by the bank. Signals are binary and imperfectly informative. A bank does not observe other banks' information acquisition decisions and can decide to participate (or not) in a first-price, common value, sealed-bid auction and may, in particular, bid even if it has not obtained information on the applicant. The authors are looking for a symmetric equilibrium where the bid depends on the own test result (if any) and the common probability of acquiring an informative signal on the applicant's type. Their main concern is the likelihood that a good project will not obtain funding (the other type of error is not discussed in detail). The authors point out the possibility that an increase in competition (as proxied by an increase in the number of banks) makes it less likely that a good project will obtain finance, because the informational externalities of an increase in the number of banks may lead to a decrease in the number of banks that are actually bidding. This is the standard 'winner's curse' effect.

Hauswald and Marquez (2005) assume that symmetric banks compete in a spatial competition model à la Salop (1979). (One section in the paper considers the case of asymmetric banks in the context of a discussion of mergers. However, the main results discussed in the following are obtained for the case of ex ante symmetric banks.) The stages of information acquisition and price competition are preceded by a third stage at which banks decide whether or not to enter (though entry decisions are hardly the main concern of the paper). Banks can purchase an informative signal at a constant variable cost; types and signals are binary. The accuracy of the signal (captured as the probability that an assessment is correct) is increasing in the bank's investment in a screening technology and decreasing in the 'distance' (a proxy for specialization, say, by industry) of the borrower from the bank. Note that the spa-

tial competition setup does not serve to model loans as products that are horizontally differentiated by transportation costs; competition for borrowers is pure price competition between one informed and one uninformed bank (see below) and has only an equilibrium in mixed strategies. The distance between bank and borrower does, however, affect the accuracy of the signal. A crucial assumption is that whenever two competing banks obtain signals and their accuracies (affected by both the borrower's location and the bank's screening investment) with respect to a borrower differ, the less accurate bank only gets a noisy signal of the more accurate bank's signal. It is this assumption that ensures that a borrower will be screened at most once. Banks only incur the variable cost of screening for applicants located up to some maximum distance away from them; the reason being that the informed bank's expected profit decreases in distance, so that beyond some distance it is not worth incurring the cost of becoming informed. This creates the theoretical possibility of gaps between banks; i.e., market segments in which applicants are not screened at all. However, throughout much of the paper the authors assume that the cutoff distance is exactly the halfway point between two banks; this effectively ensures that each applicant gets screened exactly once (see p. 11). Interest-rate competition is modeled as one informed bank bidding against one uninformed bank. As the first author kindly pointed out to me in personal communication [Hauswald (2005)], this is not due to the Salop setup (where each firm competes only against its neighbors), but rather is derived from a result in Engelbrecht-Wiggans, Milgrom and Weber (1983). The informed bank is, in fact, competing against all uninformed banks; however, Engelbrecht-Wiggans, Milgrom and Weber (1983) show that this is equivalent to one better informed bidder competing against one less informed bidder. In the interest rate competition sub-game, the informed bank denies credit to applicants with a bad signal if its accuracy exceeds some threshold level; also in that case the uninformed bank offers a loan only with some probability. If the informed bank's accuracy is less than the threshold level, both banks always offer credit. Interestingly, the more accurate the informed bank, the higher the (expected) interest rate paid by both types of borrowers (although in this paper as in most others in this literature, interest rates are completely without allocative significance). The intuition behind this result is that a better-informed competitor worsens the adverse selection problem and so forces the uninformed bank to ask for a higher interest rate (in expectation) with the informed bank taking advantage by raising its interest rate as well. In the same vein, the interest rate is decreasing in the dis-

tance between an applicant and the informed bank (and conversely for the uninformed bank) because the informed bank's informational advantage diminishes with distance. The optimum investment in the screening technology is analyzed and it is shown that it decreases in the intensity of competition (the number of banks) and increases in the fraction of bad types. Banks invest strategically to, as it were, worsen the adverse selection problem for their competitors (in turn enabling them to charge higher rates) and also to steal business from the neighboring banks by becoming better informed about those borrowers for whom competition is most balanced (in the sense that they are located close to equidistantly between the two rivals, so that no bank has an overwhelming informational advantage). In equilibrium with symmetric banks such attempts do not bear fruit in that no bank gains any market share; the authors show that bank profits would be higher if banks colluded, divided up the market '50:50' by agreeing not to screen applicants located beyond the halfway point and reduced their investment in the screening technology accordingly. Does greater accuracy improve the allocation of capital in the sense that less type I and II errors are made? Assuming that accuracy is high enough for the informed bank not to lend to an applicant with a bad signal and thus for the uninformed bank to be bidding only with some probability, type I errors (good types not obtaining loans) decrease in accuracy only if it exceeds some threshold; below the threshold type I errors become more likely the greater accuracy! (Type II errors always decrease in accuracy). What explains this at first glance counterintuitive result is the effect of improved accuracy in worsening the adverse selection problem for the uninformed bank. The likelihood that a good applicant will obtain a loan from an informed bank is increasing in accuracy indeed; however, the uninformed bank becomes disproportionately more cautious in granting loans as the informed bank's information improves and refuses to make an offer with greater probability. That there should be a threshold above which the beneficial effect of more accurate information dominates becomes clear when one imagines the error probability going toward zero. Then every good project that is screened also (in the limit) obtains funding from the informed bank. The threshold is a function of the number of banks (i.e., the intensity of competition); the authors show that when competition is already intense, good borrowers will benefit from further entry. Conversely, when the number of banks is small, entry makes it less likely that a good borrower will obtain a loan. Since entry always reduces equilibrium accuracy, type II errors become more likely as competition becomes more intense. The authors show that

an increase in the intensity of competition has a non-linear effect: If the number of banks is already high, more entry is good (in the sense that gross surplus – leaving aside the costs of investing in the screening technology – increases). If the number of banks is low, entry is not beneficial. The surplus ("welfare") effects are the main concern of this review, so will the investment in accuracy and the resulting allocation of capital be efficient overall? Once the costs of investment are taken into account, the authors show that – in line with an earlier observation – investment in the screening technology is socially excessive under free entry. The authors derive from this result the remarkable policy implication that it may be socially beneficial to limit banks' investment in accuracy. What is the bottom line with respect to competition (entry)? In line with the preceding comments, the effect is not clear-cut and depends on how competitive the market already is. One possible outcome that contradicts much of the received wisdom in the literature is that further entry may be surplus-increasing precisely because it leads to a reduction in socially excessive screening investment.

Leaving behind the auction setup, in Chan, Greenbaum and Thakor (1986) competition is not modeled explicitly. An exogenous interest rate term parameterizes the intensity of competition. Lending occurs at two dates and the authors' main interest is in the reusability of information and how the degree of reusability affects screening incentives ex ante. For information to be reusable, there needs to be some positive correlation between borrower type at the two dates (and the bank also needs to be still solvent at the second date). It is assumed that there is no additional screening at the second date (and that no new borrowers enter the market). At the first date, the bank has access to a costly screening technology which determines the mix of applicants in terms of good and bad projects. The technology is not modeled explicitly, rather the proportion of applicants that repay is assumed to have a distribution that is affected by the screening intensity. Note that this model – in contrast to most other papers reviewed in the present section – features aggregate uncertainty in the sense that the bank may go bankrupt; screening, by changing the composition of the applicant pool, merely changes the likelihood of this happening. The intensity of screening is shown to be greater the greater the correlation between type at the first and second dates (which increases the durability and therefore the reusability of information). Macroeconomic fluctuations could be thought to be one factor leading to less durable information. Higher lending (and lower deposit) rates also increase the intensity of screening. In so far as more intense competition

leads to lower interest rates, competition is then bad for screening incentives. Because project characteristics and loan demand are not modeled in detail, it is not clear whether more intense competition will reduce the surplus from investing, but there is certainly a presumption that it will. At any rate, a lower screening intensity translates directly into a higher likelihood of bank failure. The increased (expected) cost of bank failures would then need to be taken into account in any assessment of the effects of more intense competition.

Gehrig (1998a) endogenizes the accuracy of banks' creditworthiness tests explicitly. In his model, projects have a binary cashflow distribution (successful – unsuccessful). There are two types of projects, good and bad, which differ in their success probabilities. Conditional upon success, cashflows are drawn from a uniform distribution. Each entrepreneur is assumed to know his conditional-upon-success cashflow, but the bank cannot learn anything about it. Whether entrepreneurs know their type, however, turns out to be irrelevant. Bad projects' success probability is by assumption so low that no bad project should be financed. Good projects are socially valuable if their cashflow exceeds a certain threshold. The opportunity cost of capital is captured by the banks' (constant and exogenous) deposit rate. Despite the absence of bankruptcy, search (or application) costs, collateral, or equity and despite limited liability, demand is elastic because entrepreneurs by assumption only wish to borrow if their conditional-upon-success cashflow is strictly positive. Thus, if a project is successful, the entrepreneur is willing and able to repay the bank in full. An unsuccessful project does not repay anything. Entrepreneurs' effort does not matter and as the interest rate increases, it is the lower quality projects that drop out first (where quality is measured by the conditional-upon-success cashflow; the success probabilities being homogeneous within each group). The screening technology available to banks takes the following form: By investing resources, a bank can obtain increasingly accurate (but imperfectly informative) binary signals about the true type of an applicant. If the bank does not screen, signals correspond to population proportions (assumed to be common knowledge). One important technical assumption in Gehrig's model (which drives one of his results) is that the total shares of 'good' and 'bad' signals correspond to population proportions regardless of screening effort: This allows the determination of both type I and type II errors with a single choice variable. Gehrig uses a general function and does not impose a specific functional form. He considers a monopoly and a symmetric duopoly (albeit in one sub-case allowing for a first-mover

advantage of the incumbent); each entrepreneur is assumed to borrow from a single bank only. The structure of the game is not such that banks first decide on their screening effort and then compete in prices: The two choice variables are chosen simultaneously. With the implicit assumption of a continuum of borrowers, there is no aggregate uncertainty in this model. Gehrig obtains the following results: He derives a condition under which a monopolist's screening effort is increasing in the loan rate and notes that this depends on the relative value of identifying a good project and avoiding a bad project (this is driven by the assumption with respect to the screening technology mentioned above). Gehrig considers two duopoly settings: In the first, Gehrig assumes that in a first round, a monopolist has the market to himself, that borrowers deciding to borrow from the monopolist are not allowed to switch to the entrant in round two, and that the entrant cannot credibly commit to a 'low' interest rate. Gehrig shows that with the right assumptions on parameters, one may obtain equilibria where the incumbent continues to charge the monopoly rate and continues to screen at the monopoly intensity even after entry. In that case the entrant charges a rate higher than the monopoly rate and screens at a lower intensity than the monopolist. This is because the entrant suffers from an adversely selected pool. Gehrig does not discuss what happens when these restrictions with respect to the parameters do not hold. When considering simultaneous competition between two symmetric duopolists, Gehrig finds that this game has no equilibrium in pure strategies [for the reasons analyzed by Broecker (1990)]. However, once again under certain assumptions about parameters, Gehrig shows that there may be a symmetric pure-strategy equilibrium with zero screening effort in which the competitive interest rate may be higher than the monopoly rate. Given the sensitivity to parameter values which do not have clear interpretations, generally valid conclusions with respect to the allocation of capital and the impact of competition (switching from monopoly to duopoly) thereon are hard to draw. [Gehrig (1998b) extends the basic framework by allowing successful applicants to renegotiate their offers by playing off one bank against the other and likewise finds ambiguous effects of competition on surplus.]

Schnitzer (1999a) also endogenizes the bank's screening decision. There are two types of projects, good and bad, with deterministic cashflows. In the first part of the paper, only the impact of market structure on screening decisions is considered. In the second, rather brief, part, market structure is allowed to have an impact on surplus via the entrepreneur's incentives to invest (non-verifiable, non-contractible) effort into

the project (this activity is called 'restructuring' and translates either into a higher probability of success or a higher conditional-upon-success cashflow). The basic link is via interest payments: If a monopolist can expropriate all the surplus, it is not worthwhile for the entrepreneur to exert effort even when that is efficient. As for the screening part, banks have access to a technology that allows them to find out the true type of a project (with certainty) by paying a fixed fee. Schnitzer considers a monopoly and a duopoly. In the duopoly case, competition is à la Bertrand. Two sub-cases are analyzed: In the first, the results of bank's creditworthiness tests become public (with the result that neither bank screens); in the second, the test results are private information. The model has a screening stage followed by a price-competition stage (at which the competitor's screening effort is assumed known). The first result follows directly from the assumption that it is efficient to screen projects and states that a monopolist has optimal screening incentives so that the resulting capital allocation is efficient. If test results cannot be kept secret neither bank screens and the resulting capital allocation is inefficient (because a bad project would be funded). When one bank is informed and the other bank is not, there is no pure strategy equilibrium, but Schnitzer characterizes the properties of the mixed strategy equilibrium (including bidding functions and expected payoffs). Building on this result, the outcome when test results are private information is described. In that case there are two pure strategy equilibria with one bank screening and the other not screening (but nevertheless bidding) and one symmetric mixed strategy equilibrium. In each of these equilibria the capital allocation is not as efficient as under monopoly (because the screening cost is incurred more than once or because a bad project would be funded). Schnitzer claims that even under competition, a bank can have optimal screening incentives – but this is only the case if with probability one the rival does not screen. Two different technologies assumed to be available for 'restructuring' by the entrepreneur are considered: Schnitzer points out that when the effect of the interest rate on the entrepreneur's effort level matters, the presumption in favor of monopoly will, in general, no longer hold. Various settings are outlined in which one of the three market structures considered here (monopoly, duopoly with private test results, duopoly with public test results) could be said to dominate. This depends on the parameters in a straightforward way: If 'restructuring' by the entrepreneur doesn't matter (much), monopoly is best. If it is important that the entrepreneur be given incentives to 're-structure' and screening is costless, then duopoly with private test results

is best. Information spillovers plus competition are best when screening is costly. It is interesting how adding the second stage ('restructuring') and moving away from an entirely passive borrower makes an important difference here. Competition in the first ('screening only') part of this model really cannot have much of a positive effect on the total surplus generated. (In particular, since screening leads to accurate information and a monopolist has optimum screening incentives there is no benefit from giving a project a second chance.) This may be different when the intensity of competition influences surplus via the effect of the interest rate on the entrepreneur's effort incentives.

In a companion paper, Schnitzer (1999b) studies bank competition in a horizontal differentiation model. The applicants are distributed uniformly along a Salop (1979) circle and banks are assumed to be located equidistantly. The market is assumed to be covered and transportation costs are proportional to distance. (Note, though, that bad types are assumed not to care about either interest rate or location.) As in Schnitzer (1999a), banks initially face an adverse selection problem which they can counter by screening applicants (testing their creditworthiness). The testing technology again is such that by paying a fixed fee a bank learns the type of all applicants with accuracy. A crucial assumption is that it is not actually necessary to screen a borrower; random lending is feasible. Schnitzer also assumes that there is no free-riding on rivals' test results (either through direct information spillovers or by borrowers going around with evidence of the offers they have received). It is shown that a monopolist bank has optimal incentives to invest in screening. With more than one bank, banks' profits are decreasing in the number of banks that choose to screen, so the more banks are informed, the greater the incentive for an uninformed bank to screen as well. For low enough screening costs, the unique equilibrium is for all banks to become informed. For large enough screening costs, there are two pure strategy equilbria in which all banks either screen or do not screen. The 'no screening' equilibrium may generate higher surplus than the equilibrium with screening. The tradeoff in terms of surplus is between the inefficient multiplication of screening and the loss from bad projects being financed. Schnitzer shows that even without exogenous costs of entry the number of entrants is bounded. While there is no explicit fixed cost of entry (as is traditionally assumed in this kind of horizontal differentiation setup) this model features an implicit cost of entry: The bank either has to pay to screen its applicants or accept a certain amount of loan losses. Schnitzer shows that the number of banks under free entry may exceed the socially

optimal number of banks: This will be the case whenever screening costs are low enough. In an extension, the effect of the interest rate on the incentives of the borrower to exert effort is sketched.

The banks in Dell'Ariccia (2000) can screen applicants to filter out bad projects. The applicant pool consists of a fraction of new projects with characteristics unknown to all banks and a complementary fraction of 'old' applicants whose type is known to one bank. Banks will try to hang on to their (known) good borrowers, but get rid of the bad types. (Bad types do not go bankrupt and disappear, but rather re-enter the market, worsening the pool of applicants.) The proportions of good and bad projects within each group, however, are the same. Screening is assumed to be costless to the bank and perfectly accurate, but it imposes a disutility on the entrepreneurs. In the initial version of the model (extended later in the paper), the interest rate is exogenously given and banks compete on screening requirements! Crucially, the average creditworthiness is assumed to be sufficiently good for random lending (i.e., without screening) to be profitable. Dell'Ariccia shows that for a high enough proportion of new applicants, there may be equilibria (in either pure or mixed strategies) in which banks do not screen. This reduces surplus from investing in that bad projects would be financed. Competition in this setup is undesirable because an increase in the number of banks makes it less likely that banks will screen their applicants in equilibrium. However, in an extension Dell'Ariccia shows that this finding is non-robust to how competition is modeled.

The fundamental point of Manove, Padilla and Pagano (2001) is to argue that – contrary to what much of the theoretical and empirical literature is claimed to suggest – there is a downside to strong creditor-protection laws enabling the extensive use of collateral. The authors' argument is of some relevance for the matter of bank competition, even if the optimum intensity of competition is hardly their main concern. Manove and co-authors point out that a bank that is fully protected by collateral has less incentive to accurately assess project quality – if the project fails, the bank does not stand to lose anything:

'Collateral and screening are substitutes from the point of view of banks.' (p. 728)

However, since project failure is costly to society, banks' screening increases surplus (leaving aside the costs of screening). In this view, weaker protection for lenders may act as a commitment device that forces banks to screen since they cannot rely on collateral to protect themselves against default. Weaker legal protection for financiers may then lead to less credit

being granted, but the reduction in lending is actually efficient! In the authors' model, screening is non-contractible and can thus only be 'sold' packaged together with a loan. If the applicant posts full collateral, the bank has no incentive to screen and identify project type. The screening cost is assumed to be small enough for screening to be efficient and the question arizes why contracts would not emerge endogenously that give banks an incentive to screen – by specifying a low enough collateral value – and thus give rise to an efficient equilibrium. That happy outcome would indeed obtain if the information problem were simply one of entrepreneurs not knowing the type (good or bad) of the project which they are assumed to choose at random. However, things are more complicated when – as the authors assume – there is an element of adverse selection with the applicants falling into one of two categories unobservable to the bank. Applicants know their category with high quality applicants having a greater likelihood of choosing a good project (although applicants do not know the type of the project they have chosen). In this scenario, high quality applicants have an incentive to signal their type by posting full collateral. The reason is the following: In any (competitive) equilibrium in which the banks make zero profits, those borrowers that obtain a loan will have to cover the costs of screening, including the cost of screening those projects that are rejected. Since high quality applicants have a lower likelihood of choosing a bad project and being rejected, the pro-rated screening costs are lower for them in the separating equilibrium. Given certain assumptions with respect to the parameters, the low quality entrepreneurs have no incentive to mimic the other group and post full collateral, because their likelihood of losing the collateral is (much) higher. However, the separating equilibrium is socially inefficient, because the group of high-quality entrepreneurs still chooses a sufficient number of bad projects such that screening that group would be efficient. In the light of the preceding comments, however, that does not happen because the banks are fully protected against any loss. One possible policy implication is that weaker protection for creditors might lead to a more efficient outcome. (Having said that, the authors do acknowledge that collateral also serves to reduce moral hazard which, however, this model explicitly does not feature.) Competition enters the picture in that a monopoly would lead to efficiency. The reason for this is straightforward: A monopolist manages to appropriate the entire surplus generated by a project and so has the right incentives to not use collateral and instead screen every project, thereby avoiding inefficient financing decisions. The main advantage of restricting competition in

this model, therefore, is that market power increases the incentives to screen.

2.4.3 Monitoring

In Caminal and Matutes (2002) the moral hazard problem is double-sided: The bank cannot commit to the efficient level of monitoring (and there is nothing that can be done about it other than to provide the bank with different degrees of market power); moral hazard on the part of the entrepreneur arizes via the choice of production technology. The entrepreneur's production technology is modeled in a richer fashion than in most of the literature: The entrepreneur picks both scale and the type of project, where the type of project determines the success probability and the conditional-upon-success cashflow (which is also influenced by the decreasing-returns-to-scale production function). Specific functional forms are assumed and justified later in the paper. The greater the success probability chosen by the entrepreneur the greater the expected return (but the lower the conditional-upon-success cashflow). The efficient choice of success probability is therefore one and the efficient scale of investment is to set the marginal productivity of capital equal to the bank's refinancing interest rate. The assumptions with respect to the production technology also allow for the reasonably straightforward parameterization of the entrepreneur's moral hazard problem. With a certain probability, increasing in the bank's effort decision (with an assumed quadratic cost function), the bank can observe the choice made by the entrepreneur. Effort decision and whether the monitoring has been successful are not known to third parties. In particular, the fact that the entrepreneur's choice observed by the bank is not verifiable means that the gain from monitoring is not that this would allow for contracts to be written contingent on the entrepreneur's choice. Rather, the assumption is that the success or otherwise of a bank's monitoring effort is known to both borrower and bank before project type and the scale of the investment are chosen. In other words, if the bank's monitoring is successful, the entrepreneur and the bank enter negotiations and will typically agree on a larger (more efficient) investment scale. The bank is assumed to be able to find out whether the project has failed or not (regardless of whether it has learnt the entrepreneur's choice), but is assumed not to find out about the cashflow. This is a somewhat tricky combination, but the main idea behind it is simply to justify the use of standard debt contracts. Given the informational assumptions, the optimal contract is sketched out. That contract specifies a default in-

vestment level and interest rate and becomes effective if monitoring has not been successful. If monitoring was successful, the two parties start bargaining and split any gains relative to the default contract equally. The bank will set monitoring effort in anticipation of these outcomes; it is shown that the bank's monitoring effort increases with the interest rate specified in the default contract. Competition between banks enters the picture in that prior to the monitoring, negotiation and investment stages of the model, banks compete for borrowers. The market structures compared are pure price competition under 1. monopoly and 2. with two or more banks as well as 3. price-competition with horizontal differentiation between a number of banks on a Salop circle. It is shown that no market structure unambiguously dominates the others in terms of the surplus generated. The basic tradeoff is the following: The more market power banks have, the stronger their incentives to monitor and the more banks monitor, the fewer projects will go 'undetected' and suffer from inefficiently low investment. However, given that an entrepreneur is not detected, the shortfall of investment relative to the efficient level will be more severe the more market power banks have. Depending on the relative importance of these effects, it may be socially optimal for banks to have a degree of market power. In his survey, Vives (2001b) has succinctly summarized the tradeoff as follows:

'More market power for the bank diminishes the moral-hazard problem faced by the bank, but aggravates it for the entrepreneur.' (p. 539)

2.4.4 Liquidity assistance

The mechanism through which banks influence the allocation of capital in Petersen and Rajan (1995) is similar to 'liquidity assistance' (in both cases banks could be thought to cross-subsidize intertemporally), so the present sub-section seemed the most appropriate place for a discussion of this contribution. Competition is not modeled explicitly, it is parameterized by the interest-rate markup that the bank can demand. The model features

good and bad applicants as well as a moral hazard problem on the part of the borrowers, but banks cannot screen or monitor. Rather, banks' role is to intertemporally shift returns from lending to the later phase of the lending relationship. This allows banks to charge low initial interest rates which, in turn, makes it more likely that the moral hazard problem will be defused. A crucial assumption, therefore, is that firms can invest at two dates. Importantly, the bad projects drop out after the first period (they fail and return nothing). In terms of the notational

conventions used in the paper, banks can thus costlessly detect firms' types at date 1, but have no means of telling them apart at date 0. The bank is constrained in the interest rate that it can charge between date 0 and date 1 because of the moral hazard problem on the part of the good entrepreneurs who, if charged too high an interest rate, have an incentive to shift risk and pick a risky negative NPV technology rather than a safe one. The more market power the bank has, the more easily it can extract rents from surviving entrepreneurs (by way of 'implicit equity') between dates 1 and 2 (when, by assumption, entrepreneurs do not have the option of choosing a risky project; the existence of a moral hazard problem at date 1 would limit the ability of the bank to charge a high interest rate and complicate the message of the model somewhat) allowing it to charge lower rates in the first period and still (at least) break even. Where is the social gain from that? The authors don't state this explicitly, but if the proportion of good entrepreneurs were too low, the market would shut down. What the model shows is that the more market power banks have, the lower the proportion of good projects required for lending to be viable. For this reason, more intense competition impairs the allocation of capital. (That banks learn the type of their borrowers is irrelevant; bad types cannot be filtered out ex ante.) The article has been influential and derives much of its recognition from its empirical part (briefly discussed in section 2.6) which would appear to provide support for the model's assertions.

The banks in Dinç (2000) influence the allocation of capital via two channels: They screen borrowers and they (potentially) assist borrowers with a cash injection in distress states. It is quite clear that Dinç is primarily interested in the liquidity assistance mechanism (which he plausibly equates with relationship lending), so I shall concentrate my comments on this point, too. [As for screening, bank competition is once more modeled as a first-price, sealed-bid, common-value auction in which banks costlessly receive continuous signals; i.i.d. conditional on type. There are two types of borrower and signals are assumed to satisfy the monotone likelihood ratio property. However, the surplus ("welfare") effects of screening are not investigated in detail in the article. That analysis would involve an assessment of aggregate type I/II errors under different competition regimes.] Dinç takes as the starting point the observation that much lending happens on the basis of implicit contracts between banks and borrowers and that a desire to build or keep a good reputation may be one reason why a bank might honor such implicit, non-enforceable obligations even when this is costly to the bank. Dinç's

main question is

'how credit market competition changes the effectiveness of bank reputation in enforcing a bank's commitment.' (p. 782)

The setup here is truly dynamic and modeled as an infinitely repeated stage game (with three dates / two investment sub-periods). Banks are infinitely lived, entrepreneurs come and go. This latter assumption ensures that the commitment problem is one-sided: There is no question of the entrepreneur undertaking not to borrow from other banks. However, whether the bank will honor an implicit commitment to provide rescue funds to a distressed entrepreneur at an intermediate stage of the project is at the heart of the paper. Such a cash infusion is assumed to be socially efficient (though it leads the bank to incur a loss); such liquidity assistance clearly is a source of 'surplus from investing'. Entrepreneurs prefer to borrow from a relationship lender; i.e., a bank that promises to rescue them should their project go into distress. Should a bank ever break this promise, no entrepreneur will borrow from that bank on a commitment basis again. The gain to the bank of being known as trustworthy is captured as such a bank being able to charge a higher interest rate (having greater market power) than an arm's length lender. Whether relationship lending (with commitment) is viable depends on the difference in market power between the two types of lender; i.e. the size of the interest rate differential. This is of crucial importance for the results of the model, as this opens the possibility of relationship lending becoming 'more viable' as competition becomes more intense because more intense competition might reduce the market power of an arm's length lender (the interest rate that an arm's length lender can charge) by even more than that of a relationship lender. An increase in the number of banks (higher intensity of competition) affects both the arm's length and the relationship market with the market power (and thus the interest rate) of each type of bank endogenously determined as the outcome of competition between banks (modeled as an auction). In the one-shot symmetric equilibrium, no distressed borrower is assisted. In the infinitely repeated game, three regimes can be distinguished: For very good signals, relationship loans are offered, for very bad signals credit is refused, and for intermediate signals arm's length loans are proposed. Interestingly, relationship lending does not become monotonically more difficult to sustain as the number of competitors increases (competition becomes more intense). While in the limit (for very large n) relationship lending is only viable when there is no discounting, there is no relationship lending with monopoly either: The monopolist is already in a position where she can extract the borrower's

entire surplus and has no incentive to help in a distress state. Even two banks may not be enough to sustain relationship lending. If competition becomes more intense in the arm's length market, the relative return to providing loans with a rescue commitment increases, leading relationship lenders to lower the threshold of the signal at which they are prepared to offer relationship loans. (Dinç also considers competition from bond markets, but in line with my earlier comments I omit that passage.) The bottom line as far as my topic is concerned is that 1. liquidity assistance in distress states may be an important source of surplus ("welfare") gains (an uncontroversial assertion) and that 2. relationship lending may not become monotonically harder to sustain as competition becomes more intense. Thus surplus ("welfare") gains from liquidity assistance may be maximized for intermediate degrees of competition. This is in contrast with the result in Petersen and Rajan (1995) where less market power for the lenders always requires a better composition of the applicant pool for relationship lending (indeed, any lending) to be viable. That commitment should become monotonically harder to sustain as competition becomes more intense is a theme beyond the banking literature [see the discussion in Petersen and Rajan (1995)], making Dinç's somewhat counterintuitive result especially interesting.

I next review two contributions that should arguably be considered 'related contributions', because they do not feature liquidity assistance and do not focus on surplus. However, because of the similarities with Dinç (2000) it seemed more appropriate to discuss them here.

Boot and Thakor (2000) ask how more intense competition will affect relationship lending. They distinguish between more intense competition between banks and more intense competition between banks and the bond market and argue that the effects on relationship lending are different. (In line with my earlier remarks, I concentrate on inter-bank competition.) As in Dinç (2000), banks can offer either relationship loans or arm's length finance. A relationship loan in Boot and Thakor (2000) does not involve a promise to help with liquidity problems, but (following an investment by the bank in what the authors call 'sector specialization') is simply worth more to the borrower than a transaction loan (this is captured as an increase in the borrower's cashflow in that the success probability of the project is higher when it is funded with a relationship loan). A borrower's valuation of the benefit of a relationship loan over and above that of a transaction loan is assumed to decrease in the borrower's quality (creditworthiness). More intense competition leads to lower investment in 'sector specialization' because of lower mar-

gins. However, as in Dinç (2000) competition reduces the return from both relationship and transaction lending. In contrast to much of the literature, banks' lending capacity in this model is fixed and the asymmetric effect of more intense competition leads banks to dedicate a larger share of their funds to relationship lending rather than transaction lending (even though the value of a relationship loan to borrowers is lower than in the absence of competition). The effects on surplus ("welfare") – which is not considered explicitly – differ according to a borrower's creditworthiness. The higher the quality of a borrower, the more the borrower will benefit from more intense competition. Borrowers of sufficiently low quality may be made worse off. The impact on the surplus from investing in this model does not derive from screening (there are no negative NPV projects) or monitoring, but rather from relationship lending increasing the size of the pie that can be shared between bank and borrower. The results are partly driven by a refinancing side that is a little more elaborate than in most other models in that relationship loans are costlier to refinance than transaction loans (which are costlier to refinance for very high-quality projects than direct borrowing on the capital market would be). The intensity of competition is parameterized by the number of banks (which determines the likelihood that an applicant will receive more than one offer) which in turn is driven by a fixed cost of entry.

Yafeh and Yosha (2001) also contribute to the literature that examines how relationship banking specifically is affected by more intense competition. They tackle the common presumption that market power is required in order for relationship lending to be feasible:

'Without market power, banks would not be able to extract rents generated by investment in ties with firms, and consequently there would be little or no relationship banking.' (p. 66)

Yafeh and Yosha argue that if one takes into account interactions between the markets for relationship lending and arm's length finance, respectively,

'under certain circumstances, greater potential competition in the arm's length market may actually induce more relationship lending, not less.' (p. 66)

Investment in relationship lending is parameterized as the fraction of entrepreneurs that could potentially obtain a relationship loan. This share is a choice variable for the bank. The basic tradeoff is between two effects: On the one hand, more intense competition in the arm's length market reduces profitability in that market and induces banks active in

both market segments to invest relatively more into relationship lending. On the other hand, more intense competition in the arm's length market (which in the basic setup is modeled as an increase in the number of Cournot competitors) leads to lower interest rates on that type of loan. A bank offering relationship loans (which in this model do not have any explicit dynamic element and could be thought to be simply loans of higher quality) then needs to adjust the terms on such loans by leaving more of the surplus to the borrowers which makes relationship loans less profitable and thus makes the bank invest less in relationship lending. For high degrees of competitive intensity (large numbers of Cournot competitors in the arm's length market), the latter effect dominates and additional competitors lead the bank to invest less in relationship lending. As far as surplus is concerned, the authors show that investment in relationship loans would actually be excessive, although surplus ("welfare") is not the primary variable of interest. In an extension, the authors also investigate entry deterrence in the arm's length market in the context of this model.

2.4.5 Bank competition in general equilibrium models

Most models of bank competition are partial-equilibrium models. However, a strand of work embeds bank competition in an overlapping-generations (OLG) framework à la Diamond (1965). It seemed sensible to group together these general equilibrium models regardless of the channel through which banks affect surplus.

Screening intensity is arguably endogenous (if not modeled in great detail) in Cetorelli (1997) who considers the polar cases of monopoly and perfect competition; the effect of competition on surplus is captured via its impact on capital accumulation (steady state level of per-capita capital). Banks can assess an applicant's type with accuracy at a cost that is linear in the amount of capital intermediated. In terms of the economics, the key drawback to competition is that banks are assumed to be able to free-ride on a competitor's information which destroys ex ante incentives for assessing creditworthiness and leads banks to lend to all applicant types. As Cetorelli himself points out, the assumption that the information be available to competitors for free is not critical; any cost lower than the cost of the bank that first screens the applicant would allow the results to go through. However, the assumption of monopoly is critical; i.e., this is not a model about different degrees of market power. Even a second bank is enough to destroy screening incentives. The monopolist

bank wastes less capital lending to bad types, but lowers capital accumulation through the exercise of market power in the deposit market: It offers a lower interest rate to savers than a competitive banking system and mobilizes less capital for intermediation. Monopoly can lead to a higher per-capita steady state level of capital only if the proportion of good types is relatively low and the value of screening correspondingly high. Whether monopoly actually is better in this sense than perfect competition depends on the parameters (cost of screening, interest elasticity of savings, interest elasticity of loan demand). Cetorelli argues that developing economies' characteristics imply that monopoly may be beneficial there.

The same tradeoff is studied by Cetorelli and Peretto (2000) in a different framework that models competition as Cournot competition (rather than either monopoly or perfect competition) and so parameterizes the intensity of competition more finely. Surplus in that model may be maximized by intermediate degrees of competition.

Guzman (2000b) also studies the effects of banking market structure on growth in an OLG framework. One source of surplus losses from monopoly in his model is that monopoly makes it more likely that credit will be rationed. If the conditions give rise to credit rationing, monopoly will exacerbate it. Second, a monopoly charges a higher lending interest rate which requires more ex post auditing on the part of the bank with an associated loss of resources. Screening does not play a role here in that all projects are identical ex ante. The contribution by Smith (1998) is also built around an OLG model, albeit augmented by stochastic shocks [following Bernanke and Gertler (1989)]. Banks do not screen (all projects are ex ante identical), but rather act as ex post 'cashflow-verifiers' as in Diamond (1984) and arize endogenously. Imperfect competition (here driven by switching costs) is unambiguously bad in that it drives up the lending interest rate, reduces output and amplifies business-cycle fluctuations. Guzman (2000b) and Smith (1998) are the only papers I am aware of which model bank competition in a costly-state-verification setup. 'Quality of information' is not really endogenous here and the main effect of competition in these papers operates via allocative efficiency.

In Paal, Smith and Wang (2005), banks act as liquidity providers along the lines of Diamond and Dybvig (1983) in an endogenous growth model with money and overlapping generations of consumers. The authors compare a competitive and a monopolistic banking system. At low interest rates, depositors can self-insure against liquidity shocks at

low cost (the opportunity cost of holding cash is small). The higher the interest rate, the more market power a monopolist has. A monopolist bank can hold lower cash reserves and invest a larger portion of assets in productive real projects. This would suggest that surplus is greater under monopoly. However, a monopolist bank will also hold down the rate of return paid to depositors, leading to a lower savings rate, less capital accumulation, and lower rates of growth. The net effect is determined by the nominal rate of return and depositors' preferences. Even after taking into account that risk-sharing is always at least as good under competition as under monopoly, neither market structure unambiguously dominates. In particular, for low enough nominal interest rates, both systems create an equivalent amount of surplus.

2.4.6 Related literature

I consider an article as a 'related contribution' if it has only some feature(s) of interest and is thus somewhat peripheral to the main topic of this review. For example, a paper might be included if it models bank competition in an interesting way, but with a view to some outcome of interest other than the surplus from investing.

Marquez (2002) considers a pure adverse selection problem, albeit one where the borrowers do not (initially) know their own types. Learning about borrower types by banks in a first stage is automatic and accurate; the main interest of the paper is in investigating the impact of the extent of information asymmetry in the second-stage market. The information asymmetry is parameterized by the exogenous fraction of borrowers that look for a new lender for reasons other than them being of poor quality. More intense competition (which in this model cannot merely be equated with an increase in the number of banks, but rather requires looking at a number of parameters) has ambiguous effects on interest rates.

Dell'Ariccia and Marquez (2004) also consider a pure adverse selection problem. Their model features a banking duopoly where one bank has an informational advantage. The idea is that a fraction of the applicants has borrowed from one of the lenders before and its type is therefore assumed known to the lender [i.e., the bank is essentially passive; there is no screening or monitoring technology, the lender identifies type as a costless by-product of lending as in, e.g., Petersen and Rajan (1995)]. It is of critical importance to the results that the borrowers have no way of credibly conveying this information to the uninformed lender (i.e., no free-riding on the informed bank's information). The uninformed bank is assumed to have a funding advantage – this implies that there is no clear

benchmark interest rate, making statements about surplus problematic. The authors are anyway not concerned with surplus, their main interest is in investigating the impact of changes in the severity of the adverse selection problem and in how the impact of changes in the uninformed bank's cost of funds differs across markets with different degrees of informational asymmetry. The authors argue that the uninformed bank's cost of funds parameterizes the intensity of competition and that competition that is in that sense more intense leads to a 'flight to captivity' of the informed bank; i.e., more lending to informationally captive borrowers.

Almazan (2002) embeds two banks competing on a Salop circle into the model in Holmstrom and Tirole (1997). Just as in Holmstrom and Tirole (1997), there is a double moral-hazard problem: Borrowers have the choice between two actions (one 'good', but not providing them with private benefits; the other 'bad', but giving them private benefits). The only way for them to commit to take the good action is for them to be monitored by a bank which reduces the private benefit from making the 'bad' choice. The bank in turn has the same problem: Monitoring imposes private costs on the bank and if the bank does not inject some of its own capital, a promise to monitor the borrower is not credible. Bank capital is assumed exogenous (see p. 97 of the paper for a defense of that assumption). A bank's cost of monitoring is linear in its distance from the borrower; that distance is interpreted as the bank's 'expertise' (location corresponds to specialization; the location choice is endogenized later in the paper).

'The major question addressed with this framework is the following: How do specific forms of deregulation affect banks, and how do these effects differ across high-capital and low-capital banks? The different forms of deregulation I consider here are those that result in an increase in capital requirements, in an increase in interest rates, and in the lifting of geographic restrictions on the operations of banks.' (p. 88)

The last of these measures in particular could be seen as parameterizing the intensity of competition. However, this paper only touches on bank competition and capital allocation since the major source of surplus loss considered results from an entrepreneur borrowing from bank A when fewer resources would be required for the borrower to borrow from bank B. There is no heterogeneity in projects: Banks do not screen; as far as monitoring is considered, monitoring intensity is the same across borrowers. In the conclusion Almazan notes that in addition to endogenizing bank capital, allowing for agency conflicts within the bank would be a worthwhile extension. This points to the nature of the moral-hazard

problem on the part of the bank in Holmstrom and Tirole (1997) and, by extension, Almazan (2002): the bank manager, implicitly, is claimant to 100% of the bank's profits, so moral hazard is not of the standard 'effort incentive' kind.

Another model in which banks reduce the ability to extract private benefits is Almazan and Suarez (2003). They introduce separation of ownership and control on the part of the borrowing firm and investigate the optimal incentive contract for management. One aim of their article is to provide a theoretical rationale for the positive stock price reaction to the announcement of bank loan agreements that is empirically observed.

Manove and Padilla (1999) deal with capital allocation, but the intensity of competition plays no part in their analysis. Rather, they consider the case of overconfident loan applicants:

'It is widely thought that market considerations lead banks to be overly conservative and cautious, discouraging the rational buyer more than is desirable. Here, we demonstrate that in markets with optimistic entrepreneurs, the opposite may well be true.' (p. 326)

This article is less relevant from my point of view for two additional reasons: First, investment scale (which I would like to abstract from) plays an important role in the model setup and, second, the authors consider the role of collateral (which I likewise do not wish to stress). Just as in Manove, Padilla and Pagano (2001), collateral makes banks less interested in the intrinsic quality of the underlying investment project (and might therefore be thought to reduce incentives for screening and monitoring).

Hyytinen (2003) studies ex ante screening by banks, though he neither investigates endogenous screening accuracy nor changes in the intensity of competition and his model does not aim at generating clear-cut implications as to the allocation of capital. Hyytinen focuses on the effects of (1) test results becoming more correlated across banks (actually, it is the two polar cases of no correlation and of perfect correlation that are investigated) and (2) an exogenous increase in the accuracy of the screening technology. The basic setup is the horizontal differentiation model of Salop (1979); banks are assumed not to be able to discriminate by location. All banks are assumed to have access to the same imperfectly informative screening technology where accuracy is parameterized as the probability that an assessment is correct, so the likelihood of making a type I error is exogenously restricted to be identical to that of making a type II error. In terms of results, Hyytinen goes through a number of comparative statics exercises and explores the impact of screening accuracy on

loan margins (decreasing in accuracy) and lending volumes (increasing in accuracy, assuming that the proportion of good types is sufficiently high). He shows that, everything else equal, margins, lending volumes, and profits are higher under independent than under perfectly correlated test results. When banks shift from independent to perfectly correlated tests (say, because of the implementation of an identical credit scoring tool) and the accuracy of the screening technology simultaneously increases, Hyytinen shows that margins, lending volumes, and bank profits unambiguously decrease (assuming that the proportion of good borrowers is sufficiently high).

The changing intensity of competition in Hainz (2003) is captured as pure Bertrand competition, Bertrand competition with cost differences (referred to as oligopoly) and monopoly. The borrower has the choice between exerting high and low (but at any rate non-contractible) effort and in order to satisfy the entrepreneur's incentive compatibility constraint, the interest rate that the lender can demand is bounded above. However, the banks can ask for collateral; a bank with market power can ask for more collateral which allows it to also ask for a higher interest rate without violating the borrower's incentive compatibility constraint. Collateral is uniformly distributed in the economy: The borrower does not have the required liquid assets to finance the investment, but does have other assets that can be used as collateral. In the first part of the paper every entrepreneur's endowment with pledgeable assets is common knowledge. The use of collateral is costly, because the liquidation value of collateral is less than the going-concern value, and may be different for different banks (this is the cost difference between banks). Surplus is driven by the threshold of collateral that is required to be able to borrow (the lower the threshold, the more surplus will be generated) and by the costs of liquidating collateral. When asset endowments are common knowledge, the same projects would be funded under the three different market structures, but surplus is shown to be greatest under pure Bertrand competition, intermediate in the case of Bertrand competition with different costs, and lowest under monopoly because of liquidation costs that are decreasing in the intensity of competition. In the second part of the paper, the asset endowment is unknown to the lender. An entrepreneur can understate (but not overstate) her pledgeable assets. This part of the paper is not about banks' acquiring information about their borrowers, it is an exercise in contract theory. The results with respect to surplus extend to the case where information is asymmetric: Monopoly generates the smallest amount of surplus.

Gehrig and Stenbacka (2001) straddles the literatures on bank competition and amplification and propagation of shocks by the financial system. They propose a mechanism built around banks' screening incentives that generates entirely endogenous cycles: In their model, screening of applicants by banks may worsen the pool of applicants so much (via the negative externality caused by rejection decisions) that banks cease lending until the pool (through an inflow of new good types) has improved sufficiently. The screening technology that banks use to deal with the adverse selection problem (there is no moral hazard) is perfectly accurate in their baseline model, but argued to be robust to imperfect signals in an extension. The overlap with the bank competition literature arises because only a competitive banking system generates screening cycles. A monopolist bank always knows which borrowers it has rejected previously, whereas competitive banks are assumed not to know the rejection histories of their applicants. Ruckes (2004) also examines how the intensity of price competition and screening intensity evolve over the cycle. He endogenizes the banks' screening and discusses type I and II errors, but the focus is clearly on the cyclical aspects.

The lending model in Chan and Thakor (1987) is characterized by both adverse selection and moral hazard. The intensity of competition does not vary since banks earn zero profits under both specifications of competition that the paper considers. In one case, all rents are obtained by borrowers; in the other case, by depositors. The authors allow for the availability of unlimited collateral (assumed to be worth the same to both borrowers and banks) and they investigate how the outcomes (notably the extent of credit rationing, if any) differ under the two types of competition. The authors' main interest is in exploring the robustness of existing results in the rationing literature to the different specifications of competition.

Bouckaert and Degryse (2004) take a closer look at the informational advantage that an incumbent lender is often assumed to have relative to outside lenders. In their model, information sharing may emerge endogenously in the sense that an incumbent lender may voluntarily relinquish her advantage in order to soften competition (for initial market share) ex ante. Incumbents' informational advantage is also the subject of Padilla and Pagano (1997) who show that lenders may commit to sharing their private information with other lenders in order to ease the borrower's fear of exploitation following information lock-in. This makes the borrower more willing to supply effort, increasing the lender's profit in turn.

Hoppe and Lehmann-Grube (2002) and Hoppe and Lehmann-Grube

(2003) embed Broecker's screening model (with exogenous accuracy) in a Hotelling setup. Their interest is not in capital allocation but in endogenizing the banks' location decisions and game-theoretical methodological advances generating pure-strategy equilibria.

Thakor (1996) investigates endogenous costly screening and monitoring in the same model, but concentrates on an analysis of risk-based capital requirements and the impact of monetary policy on lending. Monitoring by multiple banks has more value than being monitored by a single bank.

Winton (1995) examines the case for restricting entry into the banking industry as a by-product of an investigation into more fundamental issues (to do with whether intermediation dominates direct lending); there are no clear-cut results as to the best market structure of the banking sector. The fundamentals of intermediation are also analyzed by Stahl (1988), Yanelle (1989), and Yanelle (1997) with implications that are relevant beyond the field of financial intermediation. One question these contributions address is whether it is a legitimate modeling shortcut to consider in isolation only one side of the market that the intermediary operates in. In general this is not the case for, as Yanelle (1997) points out,

'competition may lead the intermediaries to corner one of the two markets in an attempt to achieve a monopoly outcome.' (p. 215)

Besanko and Thakor (1992) investigate deregulation in a spatial competition model, but their model concentrates on allocative efficiency aspects. Banks take deposits and extend loans, but the results have little bearing on the question of capital allocation in that banks (in contrast to other providers of funds) are simply assumed to be able to observe project returns and are able to distinguish between borrower types. Chiappori, Perez-Castrillo and Verdier (1995) study the effects of a deregulation of deposit interest rates, likewise in a spatial competition model. Winton (1997) also investigates bank competition for deposits. In Purroy and Salas (2000), banks likewise do not compete for loans, but for deposits. The paper is, however, rather interesting in the overall context of this thesis, in that it features a duopoly model where a profit-maximizing bank competes against a bank with 'expense preference' which is here taken to mean that this institution maximizes a weighted combination of profit and labor expenditure. As is consistent with the empirical evidence from the Spanish banking market (which inspired their paper and where not-for-profit savings banks very successfully compete against profit-maximizing commercial banks), the 'expense preference' bank ac-

tually achieves greater profit and market share, even though cost and demand functions are the same for both banks! This counterintuitive result is due to the not-for-profit bank effectively having lower marginal cost. Depending on the nature of competition (Cournot or price competition with horizontal differentiation) the owners of the profit-maximizing banks will design the incentive scheme of the manager either such as to mimic the preferences of the non-profit and behave very aggressively (Cournot case) or go in the opposite direction (price competition with horizontal differentiation).

Gual (1999a) [also see Gual (1999b)] applies the endogenous sunk costs model developed by Sutton (1991) to the European banking market and predicts that as the market size increases with integration, one should not expect to see an increase in concentration.

Dell'Ariccia (2001) shows that when banks cannot distinguish between applicants that have been rejected by a competitor and 'new' projects, the resulting adverse selection problem creates an endogenous fixed cost that limits entry into the banking industry. In the model, learning (of borrower type) is costless and is also referred to as 'learning by lending'. The result hinges on the assumptions of bad projects not leaving the market (although they default on their first lender's loan) and banks not being able to tell apart 'old' bad projects and untested 'new' projects. The same basic adverse selection mechanism built on 'learning by lending' is analyzed in Dell'Ariccia, Friedman and Marquez (1999) who show that this information problem may give rise to what they term a 'natural duopoly'.

Banerjee (2005) studies a banking duopoly in which banks can choose between two technologies for assessing the creditworthiness of applicants. The focus is on the different effects of type I and type II errors on technology adoption incentives as well as the impact of uncertainty with respect to the extent to which the superior technology is indeed more accurate.

Screening is analyzed in the context of a traditional adverse selection model (i.e., the potential borrowers know their type, but the potential lenders do not) by Wang and Williamson (1998). In contrast to standard adverse selection problems, borrowers are risk-neutral here. The authors study a contracting problem rather than bank competition; lenders are implicitly assumed to be competitive. The punchline of the model is that debt contracts and financial intermediaries emerge endogenously. Debt contracts, in fact, are robust to randomization; under certain parameter restrictions a separating equilibrium exists, borrowers self-select and only good types submit to screening with some probability (which al-

lows lenders to establish type against the 'payment' of an effort cost). The authors plausibly argue that lenders incurring costs in the form of screening costs ex ante is a more appropriate description of reality than the ex post auditing costs in Diamond (1984).

Freixas and Rochet (1997) survey a number of bank competition models none of which, however, are of much interest in the context of the present review. For example, the Monti-Klein model (section 3.2, pp. 57-61) is really just a Cournot model in disguise.

The contribution by Schargrodsky and Sturzenegger (2000) was inspired by the evolution of the Argentine banking sector during the 1990s. The authors take issue with the view that stricter capital requirements that reduce the number of banks will necessarily lead to less intense competition. That the relationship between the number of banks and/or concentration measures and the effective intensity of competition should be indeterminate is not news; however, the authors' explanation is novel. They document that as capital requirements in Argentina were increased during the 1990s, the number of banks fell and the 10-bank (deposit) concentration ratio increased substantially. The intermediation margin still fell from 11% to 5%. The authors' punchline is that banks may respond to tougher capital requirements by reducing the extent of horizontal differentiation, thus reducing margins (under certain assumptions with respect to the marginal benefit from differentiation). Alternative explanations are given fairly short shrift even though the evidence that the authors present to document that differentiation has decreased does not strike me as all that convincing. Surplus is considered explicitly and derives from borrowers' valuation of banks' differentiation, roughly corresponding to travel costs. (All entrepreneurs are identical and screening does not play a role.) For some parameter values, tougher capital requirements are shown to both enhance stability and increase aggregate surplus.

2.4.7 Some real-world illustrations

Following this rather technical discussion, I would like to motivate this work by presenting some illustrations of the kinds of real-world issues that the models reviewed arguably address. For example, banks' ability to select profitable projects and effectively monitor borrowers ought to be an important determinant of the rate of return.

The macroeconomic return on capital – while seemingly a straightforward concept – is actually tricky to calculate. The technical difficulties are compounded when one seeks to compare rates of return internation-

ally; for example, because of different conventions and assumptions employed in different sets of national accounts (which constitute the basis for the estimation). Walton (2000) and Citron and Walton (2002) provide excellent introductions and also state rates of return for a number of countries. For the years 2000 or 2001 (depending on data availability), rates of return ranged from 20.8% in Norway via 11.6% in the UK, 6.9% in the US, and 6.5% in Germany to 0.8% in Iceland. [Citron and Walton (2002), p. 21] One would of course expect there to be several major factors influencing rates of return, but banks' ability to select projects and monitor borrowers likely is one of them.

Agrawal et al. (1996) compare returns on capital for the US, Japan, and Germany. What makes their study particularly interesting is the detailed look at five industries in addition to the aggregate analysis. These case studies suggest that – at least in the industries considered – different degrees of management sophistication (rather than factors not under the control of management, such as more or less intrusive labor market regulation) explain a large part of the observed differences. For example, high capacity utilization driven by astute marketing and pricing strategies will increase the return on capital. Management quality may, of course, be driven by any number of factors, but it does not seem too far-fetched to view banks' financing decisions as one contributing factor (poor management being starved of funding and conversely for good management). Monitoring by banks may also help to keep management on the straight and narrow. From personal observation, I would argue that banks aid in the diffusion of effective management practices through their regular discussions with borrowers.

If anything, the rate of return on capital might be an even greater concern for developing countries. For example, a recent article in The Economist (2005a) investigates

'the argument that India is much more efficient than China at using capital. Having invested an average of 22-23% of GDP for a decade, it has seen average economic growth of about 6% a year in real terms. China has invested twice as much, but its average growth rate has been only about 50% higher than India's. It is indeed staggering how much investment is needed to power Chinese growth. (...) India, meanwhile, is grappling with (...) how to raise investment rates. To emulate China's growth, India would need to increase its investment to 30-35% of GDP, and there is little sign it can do that. (...) Suman Bery, head of the National Council for Applied Economic Research in Delhi, says both countries face the same challenge: a failure of financial intermediation. Neither

has found efficient ways to translate high rates of private savings into productive investment. China's answer has been FDI, leading to what Mr Bery calls the "idiocy" of Chinese peasant savings financing the American Treasury. In India, the consequences have been underinvestment and a big fiscal deficit, financed by the banking system.'

To be fair, one ought to note that Chinese (state-owned) banks are still to some extent tools of fiscal policy rather than independent entities with a hard budget constraint, so perhaps the above passage is not comparing like with like[12]. The fact remains that where banks fail in their task of allocating capital, the (macroeconomic) implications are considerable. Of course, it is not only banks that count since they allocate only a portion of capital. However, I only meant to illustrate how banks matter for the allocation of capital and why, therefore, the intensity of competition between banks may have a bearing on a number of first-order real-world issues.

2.5 Summary of the Theoretical Literature

In the following I briefly summarize and comment on the various mechanisms linking the intensity of competition and surplus proposed in theoretical work. There is no consensus in the literature as to their relative importance and results are often ambiguous and critically dependent on parameters that do not necessarily have clear-cut interpretations. Still, it seems fair to say that the literature overall is not enthusiastically in favor of more intense bank competition and considers competition-increasing interventions with a degree of scepticism. What is this scepticism built on?

First, the model setup may ensure that monopoly is the optimum market structure. This is easily seen by recalling a number of common assumptions: With a perfectly elastic supply of deposits and perfectly inelastic demand for loans allocative efficiency is not a concern. When cashflow distributions are invariant (i.e., higher interest rates do not, for example, worsen moral hazard on the part of the borrower) lending interest rates merely redistribute surplus between lender and borrower. A profit-maximizing monopolist is then practically guaranteed to maximize surplus: Screening incentives depend on how profitable it is to assess project type; a monopolist can ask for the entire surplus (unproblematic in a setting where distributional considerations do not matter). More

[12]On China's banks, see Lardy (1998) whose analysis is unlikely to have become superseded by developments in the last few years.

intense competition leads to lower margins and thus less incentive to exert screening effort which may adversely affect surplus. In the extreme, adding even a second lender may completely destroy incentives to screen; for example, if test results become public. (The same argument in principle applies to incentives for monitoring, although for monitoring to have a rationale, one would typically have to relax the assumption that the cashflow distribution is invariant.)

A related effect encountered in the literature is the presumed loss of surplus due to the (inefficient) duplication of testing costs: Each test can be thought to consume resources and may lead to additional good and bad projects being funded. The net effect is ambiguous although in theory at least it is possible that competition leads to a socially excessive amount of testing and information acquisition. A lot hinges on the exact specification of the screening technology. Barring different assumptions, it should be noted that type I errors [incorrectly rejecting a positive NPV project] and type II errors [incorrectly accepting a negative NPV project] are logically distinct. However, normally – and quite plausibly – a single variable is used to parameterize screening accuracy and determine both type I and II errors simultaneously. In that case, the way in which the two types of errors are assumed to be related may well drive the results.

However, an additional effect [which appears, for example, in Schnitzer (1999b)] may offset the diminished incentive to exert effort somewhat: Assuming that more intense competition worsens the composition of the pool of potential borrowers, it might become more worthwhile to filter out bad projects. In other words, as in Broecker (1990), the competitors' decisions create a negative externality, but this might induce a lender to exert more effort rather than less and possibly increase surplus.

The literature hardly considers the possibility that lower lending rates might alleviate a moral hazard problem on the part of the borrower [Schnitzer (1999a), (1999b), and Caminal and Matutes (2002) are exceptions.] Even with an invariant cashflow distri-bution, the default probability should be increasing in the interest rate [leading to higher bankruptcy costs from 'auditing' in Guzman (2000b)]. However, much of the literature models borrowers' default probability as independent of the interest rate.

Where banks obtain costless signals of exogenous quality, free-riding on others' information acquisition and reduced incentives to exert effort as margins shrink are not a concern. However, as I show in the companion chapter extending Broecker (1990), the properties of the equilibrium still depend critically on whether signals are private or public informa-

tion. When tests are less than perfectly informative and test results are less than perfectly correlated, private signals give rise to the winner's curse effect and lead to the non-existence of pure-strategy equilibria. Pure-strategy equilibria may obtain if test results are public. But what are the implications for surplus? Additional competitors worsen the winner's curse. This worries some authors even though the effects on surplus are far from clear-cut: The average creditworthiness of applicants may decrease, expected interest rates may increase and banks may become more cautious, but surplus might nevertheless increase. Still, a worsening of the winner's curse due to increased competition is held by some authors [e.g., Riordan (1993)] to be undesirable. At any rate, modeling information quality as exogenous is both empirically less plausible and conceptually less interesting than endogenous accuracy.

The discussion so far centers on screening because that is what the literature concentrates on. I do, however, wish to mention some of the other channels discussed in the literature: Competition might make lenders less willing to help borrowers in a liquidity crisis which may be inefficient from a surplus perspective [Dinç (2000)]. Competition may also affect the extent of the socially harmful use of collateral [e.g., Villas-Boas and Schmidt-Mohr (1999)] and the higher margins associated with less intense competition may allow banks to intertemporally subsidize their borrowers with potentially beneficial effects on surplus [Petersen and Rajan (1995)].

All in all, the theoretical literature is thus sceptical with respect to the benefits of more intense competition. As I will show in the following section, the empirical work, painting with a broad brush, finds more intense competition to be rather beneficial. The possibility that the two literatures are not asking the same question aside, this contradiction ought to be resolved. I think that the major omission in the existing theoretical literature may be the failure to look at the role of competition as a managerial discipline device when other corporate governance devices are not functioning properly. I explore this idea in chapter 5 below.

2.6 Bank Competition and Capital Allocation - Empirical Work

In the present section I would ideally like to accomplish two things: Review the evidence on the effects of more intense bank competition on capital allocation in general and, specifically, take the theoretical work reviewed above to the data. Unfortunately, the second task can only be

tackled indirectly. Very few of the papers provide evidence on the particular theoretical model proposed. What empirical work there is does not look at the various hypotheses outlined in the preceding section directly. For example, a substantial part of the theoretical literature predicts that lower margins lead to less investment in information acquisition and that this reduces surplus. The empirical literature investigating this specific issue is, however, sparse indeed. It is difficult to even think of an operational proxy for information acquisition. The empirical work reviewed below tends to look at different outcome measures (e.g., local economic growth) and cannot easily be construed as a test of a hypothesis regarding a particular channel. So while ideally I would like to not only demonstrate which mechanisms proposed in theoretical work 'matter' but also quantify their relative importance, I am not aware of work on which such firm conclusions could be based. In the following I review the evidence and I extract from the available findings what is relevant for an assessment of the theoretical work surveyed above.

I should note two related complications: First, many of the theoretical models surveyed are implicitly models of lending to small and medium-sized businesses, but the empirical work surveyed here does not focus on this sub-market. Second, the theoretical literature omits allocative efficiency considerations from the analysis whereas empirical work takes these into account. This leads Allen et al. (2001), for example, to assert that the two literatures are not asking the same question (see p. 6). I will argue below that that assertion is overly pessimistic and that recent empirical work does indeed – albeit indirectly – address similar issues as the contributions reviewed above.

One plausible approach is to posit that a causal chain links policy measures with outcomes via the intensity of competition. For example, a country might liberalize market access; competition becomes more intense and – controlling for other possible explanations (cyclical factors, technological advances etc) – loan losses decrease. If such a study were carefully done, it might indicate (not prove) that competition has pushed banks to screen applicants and monitor borrowers more carefully. The work discussed below does not investigate the entire chain but only the link between, on the one hand, either policy measures (postulated, but not demonstrated to be competition-enhancing) or measures of the intensity of competition and, on the other hand, outcomes.

Aside from the liberalization of market access, policy measures of interest might include, for example, an abolition of branching restrictions or the phasing-out of privileges for state-owned banks. As for the inten-

sity of competition, Shaffer (1994) notes that

'Economists and policymakers would like to be able to look at a single number (such as profitability, price level, bank size, or number of banks in the market) for a bank or for a market, compare it with the value that would occur in a perfectly competitive market, and conclude something about the degree of competition or monopoly power in the bank or market in question. Unfortunately, it is very difficult to establish a reliable index for this purpose.' (p. 6)

I discuss the difficulties associated with the measurement of competition in Eggenberger (2006a). A banking-specific complication that I have not seen discussed elsewhere arises from the fact that interest rates also contain risk premia; intermediation margins are therefore highly problematic indicators of market power. A high interest rate may be due to the exercise of market power or the riskiness of the borrower and without access to the highly confidential risk assessment of the bank it is not possible to say which it is. In particular, even ex post profitability is not a good indicator unless one looked at an entire credit cycle: Risk premia collected in good times are needed for economic downturns, so a few very profitable years may simply reflect the squirreling away of funds to cover future loan losses. Intermediation margins might be particularly misleading if liberalization leads banks to start lending to hitherto neglected higher-risk borrowers: Margins would indicate a reduction in competition when, in fact, competition is effective and has highly desirable effects. Roe (2001) observes that rents are typically shared and partly spent on inflated salaries and/or staff numbers which will affect other types of margins.

However, in line with the earlier comments, the outcomes studied are the greatest source of concern. As noted, theoretical work ignores allocative efficiency. If this were the only outcome that empirical work looks at, I would be guilty of comparing apples with oranges. However, while the two literatures are not asking exactly the same questions, recent empirical work does test the models surveyed, albeit rather indirectly. Unfortunately, the variables examined in earlier work [notably profits and prices (see above)] do not really allow one to assess whether more intense competition increases surplus beyond improvements in allocative efficiency. Before I discuss more recent research, I should note that Shaffer (1994) provides a survey of earlier work [mainly, but not limited to, research in the structure-conduct-performance (SCP) tradition]. Neuberger (1998) also skims the empirical literature based on the SCP paradigm. In their discussion of U.S. anti-trust policy in the banking industry, Gilbert and

Zaretsky (2003) provide a table summarizing the results of about two dozen SCP studies. Their main interest is in finding out whether the empirical work supports the methodology employed by U.S. competition authorities (it mostly does, they claim), but their analysis provides useful background.

Fortunately, more recent work addresses these concerns. For example, Berger et al. (2004) note that

'The types of economic consequences of market structure that are now examined in the research literature have expanded well beyond simple price and profit measures.' (p. 439)

They give the following examples:

'Investigators have expanded the research agenda to include analysis of the effects of concentration and competition on economy-wide growth, credit availability to SMEs, and the performance of nonfinancial industries.' (p. 439)

Other variables that might capture what I am interested in are outcomes such as, for example, direct estimates of (X-in)efficiency or loan losses.

I should note that a small number of papers argue that the question is more complicated than 'Is competition good or bad?' Cetorelli (2001) presents evidence which, he argues, suggests that the degree of competition in the banking industry may have heterogeneous effects in different industries and at different stages in the life-cycle of borrowing firms. Overall, he claims,

'the empirical findings suggest that banking market structure has both positive and negative economic effects, and it is hard to establish which one ultimately dominates.' (p. 41)

Cetorelli and Gambera (2001) use a multi-industry, multi-country dataset and show that on average bank concentration is bad for industry growth. However, they also apply the Rajan and Zingales (1998) methodology and show that industrial sectors more dependent on external funding grow faster where the banking sector is more concentrated and that this is true a fortiori for younger firms. Their results are thus consistent with the view that, once we move beyond allocative efficiency considerations, less intense competition may have something to commend it. Claessens and Laeven (2005), however, also apply the Rajan and Zingales (1998) technique; their measure of competitiveness

'is positively associated with countries' industrial growth, suggesting that more competitive banking systems are better at providing financing to financially dependent firms.' (p. 181)

Bonaccorsi di Patti and Dell'Ariccia (2004) look at the relationship between concentration and new firm creation in Italian data. They find a positive (though quantitatively small) association between new firm creation and concentration and they further argue that this effect is more pronounced for sectors that they consider especially informationally opaque. These results might be interpreted as support for the view that competition destroys incentives for information acquisition. Notable among those sceptical of more intense competition are Petersen and Rajan (1995) who (for a cross-section of small U.S. firms) find that less intense bank competition [as proxied by the Herfindahl index of the local deposit (sic) market; while this choice of proxy is problematic, the authors offer a thoughtful defense] means easier availability of funding for young firms. Note that the predictions of their model are not consistent with evidence reported, for example, by Berger and Udell (1995) that suggests that loan terms get more favorable for the borrower as the lending relationship ages. Zarutskie (forthcoming) finds that deregulation has different effects on young and old firms. If one accepts the assumption that age proxies informational opaqueness, her results also might be considered consistent with the Petersen and Rajan (1995) hypothesis.

While these contributions suffer from measurement and other problems (using concentration indicators to measure the intensity of competition seems to be a particular problem), competition almost certainly has multiple effects and I do not doubt that some of these may be surplus-reducing ones. On balance, however, I would argue that the negative findings are fairly robustly contradicted by much other work.

In a series of papers [Jayaratne and Strahan (1996), Jayaratne and Strahan (1998), Jayaratne and Strahan (1997), Jayaratne and Strahan (1999)], Jayaratne and Strahan found that U.S. branching deregulation had positive effects: Non-interest costs (including wages) fell and so did loan losses. Cost reductions were passed on to borrowers. Intriguingly, economic growth accelerated. The finding on loan losses in particular suggests to me that competition has beneficial effects beyond improvements in allocative efficiency. Indeed, Jayaratne and Strahan (1996) note that

'Improvements in the quality of bank lending, not increased volume of bank lending, appear to be responsible for faster growth.' (p. 639)

They argue that a selection effect is responsible for these results: Deregulation allowed better banks to expand at the expense of less efficient ones. Working with U.S. data, with states deregulating at different times, avoids a host of econometric problems associated with cross-

country datasets and the authors' robustness checks strike me as uncommonly thorough and ingenious. Black and Strahan (2002) find that following the abolition of branching restrictions, the rate of new business incorporations increases. This result is confirmed by Cetorelli and Strahan (2006) who argue that

'the empirical evidence (...) strongly supports the idea that in markets with concentrated banking, potential entrants face greater difficulty gaining access to credit than in markets in which banking is more competitive.' (p. 437)

Cetorelli (2003) concludes that

'More competition in banking appears to promote job creation among industrial establishments at the start-up stage and to permit them to prosper in the immediate wake of their entry into the market. At the same time, more bank competition accelerates the exit of more mature establishments from the market. These results are consistent with theories suggesting that banking market power may represent a financial barrier to entry in product markets.' (p. 144)

Surveying a batch of recent contributions, Berger et al. (2004) conclude that

'The new research distinguishes between concentration[13] and competition and generally finds that bank competition is good from a social perspective.' (p. 444)

More specifically,

'The findings with respect to the measures of competition other than concentration are generally robust. More regulatory restrictions on bank competition are associated with bad outcomes—such as less favorable prices for customers, less access to credit, and reduced stability of the financial system. Less binding impediments to foreign bank ownership and entry are generally associated with more favorable prices for customers and more access to credit (good). State bank ownership is generally associated with less access to credit and reduced financial system stability (bad). Thus, policies that restrict bank competition—regulation, barriers to foreign bank participation, and direct state control of banking resources—tend to be associated with bad outcomes and diminished overall economic performance.' (p. 445)

Berger et al. (2004) note that

'Other restrictions on competition, such as barriers to new bank entry, interstate banking prohibitions, and implicit or explicit limits on cross-border banking may have the same unfavorable or favorable effects on

[13]Concentration measures have long been known to suffer from severe methodological problems, see Eggenberger (2006a)

SME credit, industry performance, and economic growth.' (p. 439)

In his survey-plus-original-work on recent U.S. financial deregulation, Strahan (2003)

'focuses on how one dimension of this broad-based deregulation—the removal of limits on bank entry and expansion—affected economic performance.' (p. 111)

Strahan argues:

'In a nutshell, the results suggest that this regulatory change was followed by better performance of the real economy. State economies grew faster and had higher rates of new business formation after this deregulation. At the same time, macroeconomic stability improved. By opening up markets and allowing the banking system to integrate across the nation, deregulation made local economies less sensitive to the fortunes of their local banks. (...) the results support the idea that competition and openness in financial markets are beneficial. This finding is accepted when applied to industrial firms—for most economists, free trade and competition are akin to motherhood—but it is much less accepted when applied to the financial sector.' (p. 111)

Strahan also briefly discusses the political economy considerations explaining the timing of deregulation. However, his main findings of interest here are the impact of competition on the growth rate and the rate of business creation. While the first effect probably captures improvements in allocative efficiency to some extent (Strahan does not provide data on interest-rates or margins, so this is hard to gauge), the second effect can plausibly be interpreted as a reduction in type I errors (i.e., fewer worthwhile projects not obtaining funding).

The contributions cited do not conclusively analyze the mechanisms through which more intense competition brings about the observed changes. While clearly not an exclusive explanation, I conjecture that a reduction in managerial slack has a first-order effect. Berger et al. (2004), for example, note that

'(...) research found that banks in more highly concentrated local U.S. markets have lower cost efficiency, presumably because of reduced effort or pursuit of other goals by managers when competition is lax.' (p. 439)

Berger and Hannan (1998) look squarely at the effect of competition on measured X-inefficiency. The authors note that

'Extrapolating these estimation results to the entire U.S. banking industry, we find that the additional operating cost attributable to market concentration appears to be several times larger – perhaps twenty times

larger – than the social loss due to the non-competitive pricing of bank outputs, as measured by the welfare triangle. (...) The dominance of the efficiency cost over the social loss associated with mispricing is robust with respect to a large number of variations in samples, specifications, estimation techniques, and controls for alternative explanations.' (p. 464)

Claessens and Klingebiel (2001) likewise note that

'Banks operate below their technical possibilities, and measures that would induce financial institutions to act efficiently from the standpoint of costs are essential. The most important of these measures is competition, which has been found to affect performance measures in a wide variety of countries.' (p. 20)

Park, Brandt and Giles (2003) in the empirical part of the paper discussed above find

'positive effects of competition on effort and financial performance.' (p. 463)

Guiso, Sapienza and Zingales (2003) study Italian data. One of the punchlines of their paper is that

'banks, in a non-competitive environment, become lazy and inefficient in screening customers and this inefficiency is exposed after liberalization.' (p. 2)

While this brief review can hardly claim to be the last word on bank competition, I conclude the discussion of my central claim – that more intense bank competition is beneficial – here. In the following section I present evidence on the extent of X-inefficiency and managerial slack in the banking sector – the reduction of which I conjecture to be an important effect of more intense bank competition.

Several additional strands of empirical work have some bearing on the subject of bank competition. One line of work that ought to be of interest to theorists analyzes how banks actually compete. Degryse and Ongena (2004) provide an overview and, in particular, discuss the role of distance and spatial competition. In a separate contribution, the authors present evidence on spatial price discrimination using Belgian data [Degryse and Ongena (2005)]. Kim, Kristiansen and Vale (2005) discuss evidence from Norwegian banks that suggests that banks might engage in vertical differentiation; a bank with lower loan losses might be able to ask for a higher interest rate as borrowers wish to signal their creditworthiness. Greenwich Associates (2002) provide evidence that the interest rate is not all that matters in lending. Evidence of price dispersion in loan pricing for given risk is provided in Oliver Wyman & Company (1996) and Oliver Wyman & Company (2000). A separate, but related strand of

work aims to determine how competitive the banking industry actually is. Examples are Bikker (2003) and Bikker and Haaf (2002b). Bikker (2003) applies the Bresnahan methodology to a number of EU loan and deposit markets and finds limited evidence of market power. Using the Rosse-Panzar technique, Bikker and Haaf (2002b) find evidence of monopolistic competition. The authors discuss some of the methodological underpinnings for measuring competition in Bikker and Haaf (2002a). Cetorelli (1999) outlines (and criticizes) U.S. regulators' current approach to bank competition issues revolving around concentration measures and presents Bresnahan's markup technique as an alternative (along with an application to Italian data). A discussion of economic and legal issues to do with the market definition in [U.S.] banking markets is offered by Amel and Starr-McCluer (2002). One additional important topic is consolidation and its implications for competition. Berger, Demsetz and Strahan (1999) review the U.S. experience, the Group of Ten (2001) provides an exhaustive study of consolidation in rich countries, and the International Monetary Fund (2001) looks at financial sector consolidation in emerging markets. Finally, an important question from a policy perspective is whether there is something special about foreign banks, foreign bank entry, and cross-border banking: Berger et al. (2000) study cross-border banking. A volume edited by Litan, Masson and Pomerleano (eds) (2001) collects a number of articles on various aspects of foreign participation in developing countries' financial markets which is also the topic of a chapter in an IMF International Capital Markets Report [International Monetary Fund (2000)].

2.7 X-Inefficiency in Banking

In the following I briefly present evidence showing that banks suffer from X-inefficiency in the sense of Leibenstein (1966): Apart from allocative inefficiencies due to price distortions, market power may create scope for managerial slack that leads to resources being wasted (e.g., by not minimizing costs). While the idea is conceptually simple, measuring such inefficiencies is involved. Frontier techniques are the workhorse used by most of the literature [see the survey by Murillo-Zamorano (2004)]. Frontier techniques for measuring (in)efficiency come in many varieties, all of which go back to Farrell (1957). The basic idea is to envisage a technology isoquant and ask how far the input bundle chosen by a producer is from this (by definition efficient) frontier and, furthermore, how far the input bundle chosen is from the cost-minimizing input bundle (which is

not only technically, but also allocatively efficient). Looking at efficiency from the input and cost-minimization perspective as sketched here is equivalent to taking an output and revenue-maximization perspective as long as returns to scale are constant. Murillo-Zamorano (2004) provides an overview of the various techniques that this basic idea has given rise to. Parametric techniques impose functional forms on the efficient frontier whereas under non-parametric approaches [notably Data Envelopment Analysis (DEA)] a functional form is estimated from the available data. Murillo-Zamorano further distinguishes between deterministic parametric techniques (which have the drawback that all deviations from optimality are attributed to the agent) and stochastic parametric techniques which allow for specification failures and exogenous factors. Murillo-Zamorano points out that where different techniques have been applied to the same data, the results are not always mutually consistent.

Bauer et al. (1998), for example, find such a lack of mutually consistent results in banking-related efficiency studies. They go beyond mutual consistency of the various techniques, provide a broader set of consistency criteria that include what they term "believability conditions" (see p. 87), and specifically study the banking industry. The authors compare DEA with three parametric techniques (stochastic frontier, thick frontier, and distribution-free approaches) and give concise, non-technical descriptions of the various techniques. The parametric techniques are found to yield similar results that differ somewhat from those obtained with DEA.

While I am mainly interested in X-inefficiency, the techniques outlined above have been deployed to assess economies of scale and scope, product mix, and risk diversification. Clearly, one would not want to erroneously attribute efficiency gains due to, say, scale to management quality, so alternative drivers of efficiency need to be considered. One particular concern in banking is the effect of risk. The case for including risk considerations in the estimation of productive efficiency is made by Hughes and Mester (1998) and Hughes, Mester and Moon (2001). These authors find that, after explicitly taking into account risk, there may be large scale economies in banking (whereas the conventional wisdom in the field is that scale economies are exhausted at very small bank sizes of maybe as little as one billion dollars in assets). Vives (2001b) notes:

'A problem arises in that empirical studies of scale economies and cost efficiency typically do not account for risk. Indeed, the studies measure the effect on cost of the joint increase in scale and risk. Noting that the lower cost of risk management of a larger, better-diversified bank may induce the bank to take on more risk, cost savings may not then

be detected, if to take on more risk is costly. Controlling for risk-taking, large economies of scale that increase with asset size are found in U.S. banks.' (p. 540)

However, to the best of my knowledge even the presence of sizeable scale economies would not invalidate the finding that banks suffer from considerable X-inefficiencies.

A large literature applies different techniques to assess economic efficiency in the banking industry, although as Bauer et al. (1998) point out, these contributions do not usually cross-check the results by using several techniques. In the following I present findings drawn from a number of surveys of this literature.

Bauer et al. (1998) note that

'Frontier inefficiency or X-inefficiency of financial institutions has generally been found to consume a considerable portion of costs on average, to be a much greater source of performance problems than either scale or product mix inefficiencies, and to have a strong empirical association with higher probabilities of financial institution failures over several years following the observation of substantial inefficiency.' (p. 86)

Neuberger (1998) has a brief section on X-inefficiency and observes that

'(...) X-inefficiencies, brought about by lacking managerial ability to control costs or maximize revenues, have been found to be especially important in banking. They account for at least 15-25% of costs (...)' (p. 109)

Saunders (1997) (reviewing the empirical evidence for the U.S.) writes:

'Efficiency studies find quite dramatic cost differences among banks (...). [T]hese studies find that sometimes as little as five per cent of the cost differences among financial intermediaries (...) can be attributed to economies of scale or scope. This suggests that cost inefficiencies related to managerial ability and other hard-to-quantify factors (so-called X-inefficiencies) may better explain cost differences (...) among financial firms (...).' (p. 262)

Berger, Demsetz and Strahan (1999) includes a section on efficiency and surveys a number of articles (pp. 157–165). In contrast to most of the literature cited thus far, Schure, Wagenvoort and O'Brien (2004) work with European data and find that

'Managerial inability to control costs (X-inefficiency) is with 17–25% [of costs] the main source of bank inefficiency in the EU.' (p. 371)

An earlier strand of work investigates expense preference behavior by banks. Important contributions in this line of research [started by

Williamson (1963)] are Edwards (1977), Hannan (1979), and Hannan and Mavinga (1980). Informal, but nevertheless very interesting evidence on agency conflicts in banks and their implications for various aspects of performance is provided by Davis (1995) for the case of the Australian banking industry and Hanazaki and Horiuchi (2003) for the case of Japan, where entrenched management is said to have actively pursued the 'ever-greening' of delinquent loans with disastrous consequences for the banks themselves, healthy real-sector firms, and the macroeconomy.

Banks suffer from particular corporate governance problems (opaqueness, accounting discretion making it easy to manipulate results, regulatory restrictions on takeovers etc) which may explain the presence of substantial X-inefficiencies. Berger and Hannan (1998) tackle head-on the hypothesis that relatively low intensity of competition is what drives inefficiencies and explicitly set out to test the 'quiet life' hypothesis:

'The basic hypothesis tested is that the market power exercised by firms in concentrated markets allows them to avoid minimizing costs without necessarily exiting the industry. The reduced pressures to minimize costs may result in lower cost efficiencies for firms in concentrated markets through one or more of several mechanisms – shirking by managers, the pursuit of objectives other than profit maximization, political or other activities to defend or to gain market power, or simple incompetence that is obscured by the extra profits made available by the exercise of market power. Using detailed estimates of the efficiency of over 5000 banks and measures of local market concentration [for the markets] in which these banks operate, we find strong evidence that banks in more concentrated markets exhibit lower cost efficiency.' (p. 464)

While concentration measures are a dubious indicator of market power, the results are nonetheless suggestive.

All in all, substantial evidence of X-inefficiency in banking (probably due to particularly severe corporate governance problems in that industry) is consistent with the hypothesis that more intense bank competition has beneficial effects primarily because it reduces managerial slack.

Chapter 3

On Broecker's 'Creditworthiness Tests and Interbank Competition'

3.1 Introduction

The allocation of capital is an important determinant of the growth rate of an economy. Whilst there are various sources for the financing of investment (the most important being retained earnings), banks (financial intermediaries) play an important role even in highly disintermediated economies. Small and medium-sized enterprises in particular usually do not have access to external finance from sources other than banks (and their trade creditors). How accurately banks assess their clients' creditworthiness thus has considerable implications for the economy at large. Given this perceived externality it is not surprising that the regulation and performance of banks is a key public policy concern in most countries. The present chapter focuses on how the intensity of competition between banks affects how well they do their job. Since the intensity of competition can be affected by government policy (e.g., via the number of banking licenses that are issued), this question is of great interest to policymakers and there is now a considerable body of evidence (discussed in the literature review) that suggests that eliminating restrictions to competition between financial intermediaries leads to an improved capital allocation.

However, the theoretical literature in this area is concerned less with the potentially problematic lack of competition than with the risk of 'too much' competition. Thorsten Broecker's paper 'Credit-Worthiness Tests and Interbank Competition' [Broecker (1990)] started one strand of the literature and has inspired many subsequent authors. Claims that more

intense competition is detrimental are often made with reference to that article.

The purpose of the present chapter is to investigate the key result of Broecker's paper, the non-existence of a pure-strategy equilibrium in a certain type of banking game, and to extend it to the case of asymmetric banks. In the original model (which assumes that banks' creditworthiness assessments are equally accurate), Broecker conjectured that if the banks were asymmetrically well-informed, only the most accurate bank would stay in the market and a pure-strategy equilibrium would obtain. In the following, this conjecture is shown not to be correct. Even if banks are asymmetric, interest-rate competition does not lead to an equilibrium in pure strategies. The intuition underlying this result is shown to be the same as that underlying the result in the symmetric case: A bank always finds it profitable to either undercut the rival by an epsilon or to charge the highest possible interest rate.

3.2 Broecker's 'Credit-Worthiness Tests and Interbank Competition'

Two related questions are of interest with regard to the allocation of capital: Do those projects that should be financed obtain funding? And conversely, do those projects that should not be financed not obtain funds? Broecker (1990) provides a model that allows a first stab at answering the question how the intensity of competition between intermediaries will affect the allocation of capital. Because so many of the subsequent papers in this literature have been inspired in one way or another by Broecker's model and its assumptions and it will be referred to throughout the present paper, it seems appropriate to summarize his contribution here[1].

3.2.1 Applicants

Banks compete for a continuum [0,1] of risk-neutral loan applicants who need the loan to be able to invest in their project. These projects are - following an investment of one unit of capital at date 0 - assumed to have a discrete cashflow distribution with the following returns at date 1: Each project will be successful and have a return of $R^{MAX} >> 1$ with probability p_i (where subscript i denotes the type of project) and will fail and have a return of zero with probability $1 - p_i$. Broecker assumes that

[1]The results of Broecker's two-stage-game, discussed in section 3 of his paper, are of limited interest for the present purpose and are therefore omitted.

there are two types of applicants: A proportion γ, $0 < \gamma < 1$, of good projects with a relatively high success probability and bad projects with a relatively low success probability which make up a fraction $(1-\gamma)$ of the applicant pool. γ is common knowledge. To simplify the analysis I will in the following assume that

$$p_i = \begin{cases} 1 & for\ i = Good \\ \beta & for\ i = Bad,\ 0 < \beta < 1 \end{cases} \tag{3.1}$$

This assumption will not substantially change anything, but it will get rid of the p_i and lighten notation.

ASSUMPTION 1: A bad project makes losses in expectation:

$$\beta R^{MAX} < 1 \tag{3.2}$$

In other words, a good project is always successful whereas a bad project is on average not profitable. Note that the applicants are extremely passive in this game. Each applicant demands one and only one loan of a fixed size of one unit of capital, but demand for the loan is totally interest-inelastic. Any borrower will simply accept the lowest interest rate offer, if they have any offers. For simplicity it is assumed that an applicant offered an interest rate so high that her expected return is zero is assumed to accept the loan nevertheless and that an applicant with several identical best interest rate offers will choose each with equal probability. If an applicant manages to obtain funds, she will invest them and obtain the realized cashflow at date 1. Moral hazard is not an issue as the cashflow realization is completely independent of entrepreneurs' effort and decisions. If their project is successful they will repay the bank. (In fact borrowers are so passive that it does not matter whether they know their own type or they do not.) A critical assumption is that borrowers do not have bankruptcy costs. Bankruptcy costs for borrowers would have an effect similar to application fees in that they affect good borrowers differently from bad borrowers. Bankruptcy costs together with the assumption that borrowers know their type would change the results completely. Broecker assumes that no self-selection devices are available, so that even if applicants knew their own type, there would be no way in which the banks could extract this information. Up-front application fees may not be used (and would not make much sense anyway since the applicants are assumed not to have any funds of their own). Collateral is assumed not to be available. Instead, the bank has to rely on its screening of the potential borrowers.

3.2.2 Banks

There are $n \geq 2$ risk-neutral banks that compete in interest-rates. Prior to the interest-rate competition stage, banks screen their applicants and obtain costless and imperfectly informative signals about an applicant's type. The screening is conceptually quite straightforward: Each bank subjects each applicant to a credit-worthiness test that either has the result {This applicant is of good type} which is denoted g or {This applicant is of bad type} which is denoted b. The true types are abbreviated in uppercase; i.e. a truly good applicant may be written as G and a truly bad applicant can be written as B. As these tests are not perfect a good applicant may be categorized as bad and conversely. If the statement {This applicant is good} is defined as the Null hypothesis, then to categorize a good applicant as bad is a type I error and to categorize a bad applicant as good is a type II error. It should be noted that there is nothing that suggests that the likelihood of making one type of error affects the likelihood of making the other type of error. However, note that plausibility suggests reasonable restrictions as in Assumption 2. The testing technology is such that a project that is truly good (truly bad) has a higher probability of being categorized as good (bad) than a bad (good) project. In particular, the probability that a good (bad) project is categorized as good (bad) is greater than γ $(1-\gamma)$, i.e. the tests are at least somewhat informative:

ASSUMPTION 2:

$$\Pr ob(g|G) = q_G > \gamma > 1 - q_B = \Pr ob(g|B) \tag{3.3}$$

This may equivalently be stated as

$$\Pr ob(b|B) = q_B > 1 - \gamma > 1 - q_G = \Pr ob(b|G) \tag{3.4}$$

The notational convention used is thus that q_i is the probability of being correctly identified as being of type i whereas $1 - q_i$ is the probability of making a type I or type II error respectively. Errors are assumed to be independent among banks (the results would still go through with correlated errors unless the correlation were perfect). Consider the composition of the groups with good and bad signals, respectively, after a single bank has screened all applicants. The group with good signals will consist of $q_G\gamma$ good types as well as $(1-q_B)(1-\gamma)$ bad types. The group with bad signals is composed of $(1-q_G)\gamma$ good types and $q_B(1-\gamma)$ bad types. Knowing the composition of the two groups, it is possible to obtain their respective average success probabilities (i.e., the average probabil-

ity that the loan will be repaid). For the group with good signals, that average success probability is $\frac{q_G\gamma+(1-q_B)(1-\gamma)\beta}{q_G\gamma+(1-q_B)(1-\gamma)}$. This expression is greater than the overall average success probability $\gamma+(1-\gamma)\beta$ which in turn is greater than the average success probability of only those applicants assigned to the bad group, $\frac{(1-q_G)\gamma+q_B(1-\gamma)\beta}{(1-q_G)\gamma+q_B(1-\gamma)}$. To see this, note that

$$\frac{q_G\gamma+(1-q_B)(1-\gamma)\beta}{q_G\gamma+(1-q_B)(1-\gamma)} > \gamma+(1-\gamma)\beta$$

$$\iff q_G\gamma+(1-q_B)(1-\gamma)\beta >$$

$$> q_G\gamma^2+(1-q_B)(1-\gamma)\gamma+q_G\gamma(1-\gamma)\beta+(1-q_B)(1-\gamma)^2\beta$$

$$\iff (...) > q_G\gamma[1-(1-\gamma)]+(1-q_B)(1-\gamma)\gamma$$

$$+q_G\gamma(1-\gamma)\beta+(1-q_B)(1-\gamma)\beta-(1-q_B)(1-\gamma)\gamma\beta$$

Deleting terms on both sides of the inequality yields

$$0 > -q_G\gamma(1-\gamma)+(1-q_B)(1-\gamma)\gamma+q_G\gamma(1-\gamma)\beta-(1-q_B)(1-\gamma)\gamma\beta$$

Grouping terms, this can be rearranged as

$$q_G\gamma(1-\gamma)(1-\beta) > (1-q_B)(1-\gamma)\gamma(1-\beta)$$

which is true by assumption 2. In the same way

$$\gamma+(1-\gamma)\beta > \frac{(1-q_G)\gamma+q_B(1-\gamma)\beta}{(1-q_G)\gamma+q_B(1-\gamma)}$$

$$\iff (1-q_G)\gamma[1-(1-\gamma)]+(1-\gamma)\beta(1-q_G)\gamma$$

$$+q_B(1-\gamma)\gamma+q_B(1-\gamma)\beta-q_B(1-\gamma)\gamma\beta$$

$$> (1-q_G)\gamma+q_B(1-\gamma)\beta$$

Once again deleting terms on both sides of the inequality and rearranging yields

$$q_B(1-\gamma)\gamma(1-\beta) > (1-q_G)(1-\gamma)\gamma(1-\beta)$$

which is true by assumption 2. Note that it does not necessarily follow that an applicant with a good signal is more likely to be good than bad. That will always be the case when $\gamma \geqq \frac{1}{2}$, but need no longer hold when γ is sufficiently low. As a counter-example that is compatible with assumption 2, consider $\gamma = 0.1$, $q_G = 0.11$, and $q_B = 0.91$.

Assumption 1 implies that it is impossible to break even lending to a bad borrower. However, the group of applicants assigned to the bad group by the test contains some good types as well and the average success probability of that group will therefore be greater than β. Broecker assumes that the average success probability of those projects categorized as 'bad' is so low that in expectation a bank cannot break even lending to one of these projects:

ASSUMPTION 3:

$$\frac{(1-q_G)\gamma + q_B(1-\gamma)\beta}{(1-q_G)\gamma + q_B(1-\gamma)} R^{MAX} < 1 \tag{3.5}$$

Lending to the group of applicants categorized as 'good' is, however, feasible:

ASSUMPTION 4:

$$\frac{q_G\gamma + (1-q_B)(1-\gamma)\beta}{q_G\gamma + (1-q_B)(1-\gamma)} R^{MAX} > 1 \tag{3.6}$$

These latter two assumptions are important insofar as they ensure that all banks offer only a single interest rate to those applicants they believe to be good. If lending to those projects categorized as bad were feasible, banks could offer two interest rates to account for the different default probabilities which would greatly complicate the analysis. An important assumption (to be relaxed below) is that banks' tests are equally informative. Although Broecker does not mention this explicitly, it must be impossible for banks to subject their applicants to the test more than once and thereby improve the accuracy of their results. Alternatively, the results of multiple tests run by the same bank must be perfectly correlated. Also, banks do not learn each other's test results, neither can they condition on their rivals' assessments. As for the banks' refinancing, Broecker assumes that banks can borrow unlimited amounts of funds at a zero interest rate. Alternatively, one could assume that banks are lending their capital. In either case, the question as to what about these financial intermediaries turns them into 'banks' is justified, but not pursued by Broecker. Neither do I pursue it here. The contracts that the bank offers can be described as standard debt contracts (SDC), the use of these contracts is assumed. However, given the assumptions about the

cashflow distribution it is not obvious in how far a debt contract differs from an equity contract.

3.2.3 Interest-rate competition

Applicants costlessly apply to all banks and borrow with probability $\frac{1}{k}$ from one of the k banks (if any) that offer them the lowest interest-rate based on their test results. Banks are thus competing in prices (given the test results) and I am therefore looking for an equilibrium in interest-rates[2]. The key result of Broecker's contribution is that such an equilibrium does not exist in pure strategies, there is only a (unique and symmetric) mixed strategy equilibrium in which banks make interest rate offers according to a distribution function over interest rates[3]. Broecker points out that

'banks face externalities caused by the interest rates and the rejection decisions of the other banks' (p. 430).

I illustrate this for the case of two banks: If bank 1 charges a lower interest rate than bank 2, everyone who gets an offer from bank 1 will accept it and the composition of bank 1's borrower pool only depends on the accuracy of its own screening. The group of entrepreneurs applying to bank 2, however, is made up of those applicants who were rejected by 1. As banks' test results are at least somewhat informative, the pool of applicants left for bank 2 then contains relatively many bad projects. There will thus be an incentive for bank 2 to undercut in turn as the improvement in the composition of its borrower pool will more than outweigh the reduction in profits due to the lower interest rate. However, it will not be optimal for one of the banks to set their interest rate lower than some value; rather than cutting the rate further it could increase its profits by raising the interest rate to the highest possible amount R^{MAX} (remember that the demand for loans is assumed to be inelastic)[4]. But

[2]The notational convention used throughout the paper is to refer to repayments (i.e. the gross interest rate consisting of the principal and a net interest margin) as the interest rate.

[3]If the number of potentially active banks exceeds a certain threshold value, banks do not necesssarily make interest rate offers to those applicants that they believe to be good, but leave the market with some probability that is strictly increasing in the number of banks that are potentially active. I will abstract from the possibility of some banks not making offers at all, as it does not add much to the analysis.

[4]Implicit in this description is the assumption that the higher-rate bank can at least break even when charging R^{MAX}. This is the equivalent of assumption 5 discussed below (for the case of asymmetric banks). Assumption 3 says that it is not possible to break even on the group of applicants tested once and rejected, assumption 4 states that the opposite is true for the group of applicants tested once and accepted. The higher-rate bank's borrowers have the same average success probability as a group of applicants that have been tested twice and have been rejected once and accepted once. It is intuitive that the average success probability for this group is greater than that

of course, if, say, bank 2 charges R^{MAX}, bank 1's best response is to charge just an ε less than bank 2. In this way, for any interest rate that bank 1 sets bank 2 has an incentive to either undercut or charge the highest possible interest rate R^{MAX} and conversely. The mechanism sketched out here is behind Broecker's proposition 2.1. which establishes the non-existence of an equilibrium in pure strategies.

3.2.4 Average Success Probability

Since banks' tests are informative, the average success probability (the average probability with which a borrower is able to repay the loan) decreases with the number of banks. To see this, note that the probability that a good (bad) borrower is rejected by n banks is $(1-q_G)^n$ (q_B^n), so that when there are n banks, $[1-(1-q_G)^n]\gamma$ good applicants and $(1-q_B^n)(1-\gamma)$ bad applicants will obtain a loan. The condition for the average success probability to be decreasing in the number of banks is therefore:

$$\frac{\partial}{\partial n}\frac{[1-(1-q_G)^n]\gamma+\beta(1-q_B^n)(1-\gamma)}{[1-(1-q_G)^n]\gamma+(1-q_B^n)(1-\gamma)}<0 \tag{3.7}$$

Noting that $\beta=[1-(1-\beta)]$, this can be restated as

$$\frac{\partial}{\partial n}\frac{[1-(1-q_G)^n]\gamma+(1-q_B^n)(1-\gamma)-(1-\beta)(1-q_B^n)(1-\gamma)}{[1-(1-q_G)^n]\gamma+(1-q_B^n)(1-\gamma)}<0$$

$$\Leftrightarrow \frac{\partial}{\partial n}\left[1-\frac{(1-\beta)(1-q_B^n)(1-\gamma)}{[1-(1-q_G)^n]\gamma+(1-q_B^n)(1-\gamma)}\right]<0$$

Now dividing both numerator and denominator by $(1-q_B^n)(1-\gamma)$ one obtains

$$\Leftrightarrow \frac{\partial}{\partial n}\left[1-\frac{(1-\beta)}{\frac{[1-(1-q_G)^n]\gamma}{(1-q_B^n)(1-\gamma)}+1}\right]<0 \tag{3.8}$$

Differentiating $\frac{[1-(1-q_G)^n]\gamma}{(1-q_B^n)(1-\gamma)}$ with respect to n yields:

$$\frac{\partial}{\partial n}\frac{[1-(1-q_G)^n]\gamma}{(1-q_B^n)(1-\gamma)}=$$

for applicants tested once and rejected, but less than that for applicants tested once and accepted. So assumptions 3 and 4 neither rule out nor imply that it is possible to charge this group R^{MAX} and break even.

$$\frac{\gamma}{(1-q_B^n)^2(1-\gamma)}[(1-q_G)^n(1-q_B^n)\ln(1-q_G)+(q_B)^n(1-(1-q_G)^n)\ln(q_B)] \tag{3.9}$$

As the logarithms of $(1-q_G)$ and (q_B) are necessarily negative, the whole expression is negative and the inequality holds. There is a natural interpretation for this result: As the number of banks n increases, the ratio of accepted good types (of whom there are $[1-(1-q_G)^n]\gamma$) to accepted bad types (of whom there are $(1-q_B^n)(1-\gamma)$) decreases. This shows that the average success probability is decreasing in the number of active banks.

3.3 Asymmetric Banks

Broecker conjectured that the non-existence result was not robust to an asymmetry in banks' information, i.e. banks that have tests of different accuracy. He claims that

'it is quite clear that a bank with a more reliable or less costly test than those of the other banks will be the only bank that stays in the market [and that] this bank will make positive profit.' (p. 445)

In the following I will show for the case of two banks that the mentioned result is in fact robust to such a modification.

3.3.1 Superior screening accuracy

Modified assumption 2 captures the superiority of bank 1's screening technology. Superscripts refer to the bank, so bank 1 has a lower probability of making errors than bank 2:

ASSUMPTION 2':

$$\Pr ob^1(g|G) = q_G^1 > q_G^2 > \gamma > 1-q_B^2 > 1-q_B^1 = \Pr ob^1(g|B) \tag{3.10}$$

Once again, this can equivalently be stated as

$$\Pr ob^1(b|B) = q_B^1 > q_B^2 > 1-\gamma > 1-q_G^2 > 1-q_G^1 = \Pr ob^1(b|G) \tag{3.11}$$

To make this question interesting, I will assume that it is indeed possible for the less accurate bank charging the higher interest rate[5] to make a profit - otherwise one would have a monopoly by assumption:

[5]Given the inelasticity of demand, the bank charging the higher interest rate wil always charge the highest possible interest rate R^{MAX}.

ASSUMPTION 5:

$$\frac{q_G^2(1-q_G^1)\gamma + (1-q_B^2)q_B^1(1-\gamma)\beta}{q_G^2(1-q_G^1)\gamma + (1-q_B^2)q_B^1(1-\gamma)} R^{MAX} > 1 \qquad (3.12)$$

3.3.2 Non-existence of a pure-strategy equilibrium with asymmetric banks

In this section I will investigate Broecker's conjecture that a bank that is unambiguously more accurate than its rivals will be the only bank to remain in the market. It turns out that not only is there no reason to think that the less accurate bank will leave the market, but that the modified game will have an equilibrium very similar to the one in the symmetric game. I summarize this in the following

Proposition 1 *The asymmetric game where one bank is unambiguously more accurate than the other bank does not have an equilibrium in pure strategies. However, under assumption 5 both banks will stay in the market.*

Proof: To see that both banks will stay in the market, note that regardless of the interest rate set by the rival, both banks can guarantee themselves a strictly positive profit by setting R^{MAX}. Assumption 5 ensures that bank 2 (the less accurate bank) will make a profit at this rate; it is clear that bank 1 must then make a profit at that rate as well.

To show that the game does not have a pure strategy equilibrium, I have to show that there is no set of strategies such that each bank is using its best response, given the rival's strategy. First, note that because the demand for loans is inelastic, a bank that decides to set a rate higher than its rival will always set R^{MAX}. Similarly, a bank that decides to undercut its rival will undercut by an ε only, as this allows it to capture all the applicants with offers from both banks. Bank j then has three possible responses to bank k charging R^k 1. It can charge R^{MAX}, 2. it can charge R^k, or 3. it can charge $R^k - \varepsilon$. Note also that the only interest rates I have to consider are from the interval $[1, R^{MAX}]$, as no bank would ever set a repayment below one or greater than R^{MAX}. I will first show that it is never a best response to set the same rate as the rival, as setting the same rate as the rival is always dominated by either undercutting the rival by ε or by charging R^{MAX}.

Note that the bank undercutting the rival will capture all the applicants that have offers from both banks, whereas it would capture only half of them if it set the same rate. Because these applicants have offers from both banks, their average creditworthiness must be higher than that

of the remaining applicants and a bank that manages to capture all of them must be better off than a bank capturing only half of them as long as both banks are making non-negative profits when charging the same rate. If one bank were not at least breaking even when charging the same rate as the rival it would anyway be in that bank's interest to raise its rate to R^{MAX} which, as noted above, will guarantee it positive profits. In sum, if banks can never find it optimal to charge the same rate, they must be charging different rates.

To complete the proof I wish to show the following: No matter which interest rate the rival sets, it is always in a bank's interest to either undercut by ε or to raise its interest rate to R^{MAX}. The other bank will always have an incentive to either undercut in turn or to raise its own rate to R^{MAX}, such that the rivals never choose mutually best responses. The following lemma establishes this result:

Lemma 2 *For bank j there exists a critical interest rate R_C^k such that*

$$\begin{aligned} R^k \in (R_C^k, R^{MAX}]: &\quad \lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) > \pi^j(R^j = R^{MAX}) \\ R^k = R_C^k: &\quad \lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) = \pi^j(R^j = R^{MAX}) \\ R^k \in [1, R_C^k) &\quad \lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) < \pi^j(R^j = R^{MAX}) \end{aligned}$$

Note that $\lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) - \pi^j(R^j = R^{MAX})$ is a continuous function of R^k and is given by $\lim_{\varepsilon \to 0+} [q_G^j \gamma(R^k - \varepsilon - 1) + \beta(1 - q_B^j)(1 - \gamma)(R^k - \varepsilon - 1)] - (1 - q_G^k) q_G^j \gamma(R^{MAX} - 1) - q_B^k (1 - q_B^j)(1 - \gamma)\beta(R^{MAX} - 1)$. Differentiating this difference in profit levels with respect to R^k shows that the derivative is strictly positive.

$$\frac{\partial}{\partial R^k} \left(\lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) - \pi^j(R^j = R^{MAX}) \right)$$

$$= [q_G^j \gamma + \beta(1 - q_B^j)(1 - \gamma)] > 0$$

, so $\lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) - \pi^j(R^j = R^{MAX})$ is monotonically increasing in R^k. I have already shown that when bank k sets R^{MAX} bank j has an incentive to undercut, i.e. $\lim_{\varepsilon \to 0+} \pi^j(R^j = R^{MAX} - \varepsilon) - \pi^j(R^j = R^{MAX})$ is positive. All I need to show is that for some R^k bank j makes a higher profit by setting R^{MAX} rather than by undercutting bank k, i.e. $\lim_{\varepsilon \to 0+} \pi^j(R^j = R^k - \varepsilon) - \pi^j(R^j = R^{MAX}) < 0$. Then by the intermediate value theorem the existence of the critical interest rate is established.

To find such an R^k, define $\underline{R}^j(R^j < R^k)$ as the lowest interest rate that bank j can set such that it just breaks even when it charges a lower rate

than its rival, i.e. $R^j < R^k$. Let bank k charge $\underline{R}^j(R^j < R^k) - \varepsilon$. If bank j now undercuts by ε it is making a loss and would, by assumption 5, do better by charging R^{MAX} which would guarantee it a strictly positive profit. What determines the critical interest rate is the profit that each bank can guarantee itself by charging the maximum interest rate. At the critical rate, the profit that the bank can obtain from undercutting is exactly what it could obtain from raising its rate to R^{MAX}, at any lower interest rate the bank must be strictly better off by charging R^{MAX}.

If the banks' critical interest rates are different (as they must be if banks achieve different accuracies), the non-existence of an equilibrium in pure strategies is immediate. I conjecture that there is once again an equilibrium in mixed strategies, albeit an asymmetric one.

3.4 Discussion

In the interest-rate competition game analysed here, there exists for both banks a critical interest rate defined as that interest rate set by the rival such that, on the one hand, undercutting the rival by ε and, on the other hand, charging the highest possible interest rate R^{MAX} yield the same profit. If the rival charges more than the critical rate, the best response is to undercut by ε; if the rival charges less, the best response is to charge the maximum interest rate R^{MAX}. Both banks have an incentive to undercut each other up to the greater of the two critical rates. The bank with the greater critical rate will then jump to R^{MAX}, prompting the rival to charge $R^{MAX} - \varepsilon$, thus starting the cycle of undercutting anew. This cycle of undercutting followed by a jump to the highest possible rate which in turn triggers a new round of mutual undercutting is the intuition behind the non-existence of an equilibrium in pure strategies. It is shown that the mechanism at work in the asymmetric game is the same as in the symmetric game: the difference in accuracy has not played a role at all in the proof (except in that it affects the critical rates). The non-existence result obtained by Broecker, contrary to Broecker's conjecture, is thus shown to be robust to an asymmetry in the testing technology. Even when banks achieve different levels of accuracy, there is no set of interest rates such that each bank is using its best response given the rival's interest rate and neither bank has an incentive to change its interest rate.

As an aside and by way of motivation for the following chapter, note that the general setup of Broecker's model is such that the validity of policy recommendations based on it can legitimately be questioned. In

the (symmetric) version of the game investigated by Broecker, the mixed-strategy equilibrium is very hard to interpret: Is there a real-world equivalent of banks randomizing over an interval of interest rates? Other features of the model also make me suspicious of its applicability to real-world issues. For example, error probabilities are exogenous and constant regardless of the number of applicants. Yet in spite of such difficulties, claims that more intense competition between intermediaries is harmful are often made with reference to this model. As shown above, one conclusion that can be drawn from Broecker's paper is that the more banks are in operation, the lower the average credit-worthiness of borrowers will be. It is presumably this result that leads to Broecker's paper being cited as showing that more intense competition between banks is bad for the allocation of capital. The operation of more banks clearly entails an increase in the number of bad projects that obtain funding, but that is the best capital allocation that can be achieved with the screening technology in this model. In sum, the set-up of the model is such that its applicability to the analysis of real-world banking issues must be in doubt. The following chapter shows that moving away from Broecker's auction framework and allowing applicants to renegotiate their offers leads to the existence of an equilibrium in pure strategies.

Chapter 4

Interest rate competition with renegotiation

4.1 Introduction

The following modification of Broecker's game is motivated by the observation that in some market segments interest rate competition between banks in practice operates in ways quite different from Broecker's auction setting. In particular, during the fieldwork reported in a companion chapter I learned that it is common for applicants to return to banks that have made them an offer and to try to convince the bank to improve the terms of the offered loan contract by showing the offers they have obtained from other banks. In the sample of banks I visited, banks do not insist on borrowers keeping the terms of the loan contract offered to them confidential and make their offers in writing, so that they are indeed verifiable. To the extent that another bank's loan offer conveys favorable information, it is then in an applicant's interest to seek to improve existing offers by making use of this information. This effectively leads to test results becoming public. There is every reason to expect this mechanism to develop endogenously: those applicants with many favorable test results have an incentive to show them; then everyone who cannot show evidence of many favorable test results must be of lesser quality (in expectation). In principle a bank would consider the average creditworthiness of those applicants whose testing histories are unknown, but because the above average applicants always have an incentive to document their high quality, the average falls and falls and eventually there is perfect separation by risk class. In order not to be mistaken for the worst category of applicant, all other applicants (i.e., all those with more than one offer) have an incentive to show all the offers they got.

It seems quite intuitive that the mechanism outlined here leads to a

pure-strategy equilibrium where the interest rate offered by each bank to each applicant when all test results have been revealed is such that the bank just breaks even. Borrowers with only a single offer will be charged the maximum interest rate - this is not surprising as, in the absence of other lenders, the bank making the offer is a monopolist.

4.2 The modified setup

I specifically assume that in the standard interest rate competition game as outlined in the previous chapter, there are now two rounds instead of just one. After the first round, applicants can reapply to all those banks who made them an offer in the first round and try to negotiate a lower interest rate. To do this, they show evidence of all the offers they have obtained. Since it is clearly in an applicant's interest to show all the offers obtained and since applications to banks are assumed to be costless, an applicant with m offers must have got $n-m$ rejections when the number of banks is n. A bank that has rejected an applicant in round 1 cannot make an offer in round 2. For ease of exposition, let β (the probability that a bad project is successful) equal zero. To simplify, I assume that all of these applicants still want to borrow. It is consistent with the model setup that applicants do not know their type.

Definition 3 *Define* $R\left(\frac{m}{n}\right) = 1+\left(\frac{1-q_B}{q_G}\right)^m \left(\frac{q_B}{1-q_G}\right)^{n-m} \left(\frac{1-\gamma}{\gamma}\right)$. *As I show below, this is the repayment at which a bank can break even lending to an applicant that has obtained m offers from n banks.*

ASSUMPTION 4':

$$R\left(\frac{1}{n}\right) \leq R^{MAX} \tag{4.1}$$

Then the following result obtains:

Proposition 4 *The interest rate competition game with renegotiation has the following equilibrium in pure strategies: In round 1 all banks ask for* R^{MAX} *from an accepted applicant. In round 2 if* $m = 1$ *(the applicant does not have any other offers) the initial offer of* R^{MAX} *is not modified. If* $m \geq 2$ *(i.e., the applicant obtains at least one additional offer), set* $R\left(\frac{m}{n}\right)$.

<u>Proof:</u> To see that $R\left(\frac{m}{n}\right)$ is indeed the repayment at which a bank breaks even lending to an applicant with m offers from n banks, note

that the number of good signals in n applications is a binomial random variable and the probability of obtaining m offers in n applications will depend on type as follows. Let $\left(\frac{m}{n}\right)$ denote that m offers have been obtained in n applications. Then

$$Prob(\left(\frac{m}{n}\right)|G) = \binom{n}{m} q_G^m (1-q_G)^{n-m} \tag{4.2}$$

$$Prob(\left(\frac{m}{n}\right)|B) = \binom{n}{m} (1-q_B)^m q_B^{n-m} \tag{4.3}$$

Also note that $\Pr ob(G) = \gamma$ and $\Pr ob(B) = 1-\gamma$. Then by Bayes' Rule:

$$\Pr ob(G|\left(\frac{m}{n}\right)) = \frac{Prob(\left(\frac{m}{n}\right)|G)\Pr ob(G)}{Prob(\left(\frac{m}{n}\right)|G)\Pr ob(G) + Prob(\left(\frac{m}{n}\right)|B)\Pr ob(B)}$$

$$= \frac{\binom{n}{m} q_G^m (1-q_G)^{n-m}\gamma}{\binom{n}{m} q_G^m (1-q_G)^{n-m}\gamma + \binom{n}{m}(1-q_B)^m q_B^{n-m}(1-\gamma)}$$

(and similarly for $\Pr ob(B|\left(\frac{m}{n}\right))$). Then

$$\Pr ob(G|\left(\frac{m}{n}\right)) = \frac{1}{1+(\frac{1-q_B}{q_G})^m(\frac{q_B}{1-q_G})^{n-m}\frac{1-\gamma}{\gamma}} \tag{4.4}$$

and noting the definition of $R\left(\frac{m}{n}\right)$, i.e. $\Pr ob(G|\left(\frac{m}{n}\right))R\left(\frac{m}{n}\right) = 1$, establishes the result.

As by assumption 4' $R\left(\frac{1}{n}\right) \leq R^{MAX}$, a bank will initially offer a loan at the repayment R^{MAX} to any applicant for whom the test result is favorable. Charging a repayment higher than R^{MAX} does not make sense and charging less than R^{MAX} cannot be optimal, as a bank can always lower but not raise its offer in round two and has no advantage from setting a lower rate in round one. If the applicant does not manage to obtain any additional offers, the bank is a monopolist and does not need to revise its interest rate offer downwards. The condition $R\left(\frac{1}{n}\right) \leq R^{MAX}$ ensures that this is (weakly) profitable for the bank, as $R\left(\frac{m}{n}\right)$ is the repayment at which a bank can break even lending to an applicant that has obtained m offers from n banks. If the applicant has at least one

additional offer, the proof showing that charging $R\left(\frac{m}{n}\right)$ is an equilibrium price simply follows Bertrand's classic argument for the case of price competition with constant unit costs. To see that $R\left(\frac{m}{n}\right)$ is indeed a Nash equilibrium, note that all banks make zero profits at this interest rate. A bank lowering its price below $R\left(\frac{m}{n}\right)$ will make losses whereas a bank charging more than $R\left(\frac{m}{n}\right)$ will not have any borrowers and still make zero profits. For $n = 2$ banks, this is a unique pure strategy equilibrium. As for additional equilibria when $n > 2$, note that charging less than $R\left(\frac{m}{n}\right)$ would result in losses. The worst that can happen to a bank charging $R\left(\frac{m}{n}\right)$ or more is to break even, so it is clearly not optimal to charge less than $R\left(\frac{m}{n}\right)$. If one bank were to be alone in charging $R\left(\frac{m}{n}\right)$ and its rivals all demanded strictly more than $R\left(\frac{m}{n}\right)$, then clearly the bank charging the lower price is not behaving optimally as it could make strictly positive (instead of zero) profits by raising its price a little. However, any combination of interest rates involving at least two banks charging $R\left(\frac{m}{n}\right)$ would be an equilibrium.

Note that assumption 4' is a generalized version of Assumption 4. To see this note that assumption 4 with $\beta = 0$ becomes

$$\frac{q_G\gamma}{q_G\gamma + (1-q_B)(1-\gamma)}R^{MAX} > 1 \tag{4.5}$$

Thus assumption 4 is a special case of $R\left(\frac{1}{n}\right) \leq R^{MAX}$ obtained by setting $n = 1$. In terms of the economics, assumption 4' (one good signal out of n tests) is more restrictive than assumption 4 (one good signal out of one attempt) in the sense that if assumption 4' holds, then so must assumption 4.

Remark 5 *It follows from the definition that* $R\left(\frac{m}{n}\right) > R\left(\frac{m+1}{n}\right)$, *i.e.* $R\left(\frac{1}{n}\right) > R\left(\frac{2}{n}\right) > (...) > R\left(\frac{n-1}{n}\right) > R\left(\frac{n}{n}\right)$ *(the higher the number of offers given the number of banks, the higher the average quality of the applicant and the lower the break-even rate). In fact, it can be shown that* $R\left(\frac{m}{n}\right) - 1 = \frac{q_G}{1-q_B}\frac{q_B}{1-q_G}[R\left(\frac{m+1}{n}\right) - 1]$.

Note that

$$R\left(\frac{m+1}{n}\right) = 1 + \left(\frac{1-q_B}{q_G}\right)^{m+1}\left(\frac{q_B}{1-q_G}\right)^{n-(m+1)}\left(\frac{1-\gamma}{\gamma}\right)$$

$$= 1 + \left(\frac{1-q_B}{q_G}\right)^m \left(\frac{1-q_B}{q_G}\right) \left(\frac{q_B}{1-q_G}\right)^{n-m} \left(\frac{1-q_G}{q_B}\right) \left(\frac{1-\gamma}{\gamma}\right)$$

$$= 1 + \left(\frac{1-q_B}{q_G}\right) \left(\frac{1-q_G}{q_B}\right) [R\left(\frac{m}{n}\right) - 1]$$

Remark 6 *An applicant obtaining m offers from $n+1$ banks must on average be less creditworthy than an applicant obtaining the same number of offers from n banks: $R\left(\frac{m}{n+1}\right) > R\left(\frac{m}{n}\right)$. In fact, $R\left(\frac{m}{n+1}\right) - 1 = \frac{q_B}{1-q_G}[R\left(\frac{m}{n}\right) - 1]$*

This can be seen as follows:

$$R\left(\frac{m}{n+1}\right) = 1 + \left(\frac{1-q_B}{q_G}\right)^m \left(\frac{q_B}{1-q_G}\right)^{n+1-m} \left(\frac{1-\gamma}{\gamma}\right)$$

$$= 1 + \left(\frac{1-q_B}{q_G}\right)^m \left(\frac{q_B}{1-q_G}\right)^{n-m} \left(\frac{q_B}{1-q_G}\right) \left(\frac{1-\gamma}{\gamma}\right)$$

$$= 1 + \left(\frac{q_B}{1-q_G}\right) [R\left(\frac{m}{n}\right) - 1]$$

Remark 7 *It is good news when both the number of offers and the number of tests increase by one: $R\left(\frac{m}{n}\right) > R\left(\frac{m+1}{n+1}\right)$, since $R\left(\frac{m+1}{n+1}\right) - 1 = \frac{1-q_B}{q_G}[R\left(\frac{m}{n}\right) - 1]$.*

Note that

$$R\left(\frac{m+1}{n+1}\right) = 1 + \left(\frac{1-q_B}{q_G}\right) \left(\frac{1-q_B}{q_G}\right)^m \left(\frac{q_B}{1-q_G}\right)^{n+1-(m+1)} \left(\frac{1-\gamma}{\gamma}\right)$$

$$= 1 + \left(\frac{1-q_B}{q_G}\right) [R\left(\frac{m}{n}\right) - 1]$$

The reason that assumption 4' is required is that with $R\left(\frac{1}{n}\right) > R^{MAX}$, the bank must be making losses on applicants to whom no other bank has made an offer. The setup of the model guarantees that every bank in the market will end up with some of these borrowers. Since - following the above reasoning - the bank would still make zero profits on all borrowers with 2 or more offers, the bank's expected profit when $R\left(\frac{1}{n}\right) > R^{MAX}$ must be negative. Then the bank is surely better off not making any offers at all. However, it is not an equilibrium for all

banks to leave the market. Assuming that all other $n-1$ banks leave the market and decide not to lend, Assumption 4 guarantees that the $n-th$ bank has an incentive to stay in the market. The conclusion is that when $R\left(\frac{1}{n}\right) > R^{MAX}$, any equilibrium must involve the use of mixed strategies which, however, I do not believe to be of much interest nor particularly plausible in the present setting. I therefore rule out that possibility via assumption 4'.

4.3 Entry

The above comments relate to a situation with a given number of banks. Assuming that banks enter sequentially, how many banks will enter the market? When there are already x banks in the market, entry will occur if a single bank charging R^{MAX} that is alone among $x+1$ banks in making an offer to an applicant can break even. This will be the case if and only if

$$R\left(\frac{1}{x+1}\right) \leq R^{MAX} \tag{4.6}$$

This can also be written as

$$(\frac{q_B}{1-q_G})^x \leq (R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B}$$

Note that the l.h.s. is increasing in x and so the higher x the more unlikely it is that an additional bank has an incentive to enter the market. (The assumption of a fixed market size is maintained.)

4.4 Surplus

As ultimately my concern is with the allocation of capital, a question worth asking is what the relationship is between the intensity of competition in the modified game and the net surplus generated by the entrepreneurs' borrowing (gross surplus generated by good projects minus loan losses on bad projects). It turns out that the effect of more intense competition is unambiguously beneficial.

Proposition 8 *Free entry maximizes the net surplus.*

Proof: Let $S(x)$ denote the net surplus as a function of the number of banks. $S(x)$ =[Number of good borrowers with x banks $(R^{MAX}-1)$+ Number of bad borrowers with x banks (-1)] = $(1-(1-q_G)^x)\gamma(R^{MAX}-1)$

$-(1-q_B^x)(1-\gamma)$ Under which condition would $S(x+1)$ be greater than $S(x)$?

$$S(x+1) = (1-(1-q_G)^{x+1})\gamma(R^{MAX}-1) - (1-q_B^{x+1})(1-\gamma)$$

$$= [1-(1-q_G)^x(1-q_G)]\gamma(R^{MAX}-1) - [1-q_B^x(1-(1-q_B))](1-\gamma)$$

$$= [1-(1-q_G)^x + (1-q_G)^x q_G]\gamma(R^{MAX}-1) - [1-q_B^x + q_B^x(1-q_B)](1-\gamma)$$

This can be split up as follows

$$= \underbrace{[1-(1-q_G)^x]\gamma(R^{MAX}-1) - (1-q_B^x)(1-\gamma)}_{S(x)}$$

$$+(1-q_G)^x q_G \gamma(R^{MAX}-1) - q_B^x(1-q_B)(1-\gamma)$$

So $S(x+1) - S(x) = (1-q_G)^x q_G \gamma(R^{MAX}-1) - q_B^x(1-q_B)(1-\gamma)$. This means that

$$S(x+1) - S(x) \geq 0$$

$$\iff (1-q_G)^x q_G \gamma(R^{MAX}-1) - q_B^x(1-q_B)(1-\gamma) \geq 0$$

$$\Leftrightarrow (1-q_G)^x q_G \gamma(R^{MAX}-1) \geq q_B^x(1-q_B)(1-\gamma)$$

$$\iff (R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B} \geq (\frac{q_B}{1-q_G})^x$$

This shows that $S(x+1)-S(x) \geq 0$ if and only if $(\frac{q_B}{1-q_G})^x \leq (R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B}$. But that is just the condition for entry to occur.

Banks' incentives and those of society are here perfectly aligned. The net surplus accrues in the form of either bank profits or profits accruing to good borrowers. In the appendix I show that aggregate bank profits are falling in the number of banks, whereas the aggregate net surplus accruing to good borrowers is increasing in the number of banks (always assuming that the number of banks has not yet reached its maximum). The preceding proposition shows that the net effect of an increase in the

number of banks on net surplus is positive. This is an encouraging result, given negative results obtained in earlier literature and discussed in the accompanying literature review. Riordan (1993), for example, using an explicit auction framework somewhat similar to Broecker's implicit auction framework in referring to his results notes that 'the model suggests two reasons why more competition in loan markets might damage market performance. First, the statistic the market uses to select loans may become less informative about the quality of the loan. Second, the loan approval practices of individual banks may be too conservative, and more competition may make them even more so.' (p. 330) These winner's curse - related phenomena cannot arise in the present setting where all test results become public.

One may now ask how the surplus is distributed and how an increase in the number of banks affects the expected profit of borrowers and banks. Loans to bad borrowers are a pure social loss in this framework, but the likelihood of a bad borrower obtaining a loan is increasing in the number of banks. The average bank must be worse off, for not only do aggregate bank profits fall, but these have to be shared among a larger number of banks. I formally show that this is the case in the appendix. For a good borrower, the expected profit is the product of the probability of obtaining a loan (which must increase in the number of banks) and the profit per borrower (about which little can be said without more detailed analysis as with more banks both good borrowers' aggregate profits and the number of good borrowers among whom these profits are distributed increase). I show in the appendix that a good borrower's expected profit is indeed increasing in the number of banks.

4.5 Expected interest rate for a good borrower

The last point in the preceding section suggests that a good borrower must ex ante, i.e. before knowing whether she has obtained at least one offer, always prefer more to fewer banks. But would she prefer more to fewer banks even conditional on knowing that she has obtained at least one offer? Phrased differently, is the average profit accruing to each accepted good borrower increasing in the number of banks? It turns out that the answer to that question hinges on whether the average or expected interest rate paid by a good borrower is decreasing in the number of banks. To see this, note that the expected profit per good borrower who invests is simply aggregate profits divided by the number

of good borrowers:

$$\frac{(1-nq_G(1-q_G)^{n-1}-(1-q_G)^n)\gamma(R^{MAX}-1)-(1-n(1-q_B)q_B^{n-1}-q_B^n)(1-\gamma)}{(1-(1-q_G)^n)\gamma} \tag{4.7}$$

The numerator in the above expression is composed as follows: All those good borrowers with only a single offer (of whom there are $nq_G(1-q_G)^{n-1}\gamma$) are made to pay R^{MAX} and so do not earn any profit. $(1-q_G)^n\gamma$ good borrowers do not manage to obtain a loan. All other good borrowers obtain a net cashflow of $(R^{MAX}-1)$. However, because banks make exactly zero profit on all borrowers with two or more offers, these good borrowers also have to pay the loan losses on all the accepted bad borrowers of whom there are $(1-q_B^n)(1-\gamma)$. As the loan losses on the $n(1-q_B)q_B^{n-1}(1-\gamma)$ bad borrowers with only a single offer are paid for by those good borrowers with only a single offer, the remaining loan losses to be covered are $(1-n(1-q_B)q_B^{n-1}-q_B^n)(1-\gamma)$.

By taking $(1-(1-q_G)^n)\gamma(R^{MAX}-1)$ out of the fraction, a good borrower's average profit can be rewritten as

$$(R^{MAX}-1)-\frac{nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)+(1-n(1-q_B)q_B^{n-1}-q_B^n)(1-\gamma)}{(1-(1-q_G)^n)\gamma} \tag{4.8}$$

$(R^{MAX}-1)$, the net surplus created by a good borrower's project, is independent of the number of banks. However, the second term denotes the average interest expense with n banks which depends on the number of banks[1]. Therefore, the average profit per accepted good borrower will increase in the number of banks if the average interest expense per good borrower is decreasing in the number of banks. The above expression for the average interest expense could in principle be differentiated directly with respect to n, but the resulting derivative cannot be signed in a straightforward way.

That the change in a good borrower's expected interest rate when

[1] An alternative way of obtaining an expression for the expected interest rate is to note that the probability that none of n banks makes an offer to a good applicant is $(1-q_G)^n$, so the probability that at least one bank makes an offer is $(1-(1-q_G)^n)$. Conditional on receiving at least one offer, a good borrower's expected interest rate is then $E(R|m\geq 1,n,G)=$

$$\frac{1}{(1-(1-q_G)^n)}\{nq_G(1-q_G)^{n-1}R^{MAX}+\sum_{m=2}^{n}\binom{n}{m}q_G^m(1-q_G)^{n-m}R\left(\frac{m}{n}\right)\} \tag{4.9}$$

However, $\{E(R|m\geq 1,n,G)-E(R|m\geq 1,n+1,G)\}$ cannot be signed in a straightforward way.

the number of banks increases should be ambiguous becomes plausible when one decomposes the expected interest rate with n banks into its two components. The expected interest rate for a good borrower when there are n banks can be rewritten as:

$$\frac{nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-n(1-q_B)q_B^{n-1}(1-\gamma)}{(1-(1-q_G)^n)\gamma}+\frac{(1-q_B^n)(1-\gamma)}{(1-(1-q_G)^n)\gamma} \tag{4.10}$$

The expression is obtained straight from the preceding fraction by simply grouping banks' revenues on those good borrowers with only a single offer, $nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)$, with their loan losses on the accepted bad borrowers with only a single offer, $n(1-q_B)q_B^{n-1}(1-\gamma)$. The first term, therefore, consists of bank profits (divided by the number of good borrowers) which - as I show in the appendix - are decreasing in the number of banks. The second term denotes the ratio of bad borrowers to good borrowers which is increasing in the number of banks.

The change in a good borrower's expected interest rate is therefore the sum of two opposing effects. The intuition underlying this result is that with an additional bank, on the one hand banks' monopoly power is reduced. In the situation in which an applicant obtains only a single offer, the bank is a monopolist and the borrower is completely expropriated (whereas with two or more offers, the borrower only pays the actuarially fair interest rate). Having to hand over the entire cashflow R^{MAX}, all else being equal, drives up the expected interest payment. As this is less likely to happen when there is an additional bank (the probability of obtaining two offers or more goes up), the effect is to reduce the expected interest payment. On the other hand, an additional lender gives bad projects an additional opportunity of obtaining funding. This increases loan losses on bad projects which have to be covered by the good borrowers through their interest payments. Therefore, the second effect of an additional lender is to increase the good borrowers' expected interest payments.

4.6 Break-even rates non-monotonic in accuracy

An observation that is somewhat puzzling at first glance is that the banks' break-even rate $R\left(\frac{m}{n}\right)$ is not monotonically decreasing in q_G and q_B (the parameters which measure the bank's accuracy in correctly identifying an applicant's type). Differentiating with respect to these parameters

gives:

$$\frac{\partial}{\partial q_G} R\left(\frac{m}{n}\right) = (-1)[R\left(\frac{m}{n}\right) - 1]\frac{n}{q_G(1-q_G)}[\frac{m}{n} - q_G] \qquad (4.11)$$

$$\frac{\partial}{\partial q_B} R\left(\frac{m}{n}\right) = (-1)[R\left(\frac{m}{n}\right) - 1]\frac{n}{q_B(1-q_B)}[q_B - \frac{n-m}{n}] \qquad (4.12)$$

Thus a bank's break-even rate will decrease in q_G iff. $q_G < \frac{m}{n}$ and it will decrease in q_B iff. $q_B > \frac{n-m}{m}$. Is there a convincing economic interpretation for this? Why might a break-even rate ever increase when the signal accuracy increases?

The answer lies in noting that the break-even rate is determined for a fixed ratio of offers (m) to tests (n). Say that $q_G = \frac{k}{n}$ initially and that it increases by a small amount: The break-even rates for all those applicants with more than k offers then decrease and the break-even rates for all those applicants with less than k offers increase. That makes perfect sense, for a low number of offers with a more accurate test is worse news than with a less accurate test and the break-even rates for these applicants should be revised upwards accordingly. Similarly, a high number of offers is better news with a more accurate test than with a less accurate test and so break-even rates for such borrowers should fall. As q_G increases from near zero at first only the break-even rates of the worst applicants will increase whereas the break-even rates for the lower-risk applicants will fall. However, as q_G increases towards one, the break-even rates on intermediate borrowers that had initially fallen in q_G will start to increase again until eventually, for very high levels of q_G almost all applicants except for the very lowest risk groups will see their break-even rates increase and not fall. To look at the extremes, the break-even rate for applicants with a single offer will only decrease in q_G if $q_G < \frac{1}{n}$, whereas the break-even rate of an applicant who obtains offers from all the banks that she applies to decreases in q_G as long as $q_G < 1 = \frac{n}{n}$.

An analogous reasoning applies to q_B (albeit in reverse): As q_B increases from zero, the break-even rates for all but the best applicants increase in q_B. However, as q_B increases further, the break-even rates for even intermediate risks start falling in q_B and as q_B nears one, the break-even rates for all but the worst risks will be decreasing in q_B.

4.7 Conclusion

In the present chapter I showed that when moving away from the auction framework used by Broecker by allowing applicants to renegotiate their offers (based on the total number of offers they have received), the modified game has an equilibrium in pure strategies where for a given number of banks an applicant's interest rate decreases with the number of offers obtained. Surplus unambiguously increases with entry as long as additional banks have an incentive to enter. Therefore, free entry maximizes surplus. The intuition behind this result is that in a setup where testing is costless and results become public knowledge, each additional bank will test all applicants and aggregate information improves.

Note that this finding is robust to differences in testing accuracy. To see this, note that in the first round of interest rate offers, banks still ask for the maximum interest rate (the entire cashflow of a successful project) regardless of their accuracy. No matter how accurate a bank, in the first round the bank has nothing to lose by setting such a high rate. It can always revise the offer down in the second round. However, if it turns out to be the only lender that has made an offer, it can get away with the high rate so there is no incentive to make a lower first-round offer. In the second round, all creditworthiness assessments still become common knowledge, because the applicants' incentives to reveal all their offers is unchanged. Creditworthiness assessments of different accuracies still become public knowledge and are taken into account by **all** banks in the revised estimate of the applicant's default probability. Under the maintained assumption that a bank's accuracy is known, having a more accurate testing technology therefore does not offer a bank an advantage. All borrowers with two or more offers will continue to be charged the actuarially fair interest rate; applicants with a single offer have to accept a repayment of R^{MAX}.

In the model presented here, competition unambiguously has a beneficial effect. However, one does well to remember that this result is obtained for a specific set of assumptions, most importantly the assumption that testing is costless. When testing is costless, free-riding on rivals' creditworthiness assessments is not a concern with respect to surplus. However, in much of the work that is sceptical with respect to competition, the various surplus-increasing activities of banks (ex ante testing etc.) are costly. If the information obtained by, for example, expending managerial effort on accurate creditworthiness assessments becomes common knowledge - as is the case in the present chapter - then this would destroy the incentive to produce this information in the first place. As it

is rather plausible that creditworthiness assessments and other types of monitoring are costly, this observation clearly points at the limits of the model presented here. The results would not carry through into a more general setting in which screening is costly.

However, the model does show that results that rely on modeling bank competition as a Broecker-style (first-price, sealed-bid, common value) auction (and the resulting winner's curse effect) must be interpreted with caution, too. As illustrated here, these may not be robust to borrowers being able to renegotiate loan offers from individual banks on the basis of all the loan offers obtained. Clearly, the question is whether an auction framework is appropriate for modeling bank competition. Based on the empirical evidence discussed in the literature review and personal observation in the course of my fieldwork, I argue that it is not, at least not as far as lending to small- and medium-sized businesses is concerned. The implication for banking theory is that the modeling of bank competition and creditworthiness tests should be robust to the possibility of test results becoming (partially) public and thus available to competitors.

4.8 Appendix

4.8.1 Aggregate profit accruing to good borrowers

I will first show that the aggregate profit (or net surplus) accruing to good borrowers is increasing in the number of banks if the number of banks has not yet reached its maximum. To see this, note that with n banks, $(1-(1-q_G)^n)\gamma$ good applicants each realize a payoff of $(R^{MAX}-1)$ as $(1-q_G)^n\gamma$ good applicants are not accepted by at least one bank. These good borrowers must be paying for the loan losses on those bad borrowers with 2 or more offers [of whom there are $\Pr ob(m \geq 2|n,B)(1-\gamma) = (1-n(1-q_B)q_B^{n-1}-q_B^n)(1-\gamma)$], a category of borrower on which the banks merely break even. All those good borrowers with only one offer (of whom there are $nq_G(1-q_G)^{n-1}\gamma$), however, have to hand over their entire cashflow of R^{MAX} which goes to cover loan losses on bad borrowers with only a single offer (of whom there are $n(1-q_B)q_B^{n-1}(1-\gamma)$) and also earns some strictly positive profits for the banks as long as the number of banks has not yet reached its maximum. The aggregate profit obtained by good borrowers when there are n banks is therefore given by

$$(1-nq_G(1-q_G)^{n-1}-(1-q_G)^n)\gamma(R^{MAX}-1)-(1-n(1-q_B)q_B^{n-1}-q_B^n)(1-\gamma) \tag{4.13}$$

To see whether this is increasing in n, the number of banks, one could in principle simply differentiate the expression with respect to n. However, the resulting derivative cannot be signed in a straightforward way. That is not surprising for there is no reason to think that the good borrowers' profits are increasing monotonically in n for *any* number of banks. However, by comparing good borrowers' profits with n and with $n+1$ banks and making use of the assumption that the number of banks has not yet reached its maximum, it is possible to show that good borrowers' profits are indeed increasing in the number of banks. Aggregate good borrowers' profits with $n+1$ banks are given by

$$\underbrace{(1-(n+1)q_G(1-q_G)^n-(1-q_G)^{n+1})}_{A}\gamma(R^{MAX}-1)$$

$$-\underbrace{(1-(n+1)(1-q_B)q_B^n-q_B^{n+1})}_{B}(1-\gamma) \tag{4.14}$$

In order to obtain the difference between good borrowers' aggregate profits with n and with $n+1$ banks, I first obtain alternative expressions for $A = [1-(n+1)q_G(1-q_G)^n-(1-q_G)^{n+1}]$ and $B = [1-(n+1)(1-q_B)q_B^n-q_B^{n+1}]$:

$$1-(n+1)q_G(1-q_G)^n-(1-q_G)^{n+1}=$$

$$1-nq_G(1-q_G)^{n-1}(1-q_G)-q_G(1-q_G)^n-(1-q_G)^n(1-q_G)=$$

$$1-nq_G(1-q_G)^{n-1}+nq_G^2(1-q_G)^{n-1}-q_G(1-q_G)^n-(1-q_G)^n+(1-q_G)^nq_G=$$

$$1-nq_G(1-q_G)^{n-1}-(1-q_G)^n+nq_G^2(1-q_G)^{n-1}$$

An alternative way of writing $-B=-[1-(n+1)(1-q_B)q_B^n-q_B^{n+1}]$ is obtained as follows:

$$-(1-(n+1)(1-q_B)q_B^n-q_B^{n+1})=$$

$$-(1-n(1-q_B)q_B^{n-1}[1-(1-q_B)]-(1-q_B)q_B^n-q_B^n[1-(1-q_B)])=$$

$$-(1-n(1-q_B)q_B^{n-1}+n(1-q_B)^2q_B^{n-1}-(1-q_B)q_B^n-q_B^n+(1-q_B)q_B^n)=$$

$$-(1-n(1-q_B)q_B^{n-1}-q_B^n+n(1-q_B)^2q_B^{n-1})$$

Substituting the expressions thus obtained in the above expression for good borrowers' aggregate profit with $n+1$ banks, then subtracting from this good borrowers' aggregate profits with n banks and eliminating identical terms, the difference between good borrowers' profits with $(n+1)$ banks and with n banks reduces to

$$nq_G^2(1-q_G)^{n-1}\gamma(R^{MAX}-1)-n(1-q_B)^2q_B^{n-1}(1-\gamma)$$

Thus if I can show this expression to be positive, I will have shown that good borrowers' aggregate profit is increasing in the number of banks. Equivalently I could show that

$$nq_G^2(1-q_G)^{n-1}\gamma(R^{MAX}-1)-n(1-q_B)^2q_B^{n-1}(1-\gamma)>0$$

$$\Leftrightarrow nq_G^2(1-q_G)^{n-1}\gamma(R^{MAX}-1)>n(1-q_B)^2q_B^{n-1}(1-\gamma)$$

$$\Leftrightarrow (\frac{1-q_G}{q_B})^{n-1}(\frac{\gamma}{1-\gamma})(\frac{q_G}{1-q_B})^2(R^{MAX}-1)>1$$

$$\Leftrightarrow (\frac{1-q_G}{q_B})^{n-1}>(\frac{1-\gamma}{\gamma})(\frac{1-q_B}{q_G})^2\frac{1}{(R^{MAX}-1)} \tag{4.15}$$

It is here that the assumption that the number of banks has not yet reached its maximum comes in. I know that when the number of banks has not yet reached its maximum, it is true that

$$(\frac{q_B}{1-q_G})^n\leq(R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B} \tag{4.16}$$

which can also be written as

$$(\frac{1-q_G}{q_B})^n\geq\frac{1-\gamma}{\gamma}\frac{1-q_B}{q_G}\frac{1}{(R^{MAX}-1)} \tag{4.17}$$

Multiplying both sides in (4.17) by $(\frac{q_B}{1-q_G})$ yields

$$(\frac{1-q_G}{q_B})^{n-1} \geq \frac{1-\gamma}{\gamma}\frac{q_B}{1-q_G}\frac{1-q_B}{q_G}\frac{1}{(R^{MAX}-1)} \tag{4.18}$$

Using this result in (4.15) above one obtains

$$\frac{1-\gamma}{\gamma}\frac{q_B}{1-q_G}\frac{1-q_B}{q_G}\frac{1}{(R^{MAX}-1)} > \frac{1-\gamma}{\gamma}(\frac{1-q_B}{q_G})^2\frac{1}{(R^{MAX}-1)}$$

$$\Leftrightarrow \frac{q_G}{1-q_B}\frac{q_B}{1-q_G} > 1 \tag{4.19}$$

which is clearly true. This shows that the good borrowers' aggregate profit is increasing in the number of banks, as long as the number of banks has not yet reached its maximum.

4.8.2 Banks' aggregate profit

A bank breaks exactly even on all those borrowers who have two or more offers. However, up to the maximum number of banks it makes a strictly positive profit on those borrowers who only get one offer - in other words, for whom the bank is a monopolist. With n competitors, banks' aggregate profits are thus simply given by

$$nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1) - n(1-q_B)q_B^{n-1}(1-\gamma) \tag{4.20}$$

Again it is not obvious from differentiating with respect to n whether aggregate bank profits are decreasing in the number of banks (as long as that number has not yet reached its maximum). However, it is possible to show that aggregate bank profits with n banks are greater than aggregate bank profits with $n+1$ banks as long as the number of banks has not yet reached its maximum. First, note that aggregate bank profits with $n+1$ banks are given by

$$\underbrace{(n+1)q_G(1-q_G)^n\gamma(R^{MAX}-1)}_{C}\ \underbrace{-(n+1)(1-q_B)q_B^n(1-\gamma)}_{D}$$

Part C can be rewritten as follows:

$$(n+1)q_G(1-q_G)^n\gamma(R^{MAX}-1) =$$

$$nq_G(1-q_G)^{n-1}(1-q_G)\gamma(R^{MAX}-1)+q_G(1-q_G)^n\gamma(R^{MAX}-1)=$$

$$nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-nq_G^2(1-q_G)^{n-1}\gamma(R^{MAX}-1)+q_G(1-q_G)^n\gamma(R^{MAX}-1)=$$

$$nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)\{n[1-(1-q_G)]-(1-q_G)\}=$$

$$\underbrace{nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)}_{*}-q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)[n-(n+1)(1-q_G)]$$

In the same way, part D can also be rewritten:

$$-(n+1)(1-q_B)q_B^n(1-\gamma)=$$

$$-n(1-q_B)q_B^{n-1}[1-(1-q_B)](1-\gamma)-(1-q_B)q_B^n(1-\gamma)=$$

$$-n(1-q_B)q_B^{n-1}(1-\gamma)+n(1-q_B)^2q_B^{n-1}(1-\gamma)-(1-q_B)q_B^n(1-\gamma)=$$

$$-n(1-q_B)q_B^{n-1}(1-\gamma)+(1-q_B)q_B^{n-1}(1-\gamma)[n(1-q_B)-q_B]=$$

$$\underbrace{-n(1-q_B)q_B^{n-1}(1-\gamma)}_{*}+(1-q_B)q_B^{n-1}(1-\gamma)[n-(n+1)q_B]$$

When subtracting bank profits for $n+1$ banks from the expression for bank profits for n banks, inspection shows that the terms marked with an asterisk cancel out. The difference is thus given by

$$q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)[n-(n+1)(1-q_G)]-(1-q_B)q_B^{n-1}(1-\gamma)[n-(n+1)q_B]$$

If this can be shown to be positive, banks' aggregate profits with n banks are greater than with $n+1$ banks. Alternatively, I can show that

$$(\frac{1-q_G}{q_B})^{n-1}(R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B}[n-(n+1)(1-q_G)] > n-(n+1)q_B$$

Substituting (4.18) on the left-hand side, the inequality reads

$$\frac{1-\gamma}{\gamma}\frac{q_B}{1-q_G}\frac{1-q_B}{q_G}\frac{1}{(R^{MAX}-1)}(R^{MAX}-1)\frac{\gamma}{1-\gamma}\frac{q_G}{1-q_B}[n-(n+1)(1-q_G)] > n-(n+1)q_B$$

Canceling terms, one obtains

$$\frac{q_B}{1-q_G}[n-(n+1)(1-q_G)] > n-(n+1)q_B$$

$$\Leftrightarrow q_B n - q_B(n+1)(1-q_G) > (1-q_G)n - (1-q_G)(n+1)q_B$$

which reduces to

$$q_B > (1-q_G)$$

which is clearly true and proves the claim that banks' aggregate profits are decreasing in the number of banks.

4.8.3 Expected profit for a good borrower

For a good borrower, the expected profit is the product of the probability of obtaining a loan (which must increase in the number of banks) and the profit per borrower (about which *a priori* little can be said as with more banks both good borrowers' aggregate profits and the number of good borrowers among whom these profits are distributed increase). It is, however, possible to show that a good borrower's expected profit is indeed increasing in the number of banks. The expected profit with n banks is given by

$$\frac{\text{Good borrowers' aggregate profits with n banks}}{\text{Number of good borrowers with n banks}} * \Pr ob(m \geq 1|G,n) \tag{4.21}$$

(4.21)

Now note that the number of good borrowers with n banks is simply $\Pr ob(m \geq 1|G,n) * \gamma$. By canceling the probabilities the expected profit with n banks can thus be written simply as

$$\frac{\text{Good borrowers' aggregate profits with n banks}}{\gamma}$$

In the same way, a good borrower's expected profits with $n+1$ banks can be written as

$$\frac{\text{Good borrowers' aggregate profits with n+1 banks}}{\gamma}$$

But it was already shown that the aggregate profits increase in the number of banks, so a good borrower's expected profit with $n+1$ banks must be greater than with n banks.

4.8.4 Expected profit per bank

As noted above, with n competitors banks' aggregate profits are given by

$$nq_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-n(1-q_B)q_B^{n-1}(1-\gamma) \tag{4.22}$$

With $n+1$ competitors, the corresponding expression is

$$(n+1)q_G(1-q_G)^{n}\gamma(R^{MAX}-1)-(n+1)(1-q_B)q_B^{n}(1-\gamma) \tag{4.23}$$

Average (or expected) profits in the two cases are then given by

$$q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-(1-q_B)q_B^{n-1}(1-\gamma)$$

for the case of n competitors and

$$q_G(1-q_G)^{n}\gamma(R^{MAX}-1)-(1-q_B)q_B^{n}(1-\gamma)$$

for $n+1$ banks. I wish to show that - as asserted in the paper - average profits with n banks are greater than average profits with $n+1$ banks. This means showing that

$$q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-(1-q_B)q_B^{n-1}(1-\gamma)$$

$$> q_G(1-q_G)^{n}\gamma(R^{MAX}-1)-(1-q_B)q_B^{n}(1-\gamma)$$

$$\Leftrightarrow q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)-(1-q_B)q_B^{n-1}(1-\gamma)$$

$$> q_G(1-q_G)^{n-1}\gamma(R^{MAX}-1)(1-q_G)-(1-q_B)q_B^{n-1}(1-\gamma)[1-(1-q_B)]$$

$$\Leftrightarrow q_G^2(1-q_G)^{n-1}\gamma(R^{MAX}-1) > (1-q_B)^2 q_B^{n-1}(1-\gamma)$$

$$\Leftrightarrow \frac{q_G}{1-q_B}\frac{q_G}{1-q_B}\frac{\gamma}{1-\gamma}(R^{MAX}-1) > (\frac{q_B}{1-q_G})^{n-1}$$

$$\Leftrightarrow \frac{q_B}{1-q_G}\frac{q_G}{1-q_B}\underbrace{\frac{q_G}{1-q_B}\frac{\gamma}{1-\gamma}(R^{MAX}-1) > (\frac{q_B}{1-q_G})^{n}}$$

Once again making use of the condition that the number of banks has not yet reached its maximum and that, therefore, (as indicated by the bracket)

$$\frac{q_G}{1-q_B}\frac{\gamma}{1-\gamma}(R^{MAX}-1) \geq (\frac{q_B}{1-q_G})^{n}$$

it is shown that the average profit for n banks is greater than the average profit for $n+1$ banks.

Chapter 5

Bank Competition and X-Inefficiency

5.1 Introduction

In the literature review I showed that the theoretical literature is sceptical with respect to bank competition. The empirical literature, however, finds substantial gains from more intense bank competition. I argued that while the two literatures are not asking exactly the same questions, there is enough of an overlap to claim that, painting with a broad brush, these results contradict each other. The main purpose of the present chapter is to offer a possible resolution for this discrepancy.

In a nutshell, my argument is the following: The main role of more intense competition may be that of a discipline device - a point overlooked in the existing banking literature. When other mechanisms of corporate governance fail for whatever reason (and, as the evidence on X-inefficiency in banks suggests, they do fail), a major benefit of more intense competition may be that it forces managers to work harder (which in turn improves the allocation of capital). The existing theoretical work cannot capture this effect because it assumes that banks are profit-maximizing entities. In the present chapter I formalize this conjecture and demonstrate that in the presence of an effort-minimizing bank, surplus unambiguously increases in the intensity of competition. The evidence is certainly consistent with this line of reasoning.

In the next section, I present the building blocks for the model: the way in which I model bank competition and X-inefficiency plus a set of standard assumptions. In section 5.3., I study the benchmark case of two profit-maximizers competing against each other and in the following section, I let a profit-maximizer compete against an effort-minimizer. This section (5.4.) is the main part of the chapter and is followed by a

brief discussion of possible extensions and a conclusion. Many derivations have been moved to the appendix to unburden the exposition.

5.2 The Basic Model: Assumptions and Motivation

In the literature review I argued that certain assumptions recur in the bank competition literature and that models differ mainly in their assumptions with respect to cashflow distributions, the technology that helps to overcome information problems (e.g., a monitoring technology), and the way competition is modeled. In the following sub-section I restate the standard assumptions that I use. I then motivate and discuss the two innovations of this paper. First, on a technical note, the way I model interest-rate competition is novel. Second, in a more substantial departure from the literature I allow for an agency problem on the part of the bank that leads the manager to minimize effort rather than maximize profit.

5.2.1 Standard assumptions

When project cashflow distributions are invariant and common knowledge, there is no role for monitoring or screening and there is no role for competition either: In particular, the interest rate is irrelevant; it merely redistributes surplus but does not affect how much surplus is generated. (This property, however, is not robust to even small modifications. A bankruptcy cost, for example, would ensure that only the lowest feasible interest rate maximizes surplus.) In the following I focus on monitoring, not screening, so I assume that cashflow distributions are common knowledge. However, monitoring alleviates a moral hazard problem which I explain in the next sub-section.

As is commonly assumed in the literature, entrepreneurs are very passive: They all require one unit of capital each to start a project, but have no funds of their own (no equity) and no collateral. There are no bankruptcy costs and no application costs (though I will allow for processing costs to be borne by the bank as part of a horizontal differentiation setup). Projects are homogeneous except for location. Since the assumption that every entrepreneur borrows from a single bank only is maintained, there is - unlike, for example, in Thakor (1996) - no question of monitoring costs being duplicated. All agents are risk-neutral.

The existence of financial intermediaries and the use of standard debt

contracts is assumed. I have three comments to make on these two shortcuts: 1. They are standard in the bank competition literature. That in itself is not a good justification, but it is indicative of the difficulties inherent in building a model that explains everything. 2. Financial intermediaries exist and standard debt contracts are the main contractual form used when these lenders put capital at the disposal of entrepreneurs. 3. Both institutions - that of the financial intermediary as well as the use of standard debt contracts - are endogenized in other contributions.

I stick to a one-shot static setup which, even though this is the approach chosen by much of the literature, strikes me as a more serious misdemeanor than simply assuming the existence of financial intermediaries. Clearly, at least in terms of the number of transactions (if not lending volume) only a small share of real-world lending to businesses occurs in the true arms-length setting that the static setup implies and these are likely transactions that are little different from disintermediated direct finance (typically involving large borrowers that are relatively transparent thanks, for example, to an agency rating). In other words, the specifics of financial intermediation are more appropriately explored by investigating a dynamic model [as, for example, in Petersen and Rajan (1995)] or, possibly, even an infinitely repeated game [as, for example, in Dinç (2000)]. However, a static framework is sufficient to make the point I wish to illustrate.

The loan portfolio of each bank is assumed to consist of a continuum of projects. This avoids the problems associated with aggregate uncertainty. Since banks' cashflows are effectively non-stochastic, it is plausible that a bank should always be able to repay its debts[1]. The rate of return $\rho > 1$ (at which banks can borrow any amount of funds) serves as the benchmark for determining whether a project is surplus-increasing (it is a positive NPV project) or it is not (it is a negative NPV project)[2]. I shall stick to the partial-equilibrium framework that is conventionally used in the bank competition literature and accept ρ as exogenously given (determined by savers' preferences etc). In the absence of a well-founded social welfare function that would allow me to make value judgements with respect to distributional effects, the concept of surplus is the best

[1] I will argue that the bank manager will not default as long as he can avoid it and that he always can avoid default indeed. Lenders know this and will, therefore, not be troubled by the assumption (to be discussed below) that the bank's return is unknown.

[2] There is no discounted cash-flow analysis in any of the following, but I shall nevertheless use the customary phrase 'positive (negative) net present value (NPV) project' for projects that are socially valuable (wasteful) in that they generate (reduce) surplus [they achieve a rate of return greater (lower) than the opportunity cost of capital ρ].

criterion available for judging the relative desirability of different allocations. (I will be using the terms 'surplus' and 'expected surplus', 'loss' and 'expected loss' etc. interchangeably; given the setup, the realized values are the same as the expected values.)

I follow the literature and omit allocative efficiency considerations from the analysis by keeping demand for loans inelastic. This simplifies the model setup and in view of strong evidence of overconfidence may not even be that outrageous an assumption[3]. One reason for this is a desire to focus on what is special about banking and that is banks' role in acquiring and processing information as well as their monitoring function. Another reason is that allowing demand for credit to be elastic means that the precise characteristics of the assumed demand function would drive the results.

5.2.2 Modeling competition

Since I wish to study the effects of more intense competition between banks, I need to find a way to parameterize the intensity of competition. A route chosen by many papers in this strand of research is to model competition as pure price (Bertrand) competition and to compare the outcomes of monopoly and duopoly (effectively perfect competition). This comparison is then said to yield insights as to the effects of more intense competition. The implicit assumption in this is that by comparing the extremes, we learn something about what is happening at intermediate degrees of competition. I do not mean to claim that this approach, for all its lack of granularity, is invalid. However, I wish to make two comments:

First, to the extent we want to take a cue from real-world bank competition, we would want to avoid the extremes of monopoly and perfect competition: The empirical evidence rather consistently suggests that banks enjoy an intermediate degree of market power. Second, many bank competition models are set up almost as natural monopolies. Surely that calls for caution in interpreting the results, especially when comparing monopoly with perfect competition. Bertrand competition will normally lead all profits to disappear which, for example, may make it impossible to acquire any information. That may be too extreme a description of reality.

I believe that limiting attention to the extremes of competition is best understood as an attempt to sidestep tractability issues. A typical setup

[3]See the references in Manove and Padilla (1999) where over-confidence plays an important role.

investigated in the literature is a two-stage game where banks compete in prices (interest rates) in the second stage and in some kind of accuracy variable in the first stage. Modeling competition as a Bertrand duopoly and monopoly, respectively, means that to all intents and purposes there is no strategic interaction at the price competition stage, as there would be if, for example, the competition model used were of the Hotelling (1929) type which easily becomes intractable in such a setup. One alternative approach is to let the interest rate enter parametrically; i.e., let the interest rate itself parameterize the intensity of competition [as, for example, in Chan, Greenbaum and Thakor (1986) or Petersen and Rajan (1995)]. The logic behind this is that more intense competition ought to lead to lower interest rates. One might gain in terms of tractability, but lose a possibly interesting part of the story, i.e., the explicit modeling of the link between competition and the interest rate.

In the following I outline the setup that I employ in this model which should go some way toward addressing the concerns raised above. I use a tweaked version of the Hotelling model in which competition is parameterized in a very smooth fashion (so that results should be somewhat more robust), but which avoids the tractability problems hinted at above. Inelastic demand is neatly captured by the standard 'covered market' assumption and the intensity of competition is quite naturally parameterized by the transport cost parameter t.

I keep the number of banks (one located at each end of the line[4]) fixed at two and assume that a bank located at $x = 0$ incurs a processing cost of xt when lending to a borrower located at x. Conversely, the bank located at $x = 1$ incurs a processing cost of $(1 - x)t$ when lending to the same borrower. Distance (which need not be taken literally and might also represent industry specialization) thus captures a cost difference. t parameterizes competition and is exogenous (controlled by a regulator, for example). In contrast to the standard Hotelling model, I do not restrict banks to setting a single interest rate. In fact, each bank will set an infinity of different interest rates, although that is an artefact of the assumption that applicants constitute a continuum (with a mass normalized to one). Entrepreneurs choose purely on the basis of the stated interest rate (price). As will become clear, that is not a completely innocuous assumption. Some entrepreneurs could achieve a higher expected return by choosing their lender strategically.

Apart from lending money, banks may also provide an additional service which, in keeping with the standard terminology in the literature,

[4]I will abstract from endogenizing the location decisions throughout.

I refer to as "monitoring". If a bank monitors a project, this increases the entrepreneur's expected cashflow as follows. Two production technologies are assumed to be available, one of which will lead the project to be successful (and thus yield conditional-upon-success cashflow $R >> \rho$) with probability p_H (the high-return technology) and one of which will lead to a lower success probability $p_L < p_H$ (the low-return technology). In the absence of monitoring, only the low-return technology is accessible to the borrower. The borrower does not consciously trade-off between the gain from a private benefit (there is none) and a reduction in the expected cashflow of the project as is often assumed in similar monitoring setups [e.g., Holmstrom and Tirole (1997)]. Effort does not play a role. The technology choice problem is exogenous; no matter how intense the competition between banks and, therefore, no matter what the interest rate, a borrower on its own is unable to make use of the high-return technology. By way of motivation, note that the entrepreneur might not even realize that the high-return technology is available. This is not a far-fetched assumption: The loan officers I talked to in the course of my fieldwork provided an abundance of examples of borrowers being totally oblivious to their decisions being misguided. An attentive loan officer could alert the entrepreneur to the scope for pursuing a more suitable course of action. (There are also, of course, situations of genuine disagreement between borrower and lender, but much of the time the loan officer's advice may not be contentious. The banker, talking to many entrepreneurs and having an inside view of many companies, simply develops specific expertise that many a borrower is not reluctant to draw upon.) Therefore, one plausible way of thinking about the monitoring service provided by the bank is to view this monitoring as the provision of advice, technical and financial know-how, and so on with a beneficial effect upon the return distribution. This is also consistent with the way that monitoring is viewed in the literature reviewed in Chapter 2.

I make the following assumptions:

1. The cost of monitoring is a flat per-borrower cost of m in terms of non-contractible effort.

2. Monitoring is assumed to always be profitable, i.e. $(p_H - p_L)\rho - m > 0$. This implies that monitoring is always efficient, i.e. $(p_H - p_L)R - m > 0$.

3. For simplicity, all projects are assumed to be positive NPV projects even when they are not monitored: $p_L R > \rho + t$.

4. Both banks are competing for all applicants. Not only do banks know that the rival is making an offer; interest-rate offers are assumed

to be common knowledge. I assume that the following tie-breaking rule is in use: If both banks offer the same interest rate, the applicant will choose the bank with the lower break-even rate (BER). The idea behind this assumption is that the bank with the lower BER could always cut its offer by an epsilon; rather than fiddle with epsilons I assume that the tie-breaking rule applies. This keeps the notation simple.

5.2.3 Formalizing X-inefficiency

A central building block of many models of bank competition is that incentives for information acquisition (screening) or monitoring are increasing in market power (margins), so that more intense competition may mean lower screening or monitoring effort and less surplus (because it leads to lower margins). My point in the present chapter is that this mechanism only operates if banks are profit-maximizers. The exact opposite may be the case if banks - because of agency conflicts - pursue objectives other than profit-maximization. The most important way in which I depart from the existing literature thus consists of relaxing the assumption of profit-maximization[5]. In view of compelling evidence of banks suffering from X-inefficiency (which I review in section 2.7), it is surely reasonable to ask whether this modification lets earlier results appear in a new light.

I model banks that suffer from an agency problem as effort-minimizers. They monitor as few borrowers as possible, subject to their attaining a profit target $\underline{\pi}$. The banks are not assumed to be totally indifferent to profit above the target level: Preferences are lexicographic in the sense that for given effort, the bank prefers more profit to less. (This feature is helpful in obtaining equilibrium bidding strategies.)

Which primitive assumptions might justify this setup? As is standard in a principal-agent setup, the owner of the bank needs the services of a manager to run the bank, the manager's effort is non-contractible, and the manager does not have the capital to purchase the bank. Of course, since there is no aggregate uncertainty the return on the loan portfolio is non-stochastic. It should be possible to find out exactly what the effort level chosen by the manager was, making it exceedingly easy to construct an incentive scheme. However, I assume that the bank's true profits are

[5] A few contributions in the literature do consider the presence of a moral hazard problem on the part of the bank [for example, Holmstrom and Tirole (1997) and the extension by Almazan (2002)], but this is quite different from an agency problem that leads the bank manager to exert only as much effort as is strictly required to keep the bank afloat. There is also an older literature on expense preference started by Williamson (1963) and adapted to banking, largely in empirical work by, for example, Edwards (1977), Hannan (1979), and Hannan and Mavinga (1980).

not observable. In view of the opaqueness of banks and the well-known scope for manipulating earnings and hiding losses in banking, this is a strong, but not outrageous assumption.

In order to be able to focus on the link between the intensity of competition and the allocation of capital without having to deal with contract-theoretic considerations, I follow Hart and Moore (1995) and Aghion, Dewatripont and Rey (1997, 1999) in assuming away the possibility of more than token incentive compensation. Rather, the contract between owner and manager is similar to a standard debt contract: the manager keeps the job (which provides him or her with some large, but unspecified private benefit) as long as the profit target is met. What makes the owner's threat to fire an underperforming manager credible is that the owner could sell the bank to some other investor who would put the assets to some other use. Bank branches, for example, might be converted to supermarkets. The rate of return that an alternative investor could achieve thereby determines the bank owner's bargaining power and the bank manager's profit target. The competitive rate of return ρ provides a floor, but the banking industry is assumed to be different from other industries in that competition is restricted through regulation of some sort, so that the rate of return in the banking industry is potentially greater than ρ. The bank manager becomes the de facto residual claimant for these potential rents.

The extent to which potential rents are converted into actual profits rather than slack depends on the banker's type. A profit-maximizing banker will not take profit in the form of slack, whereas an effort-minimizer has the lexicographic preferences sketched above and minimizes effort subject to attaining the profit target. I assume that the bank manager is the only person able to run the bank and that the only limit to his or her bargaining power is the owner's outside option of selling the bank to another investor. These assumptions may seem strong, but they correspond surprisingly well to certain features of reality. For example, in practitioner statements talk about target rates of return on equity is pervasive. However, for a given amount of equity, a target return on equity effectively boils down to a profit target.

For a more concrete illustration, consider the fact that around the world much of the banking industry is owned by the public sector. The bankers working in these public-sector banks may be civil servants; at any rate, they will tend to benefit from rather extensive employment guarantees. Remuneration structures in state-owned institutions, be they banks or ministries, are typically fairly rigid and tend to be built around

a system of fixed grades with associated, invariable remuneration levels. Whatever the deeper rationale for such a system (which may well have a justification), it makes it difficult or impossible to use financial incentives to reward performance. If such a system goes hand in hand with an employment guarantee, we are not far from the setup described above. Employment guarantees are not absolute and a banker could still be fired for cause, but it would have to be a very significant and verifiable underperformance. This minimum performance standard is the equivalent of the minimum profit level $\underline{\pi}$ introduced above. Subject to this minimum performance standard being met, it is plausible that the banker will minimize effort as is assumed here as there is no incentive to maximize profit.

Incidentally, one need not consider public-sector banks to find a setting roughly corresponding to the model setup described above. In many industries with industry-wide wage bargaining, similar systems of fairly inflexible employment grades are in use. In fact, one of the reasons why the banks that I visited for my fieldwork were interested in the collaboration was that precisely such a system of fixed grades was in place which senior management wanted to replace with a more performance-related remuneration system. Clearly, the assumption of effort minimization is a strong one, but it is not without plausibility in many real-world institutional contexts.

5.3 Benchmark case: Competition between two profit-maximizers

To establish a benchmark, I first study the case of two profit-maximizing banks. The objective is to find equilibrium bidding strategies, compute profit and surplus, and investigate how these respond to changes in the intensity of competition, parameterized with the cost parameter t. I begin by stating the following

Lemma 9 *Conditional on having won a bid, a profit-maximizing bank monitors all borrowers at the monitoring stage.*

A bank that makes a loan to a borrower against a promised repayment of P can choose between obtaining P with probability p_L and not monitoring or it can incur the cost m and obtain P with probability p_H. It will be more profitable to monitor iff $(p_H P - m) - p_L P > 0 \Leftrightarrow (p_H - p_L)P - m > 0$. It must be the case that P is strictly greater than

the refinancing cost ρ, so $[(p_H - p_L)P - m] > [(p_H - p_L)\rho - m]$. However, $[(p_H - p_L)\rho - m] > 0$ by assumption, so the lemma follows.

I can use this result in obtaining the equilibrium bidding strategies.

Proposition 10 *The interest-rate competition game has the following equilibrium in pure strategies: Let $\underline{P}(k, x, M)$ denote the break-even rate (BER) for bank k lending to a borrower located at x conditional on the bank monitoring the borrower (M). Then bank k bids as follows: For x such that $\underline{P}(k, x, M) < \underline{P}(l, x, M)$ under the tie-breaking rule bank k wins the loan with a bid of $\underline{P}(l, x, M)$, making an expected profit of $p_H[\underline{P}(l, x, M) - \underline{P}(k, x, M)]$. For x such that $\underline{P}(k, x, M) > \underline{P}(l, x, M)$, bank k bids $\underline{P}(k, x, M)$, but under the tie-breaking rule bank l wins the loan with the same bid of $\underline{P}(k, x, M)$, making an expected profit of $p_H[\underline{P}(k, x, M) - \underline{P}(l, x, M)]$*

Proof: Throughout the chapter, I will without loss of generality assume that bank A is located at $x = 0$ and that bank B is located at $x = 1$. Given that both banks will be monitoring all borrowers, the break-even rates are defined by the equations

$$p_H \underline{P}(A, x, M) = \rho + m + xt$$

and

$$p_H \underline{P}(B, x, M) = \rho + m + (1 - x)t$$

so the break-even rates are

$$\underline{P}(A, x, M) = \frac{\rho + m + xt}{p_H} \tag{5.1}$$

and

$$\underline{P}(B, x, M) = \frac{\rho + m + (1 - x)t}{p_H} \tag{5.2}$$

For an applicant located at x, the difference in BERs is given by

$$\underline{P}(A, x, M) - \underline{P}(B, x, M) =$$

$$\frac{\rho + m + xt}{p_H} - \frac{\rho + m + (1 - x)t}{p_H} =$$

$$\frac{1}{p_H}[xt - (1 - x)t] =$$

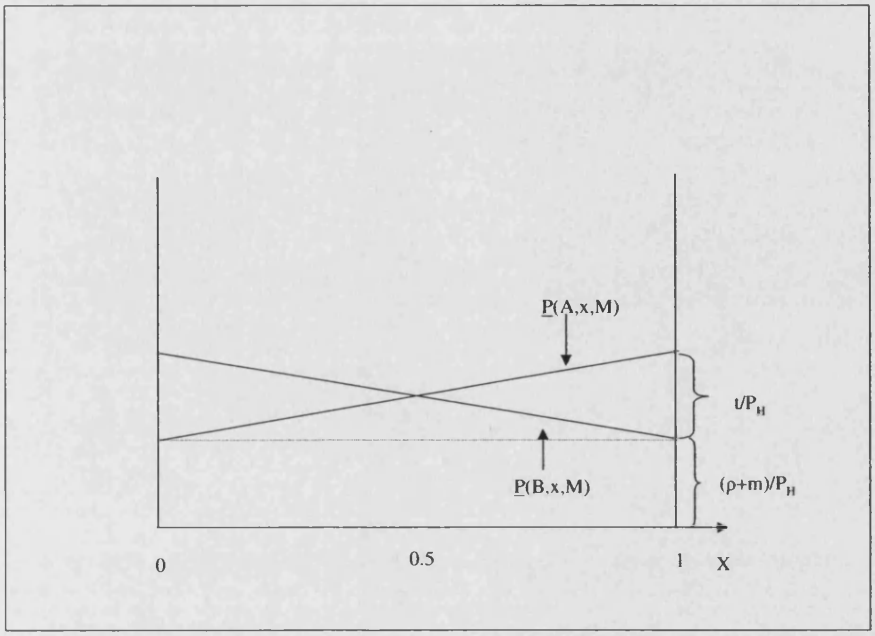
$$\frac{1}{p_H}(2x - 1)t$$

Clearly, this difference is monotonically increasing in x. For $x = 0$, $\underline{P}(B, x, M)$ exceeds $\underline{P}(A, x, M)$ by $\frac{t}{p_H}$. The reverse result is obtained for

an applicant located at $x = 1$: $\underline{P}(A, x, M)$ exceeds $\underline{P}(B, x, M)$ by $\frac{t}{p_H}$. The BERs are equal for $x = \frac{1}{2}$, i.e. each bank takes half the market[6]. The stated strategies are mutually best responses. Given that the lower-cost bank bids at the higher-cost bank's cost, it is an optimal response for the higher-cost bank to bid at cost. If the higher-cost bank bid below cost, it would make a loss so it would be better off if it raised its bid to its BER. If the higher-cost bank asked for a repayment strictly greater than its cost, the lower-cost bank could increase its profit by asking for the same rate as the higher-cost bank, so that cannot be an equilibrium either. Given that the higher-cost bank bids at cost, the lower-cost bank's best response is to ask for the same interest rate for this will ensure that it wins the loan. If it asked for a strictly higher repayment, it would not win the bid and forego a positive profit, so that cannot be an equilibrium. If it offered to lend at a rate strictly lower than the higher-cost bank's bid, it could increase profit by increasing its bid so that cannot be an equilibrium either.

Expected profit is less than the difference in break-even rates, for the winning bank will be repaid with a probability of p_H only. As an illustration, consider the expected profit that A makes on a borrower located at $x < \frac{1}{2}$. A wins the bid asking for a repayment of $\underline{P}(B, x, M)$, which, however, A will receive with a probability of p_H only. The cost of the loan to A is $\rho + m + xt = p_H \underline{P}(A, x, M)$. Taking the difference, $A's$ expected profit is $p_H \underline{P}(B, x, M) - p_H \underline{P}(A, x, M) = p_H[\underline{P}(B, x, M) - \underline{P}(A, x, M)]$. I illustrate this in the following diagram.

[6]The case of $x = \frac{1}{2}$ which is not covered by the tie-breaking rule can be ignored, because a particular point on the distribution has no positive probability mass. For the sake of notational clarity, I assume that entrepreneurs prefer to borrow from bank A when the banks' break-even rates are equal.



Break-even rates when both banks are maximizing profit

The diagram shows the break-even rates for both banks increasing in the borrower's distance from the bank. The bank with the lower break-even rate wins the loan and, conditional on the project's success, obtains the higher break-even rate as repayment. Note that the expected profit is not given by the difference in break-even rates, but by the difference in break-even rates multiplied by p_H.

5.3.1 Profits and Surplus

It is now possible to obtain profit and surplus and see how these respond to changes in t. Profit for each bank is most conveniently obtained as the product of the number of borrowers and the average profit per borrower.

Remark 11 *Alternatively, one can obtain each bank's profit by integration. Given the market shares established above, the profit expressions are then* $\pi_A = \int_0^{\frac{1}{2}} (1-2y)t \, dy$ *and* $\pi_B = \int_{\frac{1}{2}}^1 (2y-1)t \, dy$*, respectively. In view of the simple structure, I find the formulation below more intuitive. To avoid unnecessary notation, in the remainder of the chapter I only state either the integral formulation or the formulation of profit as the product of demand and average per-borrower profit, but I have in each case checked that both approaches give the same result.*

Each bank wins a loan portfolio of size one-half (with the total number of borrowers normalized to one). The average per-borrower profit is given

by the borrower located at $x = \frac{1}{4}$ for A and by the borrower located at $x = \frac{3}{4}$ for B. Therefore,

$$\pi_A = \frac{1}{2}p_H[\underline{P}(B, x = \frac{1}{4}, M) - \underline{P}(A, x = \frac{1}{4}, M)]$$

$$= \frac{1}{2}p_H \left[\frac{\rho + m + (1 - \frac{1}{4})t}{p_H} - \frac{\rho + m + \frac{1}{4}t}{p_H} \right] = \frac{t}{4} \qquad (5.3)$$

$B's$ profit is

$$\pi_B = \frac{1}{2}p_H[\underline{P}(A, x = \frac{3}{4}, M) - \underline{P}(B, x = \frac{3}{4}, M)]$$

$$= \frac{1}{2}p_H \left[\frac{\rho + m + \frac{3}{4}t}{p_H} - \frac{\rho + m + (1 - \frac{3}{4})t}{p_H} \right] = \frac{t}{4} \qquad (5.4)$$

Profit is increasing in t. This is not surprising since an increase in t corresponds to a reduction in the intensity of competition. What about surplus? All applicants obtain a monitored loan and the average processing cost is $\frac{t}{4}$. Surplus is thus given by $p_H R - \rho - m - \frac{t}{4}$ and is increasing in the intensity of competition, because the intensity of competition is parameterized by a processing cost that is a social loss. The latter effect is not surprising, since it is constructed into the model.

Why are monitoring incentives not affected by t even though less intense competition leads to higher margins and therefore makes monitoring more valuable? This result is due to the assumption that $[(p_H - p_L)\rho - m] > 0$, so that it is always not only efficient but also profitable to monitor. The benefit from monitoring is decreasing in the interest rate, but even for small t it is larger than the (fixed) cost of effort it entails. This is enough to induce a profit-maximizing bank to choose the efficient level of monitoring[7]. The properties of the monitoring technology are an empirical issue about which we know little. Here I assume that monitoring has the simple technological structure outlined above.

Lastly, note that the profit expression in the two-profit-maximizers case provides an upper bound for the profit target $\underline{\pi}$: It makes no sense to set a profit target that even a profit-maximizing bank would not be

[7] An analogy from the corporate governance literature might be monitoring by dominant shareholders. Even though a shareholder with less than one-hundred percent ownership does not reap the full benefit from monitoring management (leaving aside tunneling and similar tricks to expropriate other shareholders), the benefit that it does obtain may be sufficient to lead it to monitor efficiently (for example, if monitoring is a 'yes or no' choice as it is here).

able to attain, so it must be the case that

$$\underline{\pi} \leq \frac{t}{4} \tag{5.5}$$

5.4 One profit-maximizer and one effort-minimizer

I next consider the case of one profit-maximizing bank competing against one effort-minimizing bank. This section is at the heart of the chapter and illustrates how more intense competition can operate as a managerial discipline device. The key objective in the following is to demonstrate that in the presence of an effort-minimizing bank, surplus is monotonically decreasing in t (with higher t corresponding to less intense competition) and to analyze the various channels through which t affects surplus. In the two-profit-maximizers case discussed above, the only effect of t on surplus is the direct one: Processing costs are a social loss so higher processing costs mean lower surplus. In the present scenario, the effects of t on surplus are more complex.

I begin by giving a brief overview. I will show that three regimes defined by two particular threshold values of t need to be distinguished: For large enough values of t, the effort-minimizer will not monitor at all; for intermediate values of t the effort-minimizer will be forced to monitor some borrowers and when t is low enough, all of the effort-minimizer's borrowers will be monitored. I will also demonstrate that surplus is continuously decreasing in t.

5.4.1 The effort-minimizer's break-even rate

Without loss of generality, I assume that A is the effort-minimizer (located at $x = 0$) and B the profit-maximizer (located at $x = 1$). For the profit-maximizer, nothing changes: B will continue to monitor all its borrowers; its BER for an applicant located at x is therefore still given by (5.2):

$$\underline{P}(B, x, M) = \frac{\rho + m + (1 - x)t}{p_H}$$

Let $\frac{\rho+m+t}{p_H}$ define I and let $\frac{t}{p_H}$ define F. Then one can also write $\underline{P}(B, x, M)$ as

$$\underline{P}(B, x, M) = I - Fx$$

Given borrower location, a bank that monitors breaks even at a lower rate than if it did not. As I shall show, this gives the profit-maximizer an advantage at the interest-rate competition stage.

As for the effort-minimizer A, I need to state break-even rates both for the case where A monitors and for the case where A does not monitor a borrower. If A monitors, (5.1) applies:

$$\underline{P}(A, x, M) = \frac{\rho + m + xt}{p_H}$$

Let $\frac{\rho+m}{p_H}$ define E. Then one can also write $\underline{P}(A, x, M)$ as

$$\underline{P}(A, x, M) = E + Fx$$

Let U stand for unmonitored. In that case, $A's$ BER is defined by the equation

$$p_L \underline{P}(A, x, U) = \rho + xt$$

so by simple rearranging, $A's$ break-even rate when <u>not</u> monitoring can be shown to be

$$\underline{P}(A, x, U) = \frac{\rho + xt}{p_L} \tag{5.6}$$

Let $\frac{\rho}{p_L}$ define G and let $\frac{t}{p_L}$ define H. Then $\underline{P}(A, x, U)$ can also be written as

$$\underline{P}(A, x, U) = G + Hx$$

The changed notation will simplify some of the derivations below.

The most straightforward way to proceed is to distinguish between the three cases mentioned above (the effort-minimizer monitoring none, some, or all of its borrowers) and establish which values of t lead to which of the three cases.

5.4.2 Regime 1: No monitoring by the effort-minimizer

My approach is to assume that the effort-minimizer can attain its profit-target $\underline{\pi}$ without monitoring, proceed on that assumption and analyze the outcome that obtains to establish under which parameter restrictions the initial assumption holds. Thus by assumption the effort-minimizer A does not monitor, so its BER is given by (5.6). $B's$ BER is given by (5.2). The difference is

$$\underline{P}(A, x, U) - \underline{P}(B, x, M) =$$

$$= G + Hx - (I - Fx)$$

$$= G - I + (H + F)x$$

Bidding strategies and market shares

The difference in break-even rates is clearly a continuous function of and monotonically increasing in x. The equilibrium bidding strategies are essentially unchanged from the game with two profit-maximizers discussed above. I restate the result with the required changes in notation, but the argument of the proof is the same as above and so is not repeated.

Proposition 12 *The interest-rate competition game with one effort-minimizer (by assumption bank A at $x = 0$) and one profit-maximizer (bank B at $x = 1$) has the following equilibrium in pure strategies: Let $\underline{P}(k, x, Z)$ denote the break-even rate (BER) for bank k lending to a borrower located at x conditional on the bank monitoring the borrower ($Z = M$) or not monitoring the borrower ($Z = U$). Then bank k bids as follows: For x such that $\underline{P}(k, x, Z) < \underline{P}(l, x, Z)$,under the tie-breaking rule bank k wins the loan with a bid of $\underline{P}(l, x, Z)$, making an expected profit of $p_H[\underline{P}(l, x, Z) - \underline{P}(k, x, Z)]$ if it monitors the borrower and $p_L[\underline{P}(l, x, Z) - \underline{P}(k, x, Z)]$ if it does not monitor the borrower. For x such that $\underline{P}(k, x, Z) > \underline{P}(l, x, Z)$, bank k bids $\underline{P}(k, x, Z)$, but under the tie-breaking rule bank l wins the loan with the same bid of $\underline{P}(k, x, Z)$, making an expected profit of $p_H[\underline{P}(k, x, Z) - \underline{P}(l, x, Z)]$ if it monitors the borrower or $p_L[\underline{P}(k, x, Z) - \underline{P}(l, x, Z)]$ if it does not monitor the borrower.*

Proof. The previous argument applies analogously. ■

Knowing banks' bidding behavior, it is straightforward to establish market shares (which are equal to a bank's demand). Define $\widetilde{X}(A = U; B = M; t)$ as that value of x such that $\underline{P}(A, x, U) = \underline{P}(B, x, M)$. Then A has a market share of $\max\{0, \widetilde{X}(A = U; B = M; t)\}$ and B has a market share of $\min\{1 - \widetilde{X}(A = U; B = M; t), 1\}$. $\widetilde{X}(A = U; B = M; t)$ is given by $\frac{I-G}{F+H}$. To see that this is correct, note that $A's$ market share $\widetilde{X}(A = U; B = M; t)$ when not monitoring (which I shall refer to as the effortless market share) is given by that value of x for which $\underline{P}(A, x, U) - \underline{P}(B, x, M) = 0$.

$$\underline{P}(A, x, U) - \underline{P}(B, x, M) = G - I + (H + F)x = 0$$

$$\Leftrightarrow x = \frac{I - G}{F + H} = \widetilde{X}(A = U; B = M; t)$$

An explicit expression for $\widetilde{X}(A = U; B = M; t)$ can be obtained as follows:

$$\widetilde{X}(A=U;B=M;t)=\frac{I-G}{F+H}=\frac{\frac{\rho+m+t}{p_H}-\frac{\rho}{p_L}}{\frac{t}{p_H}+\frac{t}{p_L}}$$

$$=\frac{\frac{(\rho+m+t)p_L}{p_Hp_L}-\frac{\rho p_H}{p_Lp_H}}{\frac{tp_L}{p_Hp_L}+\frac{tp_H}{p_Lp_H}}$$

$$=\frac{(\rho+m+t)\,p_L-\rho p_H}{t(p_L+p_H)}$$

$$=\frac{\rho p_L+mp_L+tp_L-\rho p_H}{t(p_L+p_H)}$$

$$=\frac{tp_L}{t(p_L+p_H)}+\frac{\rho p_L+mp_L-\rho p_H}{t(p_L+p_H)}$$

$$=\frac{p_L}{(p_L+p_H)}+\frac{\rho(p_L-p_H)+m(p_L-1+1)}{t(p_L+p_H)}$$

$$=\frac{p_L}{(p_L+p_H)}+\frac{\rho(p_L-p_H)+m-(1-p_L)m}{t(p_L+p_H)}$$

$$=\frac{p_L}{(p_L+p_H)}-\frac{1}{t}\frac{\rho(p_H-p_L)-m+(1-p_L)m}{(p_L+p_H)}$$

Clearly, $\widetilde{X}(A=U;B=M;t)$ is increasing in t [since $[\rho(p_H-p_L)-m]>0$ by assumption] and converges to $\frac{p_L}{(p_L+p_H)}<\frac{1}{2}$ as t tends to infinity. $\widetilde{X}(A=U;B=M;t)$ is positive iff

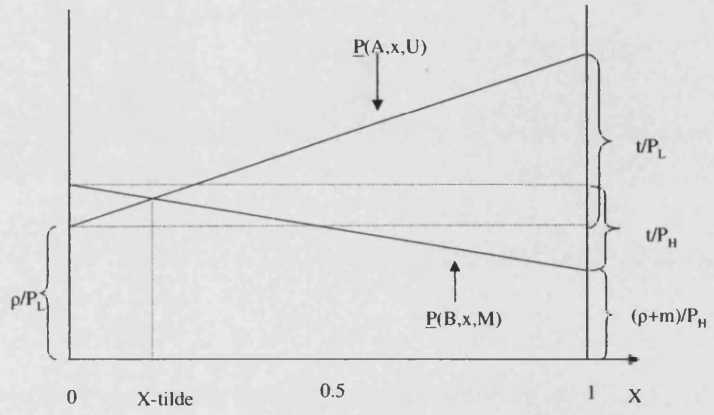
$$\frac{p_L}{(p_L+p_H)}>\frac{1}{t}\frac{\rho(p_H-p_L)-m+(1-p_L)m}{(p_L+p_H)}$$

$$\Leftrightarrow tp_L>\rho(p_H-p_L)-m+(1-p_L)m$$

Let $\rho(p_H-p_L)-m+(1-p_L)m=\rho(p_H-p_L)-p_Lm$ define $C>0$. Then $\widetilde{X}(A=U;B=M;t)>0$ iff

$$t>\frac{C}{p_L} \tag{5.7}$$

When $\widetilde{X}(A=U;B=M;t)$ is positive, A has a strictly positive market share. I have started the discussion by assuming that A can attain its profit target without monitoring so (5.7) has to hold. All that (5.7) ensures is that A actually makes some loans, so it is a necessary but not sufficient condition for A not to have to monitor at all. In the following I shall also write $\widetilde{X}(A=U;B=M;t)$ more briefly as $\widetilde{X}(t)$. I illustrate regime 1 in the following diagram



Regime 1: Break-even rates when the effort-minimizer does not monitor at all

Profit expressions when A does not monitor at all

I can now obtain expressions for π_A (the effort-minimizer's profit) and π_B (the profit-maximizer's profit). $A's$ profit from lending to an entrepreneur at x and not monitoring is

$$p_L\left[\underline{P}(B,x,M) - \underline{P}(A,x,U)\right] = p_L\left[I - G - (H+F)x\right]$$

$$= p_L(H+F)\left[\frac{I-G}{(H+F)} - x\right] = p_L(H+F)\left[\widetilde{X}(t) - x\right]$$

Given market shares, $A's$ total profits are then

$$\pi_A = \int_0^{\widetilde{X}(t)} p_L(H+F)\left[\widetilde{X}(t) - x\right] dx$$

$$= p_L(H+F)\int_0^{\widetilde{X}(t)} \left[\widetilde{X}(t) - x\right] dx$$

$$= p_L(H+F)\left[\widetilde{X}(t)x - \frac{x^2}{2}\right]_0^{\widetilde{X}(t)}$$

$$= p_L(H+F)\frac{1}{2}\left[\widetilde{X}(t)\right]^2$$

$$= p_L \left[\frac{t}{p_L} + \frac{t}{p_H}\right] \frac{1}{2} \left[\widetilde{X}(t)\right]^2$$

$$= p_L \left[\frac{tp_H}{p_L p_H} + \frac{tp_L}{p_H p_L}\right] \frac{1}{2} \left[\widetilde{X}(t)\right]^2$$

$$= t \left[\frac{p_H + p_L}{p_H}\right] \frac{1}{2} \left[\widetilde{X}(t)\right]^2$$

It is clear from inspection that π_A is increasing in t since $\widetilde{X}(t)$ was shown to be increasing in t. Substituting for $\widetilde{X}(t)$:

$$\pi_A = t \left[\frac{p_H + p_L}{p_H}\right] \frac{1}{2} \left[\frac{\rho p_L + m p_L + t p_L - \rho p_H}{t(p_L + p_H)}\right]^2$$

$$= t \left[\frac{p_H + p_L}{p_H}\right] \frac{1}{2} \left[\frac{tp_L - C}{t(p_L + p_H)}\right]^2$$

$$= \frac{1}{p_H} \left[\frac{1}{2(p_L + p_H)}\right] \frac{(tp_L - C)^2}{t}$$

Let D denote the positive constant $2(p_H + p_L)$; $D = 2(p_H + p_L) > 0$. Then $A's$ profit is given by

$$\pi_A = \frac{1}{p_H D} \frac{(tp_L - C)^2}{t} \tag{5.8}$$

π_A is always positive. That is an artefact of the property that mathematically it is quite possible for x and therefore $\widetilde{X}(t)$ to be negative. Given that, as shown above, $\widetilde{X}(t)$ is monotonically increasing in t, that will happen for low enough values of t. In that case, a negative market share is multiplied by a negative average per-borrower margin, leading to an expression that is positive overall. The average per-borrower margin will be negative for negative $\widetilde{X}(t)$ since mathematically it is equal to the margin on the borrower located at $\frac{\widetilde{X}(t)}{2}$. When $\widetilde{X}(t)$ is negative, however, it must be the case that $\widetilde{X}(t) < \frac{\widetilde{X}(t)}{2} < 0$. By definition, break-even rates are equal for $x = \widetilde{X}(t)$, so when $\widetilde{X}(t)$ is negative, B must have a lower BER than A for $x = \frac{\widetilde{X}(t)}{2}$, leading to negative average per-borrower profit for A. In terms of the economics that is, of course, nonsense. Market share and demand cannot be negative.

I had above restricted t to values greater than $\frac{C}{p_L}$ so that A has strictly positive demand; $t = \frac{C}{p_L}$ is the value of t that leads π_A to take the value of zero. I will discuss the shape of $A's$ profit function π_A in more detail in the appendix when I analyze how π_A responds to changes in t. One observation that is usefully made at this point is that $A's$ average per-

borrower profit is increasing in t. To see this, note that $A's$ average per-borrower profit is given by

$$p_L\left(\underline{P}(B, x=\frac{\widetilde{X}(t)}{2}, M)-\underline{P}(A, x=\frac{\widetilde{X}(t)}{2}, U)\right)$$

$$=\frac{1}{p_H}\left(\frac{tp_L-C}{2}\right)$$

While this term may well be negative for small enough t, it is always increasing in t.

$B's$ profit π_B can be obtained in much the same manner. $B's$ profit on a borrower located at x when monitoring (when, by assumption, A is not monitoring) is given by

$$p_H[\underline{P}(A, x, U)-\underline{P}(B, x, M)]$$

$$=p_H[G+Hx-(I-Fx)]$$

$$=p_H[(H+F)\,x-(I-G)]$$

$$=p_H\,(H+F)\,[x-\frac{(I-G)}{(H+F)}]$$

$$=p_H\,(H+F)\,[x-\widetilde{X}(t)]$$

$B's$ profit π_B is then given by

$$\pi_B=\int_{\widetilde{X}(t)}^{1} p_H\,(H+F)\,[x-\widetilde{X}(t)]dx$$

$$=p_H\,(H+F)\int_{\widetilde{X}(t)}^{1}[x-\widetilde{X}(t)]dx$$

$$=p_H\,(H+F)\left[\frac{x^2}{2}-x\widetilde{X}(t)\right]_{\widetilde{X}(t)}^{1}$$

$$=p_H\,(H+F)\left[\frac{1}{2}-\widetilde{X}(t)-\frac{\left(\widetilde{X}(t)\right)^2}{2}+\left(\widetilde{X}(t)\right)^2\right]$$

$$=\frac{1}{2}p_H\left(\frac{t}{p_L}+\frac{t}{p_H}\right)\left[1-2\widetilde{X}(t)+\left(\widetilde{X}(t)\right)^2\right]$$

$$= \frac{1}{2}p_H\left(\frac{tp_H}{p_Lp_H} + \frac{tp_L}{p_Hp_L}\right)\left(1 - \widetilde{X}(t)\right)^2$$

$$= \frac{1}{2}t\left(\frac{p_H + p_L}{p_L}\right)\left(1 - \frac{\rho p_L + mp_L + tp_L - \rho p_H}{t(p_L + p_H)}\right)^2$$

$$= \frac{1}{2}t\left(\frac{p_H + p_L}{p_L}\right)\left(\frac{tp_L + tp_H - \rho p_L - mp_L - tp_L + \rho p_H}{t(p_L + p_H)}\right)^2$$

$$= \frac{1}{2}t\left(\frac{p_H + p_L}{p_L}\right)\left(\frac{tp_H + C}{t(p_L + p_H)}\right)^2$$

$$= \frac{1}{p_L}\left(\frac{1}{2(p_L + p_H)}\right)\frac{(tp_H + C)^2}{t}$$

Once again using $D = 2(p_H + p_L) > 0$, $B's$ profit is given by

$$\pi_B = \frac{1}{p_L D}\frac{(tp_H + C)^2}{t} \tag{5.9}$$

As was the case for A, $B's$ average per-borrower profit is increasing in t. To see this, note from the above derivation that $B's$ average per-borrower profit is given by

$$p_H[\underline{P}(A, x = \frac{1 + \widetilde{X}(t)}{2}, U) - \underline{P}(B, x = \frac{1 + \widetilde{X}(t)}{2}, M)] =$$

$$= \frac{1}{p_L}\left(\frac{tp_H + C}{2}\right)$$

which is increasing in t. In the appendix I analyze how π_B responds to changes in t.

I collect these results in the following

Proposition 13 *Assume that the parameters of the model are such that the effort-minimizer, bank A, can attain its profit target without monitoring any of its borrowers. This implies that* $t > \frac{C}{p_L}$. *In that case,* $\pi_A = \frac{1}{p_H D}\frac{(tp_L - C)^2}{t}$ *and* $\pi_B = \frac{1}{p_L D}\frac{(tp_H + C)^2}{t}$ *(where C and D are positive constants as defined above). Then both* π_A *and* π_B *are increasing in t.*

An explicit lower bound for t

Thus far I have proceeded on the assumption that A can attain its profit target without monitoring. The following result states an explicit suffi-

cient condition for this to be the case.

Proposition 14 *Let $\underline{t}(A = U; B = M)$ denote that value of t such that $\pi_A(A = U) = \underline{\pi}$ when the effort-minimizer (bank A) is not monitoring at all. Then $\underline{t}(A = U; B = M)$ is given by*

$$\frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} + \frac{1}{2(p_L)^2}\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2} \tag{5.10}$$

(The derivation is provided in the appendix.)

I have now provided a necessary and sufficient condition for the effort-minimizing bank not to be monitoring at all: The lack-of-competition parameter t has to be greater than or equal to the value of t, $\underline{t}(A = U; B = M)$, established above. For values of t strictly between $\frac{C}{p_L}$ and $\underline{t}(A = U; B = M)$, $\frac{C}{p_L} < t < \underline{t}(A = U; B = M)$, A has positive market share (demand) and profit even when not monitoring, but does not attain the profit target $\underline{\pi}$. In that case, the only option open to A is to monitor some or all of its borrowers. I will investigate the scenario of $t < \underline{t}(A = U; B = M)$ in cases 2 and 3 below.

Surplus

I now turn to surplus, for this is ultimately the variable of interest. One might conjecture that surplus is invariant to changes in t as long as $t \geq \underline{t}(A = U; B = M)$ since A will not monitor any borrowers when t exceeds its critical value and B monitors all its borrowers irrespective of the value of t. However, that conjecture is not correct.

Proposition 15 *Let $t \geq \underline{t}(A = U; B = M)$. Then the effort-minimizer (bank A) does not monitor any of its borrowers. However, surplus is strictly decreasing in t.*

Proof: See appendix.

Even though the effort-minimizer A does not monitor at all when $t \geq \underline{t}(A = U; B = M)$, surplus is strictly decreasing in t. *A's* (invariant) monitoring decision apart, there are three effects that play a role: Higher t gives A more market share. This lowers the average processing cost (in terms of distance); however, the direct effect of higher t outweighs the indirect effect, so processing costs are increasing in t. Leaving aside processing costs, surplus is decreasing in t because, as noted, higher values of t allow A to take a larger market share. A, however, does not monitor at all. Since monitoring is by assumption efficient, this reduces surplus. One could argue that the positive effect of lower t (more intense

competition) via lower processing costs is assumed into the model. However, my argument here shows that the beneficial effect of competition is robust to this criticism: Surplus before processing costs is likewise decreasing in t, because less intense competition gives the effort-minimizer larger market share.

Summary

In the present sub-section, I studied regime 1 in which the effort-minimizer does not monitor at all. I derived bidding strategies for both banks as well as expressions for market shares and profits and I obtained a lower bound $\underline{t}(A = U; B = M)$ for the lack-of-competition parameter t such that the effort-minimizer can reach its profit target without monitoring. Finally, I also showed that in regime 1 surplus is decreasing in t. All of these results have been obtained for the case of $t \geq \underline{t}(A = U; B = M)$. When this condition does not hold, A will have to monitor at least some of its borrowers to attain the profit target. That is the scenario I investigate next.

5.4.3 Regime 2: The effort-minimizer monitors some, but not all borrowers

If t is less than $\underline{t}(A = U; B = M)$, A (the effort-minimizer) cannot attain its profit-target $\underline{\pi}$ without at least some monitoring. As I show below, there is another threshold value for the lack-of-competition parameter t, t^*, such that when $t \leq t^*$, A has to monitor all of its borrowers. For $t^* < t < \underline{t}(A = U; B = M)$ A has to monitor some, but not all of its borrowers. That is the regime I study in this subsection.

When $t > \frac{c}{p_L}$, $A's$ effortless market share $\widetilde{X}(t)$ and $A's$ effortless profit are strictly greater than zero but insufficient to attain the profit target. In that case, A could increase profit by monitoring in one of two ways: A could monitor thus far unmonitored borrowers already in the portfolio (i.e., with locations $0 \leq x \leq \widetilde{X}(t)$) or A could steal market share from B by monitoring borrowers located between $\widetilde{X}(t)$ and $\frac{1}{2}$ (as the earlier discussion in the two-profit-maximizers case shows, the maximum market share that A could obtain by monitoring all borrowers is $\frac{1}{2}$; for $x > \frac{1}{2}$ $B's$ cost advantage is unassailable). Below I study this decision in more detail.

(5.5) implies a lower bound for t, so for $4\underline{\pi} \leq t \leq \frac{c}{p_L}$ bank A - when not monitoring at all - has no positive market share (demand) and consequently no profits either. $t = \frac{c}{p_L}$ turns out not to be a particularly

interesting value of t; however, for expositional reasons it is nevertheless convenient to consider the two cases $\frac{C}{p_L} < t < \underline{t}(A = U; B = M)$ and $4\underline{\pi} \leq t \leq \frac{C}{p_L}$ separately.

$$\frac{C}{p_L} < t < \underline{t}(A = U; B = M)$$

The only way for the effort-minimizer A to reach the profit target is to monitor. How would A proceed? A would want to monitor as little as is necessary. Given that the per-borrower cost of monitoring is constant, A monitors in the order of profitability; i.e., first the project which it is most profitable to monitor, the next most profitable project next and so on until the profit target is attained. The earlier discussion suggested - I show this rigorously below - that the closer a project is located to $A's$ own location at $x = 0$, the higher $A's$ margin on the project. That reasoning would imply that <u>within</u> each of the two regions $0 \leq x \leq \widetilde{X}(t)$ and $\widetilde{X}(t) < x \leq \frac{1}{2}$, A would begin monitoring from the left, so to speak, i.e. at the lowest value of x. However, it is not obvious whether it is more profitable to monitor hitherto unmonitored existing borrowers or to monitor in order to gain a cost advantage and steal borrowers from B. In particular, it might appear possible that $A's$ best course of action is to begin monitoring in one region only; then monitor in both regions simultaneously until the profit target is attained. That scenario would require that the gain from monitoring in the first region is initially (i.e., for low values of x) greater than in the other region, but that - as the effort-minimizer monitors less and less profitable borrowers located at increasing values of x - the gain from monitoring the most profitable projects in the other region eventually equals the gain from monitoring additional projects located in the region first targeted.

However, the following proposition shows that the order in which A monitors projects and the location of monitored projects can be described very simply:

Proposition 16 *Let* $\frac{C}{p_L} < t < \underline{t}(A = U; B = M)$. *In order to bridge the gap between the effortless profit and the profit target, A obtains additional profit by monitoring borrowers located in the interval* $[0, x^*(\frac{C}{p_L} < t < \underline{t}(A = U; B = M))]$ *where* $x^*(\frac{C}{p_L} < t < \underline{t}(A = U; B = M))$ *is defined in a piecewise fashion as set out below. First, let* t^* *be defined by*

$$t^* = \frac{C(p_H - p_L) + (p_H + p_L)^2\underline{\pi}}{2p_H p_L}$$

$$+\frac{(p_H + p_L)}{2p_H p_L}\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + (p_H + p_L)^2\underline{\pi}^2}$$

For $t^* < t < \underline{t}(A = U; B = M)$, $x^*(\frac{C}{p_L} < t < \underline{t}(A = U; B = M))$ *is given by* $x^*(t^* < t < \underline{t}(A = U; B = M))$ *with the property that* $0 < x^*(t^* < t < \underline{t}(A = U; B = M)) < \widetilde{X}(t^*)$, *i.e., the effort-minimizer (bank A) lends to projects located in the interval* $[0, \widetilde{X}(t)]$, *but only monitors projects located in* $[0, x^*(...)]$ *where* $x^*(...) < \widetilde{X}(t^*)$. $x^*(t^* < t < \underline{t}(A = U; B = M))$ *is defined by*

$$\frac{[(p_H - p_L)t + C]}{(p_H - p_L)t}$$

$$-\frac{1}{(p_H - p_L)t}\sqrt{[(p_H - p_L)t + C]^2 - \frac{(p_H - p_L)(\underline{t} - t)\left[t\underline{t}\,(p_L)^2 - C^2\right]}{(p_H + p_L)\underline{t}}}$$

For $\frac{C}{p_L} < t \leq t^*$, $x^*(\frac{C}{p_L} < t < \underline{t}(A = U; B = M))$ *is given by* $x^*(\frac{C}{p_L} < t \leq t^*)$ *with the property that* $\widetilde{X}(t^*) < x^*(\frac{C}{p_L} < t \leq t^*) < \frac{1}{2}$, *i.e., the effort-minimizer lends to projects located in the interval* $[0, x^*(...)]$ *and monitors all of them.* $x^*(\frac{C}{p_L} < t \leq t^*)$ *is defined by*

$$x^* = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}$$

However, when $\frac{C}{p_L} < t \leq t^*$ *A monitors not only some, but all of its borrowers, so this case is discussed in more detail below.*

This result states that A always begins monitoring at $x = 0$ and monitors projects in the order of increasing distance from $A's$ own location. A threshold value of t called t^* is the largest value of t such that A monitors all the borrowers it wins effortlessly, i.e. projects located between 0 and $\widetilde{X}(t^*)$. For t smaller than the critical value, A has to steal market share from B in order to attain its profit target; this case is studied in regime 3 below.

I first give an overview of the proof: I show that A begins monitoring at $x = 0$ and moves out from there by monitoring projects located successively further away from $A's$ own location (i.e., characterized by increasing x). I then obtain the expression for the critical value $x^*(t^* < t < \underline{t}(A = U; B = M))$ such that A monitors borrowers located in the interval $[0, x^*(t^* < t < \underline{t}(A = U; B = M))]$ with $x^*(t^* < t < \underline{t}(A = U; B = M)) < \widetilde{X}(t^*)$.

By assumption $t > \frac{C}{p_L}$ so that $\widetilde{X}(t) > 0$ and, even without monitoring, A makes a strictly positive effortless profit of $\pi_A = \frac{1}{p_H D}\frac{(tp_L - C)^2}{t}$ as demonstrated above[8]. However, A falls short of the profit target $\underline{\pi}$ and so

[8]In obtaining that expression I had only used the assumption that A does not monitor at all and that B monitors all its borrowers; in particular I had not made use of assumptions with respect to t. Having said that, the profit expression is only meaningful for $t > \frac{C}{p_L}$ which, however, is true by assumption here.

needs to monitor some of the borrowers already in its portfolio and/or win additional market share from B (likewise by monitoring). The additional profit from monitoring a current borrower located at $0 \leq x \leq \widetilde{X}(t)$ is given by

$$p_H\left[\underline{P}(B,x,M) - \underline{P}(A,x,M)\right] - p_L\left[\underline{P}(B,x,M) - \underline{P}(A,x,U)\right]$$

This difference is explained as follows: I subtract from the (expected) profit that A makes on a monitored borrower the (expected) profit that A makes on an unmonitored borrower (at the same location). When monitoring, A obtains the promised profit margin $[\underline{P}(B,x,M) - \underline{P}(A,x,M)]$ with probability p_H; when not monitoring, the profit margin is given by a different expression, $[\underline{P}(B,x,M) - \underline{P}(A,x,U)]$, and is only obtained with probability p_L. Clearly, $[\underline{P}(B,x,M) - \underline{P}(A,x,M)]$ is the same as that obtained in the two-profit-maximizers case when both banks were assumed to be monitoring and is given by $F(1-2x)$. The profit margin $[\underline{P}(B,x,M) - \underline{P}(A,x,U)]$ had been obtained as $[I - G - (F+H)x]$. Note that the monitoring cost m does not appear explicitly in the difference; it is contained in the expression for $A's$ break-even rate $\underline{P}(A,x,M)$. Substituting the expressions for the profit margins, the difference can be written as

$$p_H F(1-2x) - p_L\left[I - G - (F+H)x\right]$$

$$= p_H F - 2p_H Fx - p_L(I-G) + p_L(F+H)x$$

$$= p_H F - p_L(I-G) - \left[2p_H F - p_L(F+H)\right]x$$

This is decreasing in x iff

$$\left[2p_H F - p_L(F+H)\right] > 0$$

$$\Leftrightarrow 2p_H F > p_L(F+H)$$

$$\Leftrightarrow 2p_H\frac{t}{p_H} > p_L\left(\frac{t}{p_H} + \frac{t}{p_L}\right) = p_L\left(\frac{tp_L}{p_H p_L} + \frac{tp_H}{p_L p_H}\right)$$

$$\Leftrightarrow 2 > \frac{p_L + p_H}{p_H}$$

which is clearly true. This formally proves the claim that was made earlier on: This expression is decreasing in x, so the smaller x (i.e., the closer a borrower is located to A) the more profitable it is to monitor that

borrower. Therefore, if A decides to monitor any borrowers already in its loan portfolio (i.e., with $0 \leq x \leq \widetilde{X}(t)$), the bank will begin monitoring at $x = 0$ and will monitor in order of increasing x. Note that this expression gives the actual (expected) profit from monitoring an existing borrower, not profit conditional on the success of the project. The expression stated here has already been corrected for the fact that it will be obtained with a probability of p_H only.

However, in the case of projects located at $\widetilde{X}(t) \leq x \leq \frac{1}{2}$, A can gain a cost advantage by monitoring and underbid B. What is the additional profit from doing so for a borrower located at x? Since the profit when not monitoring is simply zero (it is B that makes the loan in that case), the gain from monitoring is given by $p_H\left[\underline{P}(B,x,M) - \underline{P}(A,x,M)\right]$ which is equal to

$$p_H F(1-2x)$$

$$= p_H F - 2p_H F x$$

(This expression, too, has already been corrected for the fact that it will be obtained with a probability of p_H only.) Once again, this shows formally that the earlier claim is correct and that the profitability of monitoring is decreasing in x for projects located at $\widetilde{X}(t) \leq x \leq \frac{1}{2}$, too. Therefore, if the effort-minimizer were to monitor projects in that range, it would begin at $x = \widetilde{X}(t)$ and then monitor in order of increasing x.

Monitoring in order to gain market share superficially appears to be more profitable than monitoring to increase the return on borrowers already in the portfolio: In the former case, the additional profit is given by $p_H\left[\underline{P}(B,x,M) - \underline{P}(A,x,M)\right]$; in the latter case, it is the same $p_H\left[\underline{P}(B,x,M) - \underline{P}(A,x,M)\right]$ minus something positive $\{p_L\left[\underline{P}(B,x,M) - \underline{P}(A,x,U)\right]\}$. However, these expressions refer to different location ranges. As the proposition claims, it turns out that it is unambiguously more profitable to monitor existing borrowers than to steal market share from B.

The argument is the following: $x = \widetilde{X}(t)$ is the location of both

–the existing borrower that it is least profitable to monitor and

–the additional borrower which it would be most profitable to steal from B.

(By definition of $\widetilde{X}(t)$, borrowers located in $[0, \widetilde{X}(t)]$ are borrowing from A and borrowers located in $(\widetilde{X}(t), 1]$ are borrowing from B, so the statement is not quite correct; the borrower which it would be most profitable to steal from B is located just to the right of $\widetilde{X}(t)$. However, that does not affect the argument.)

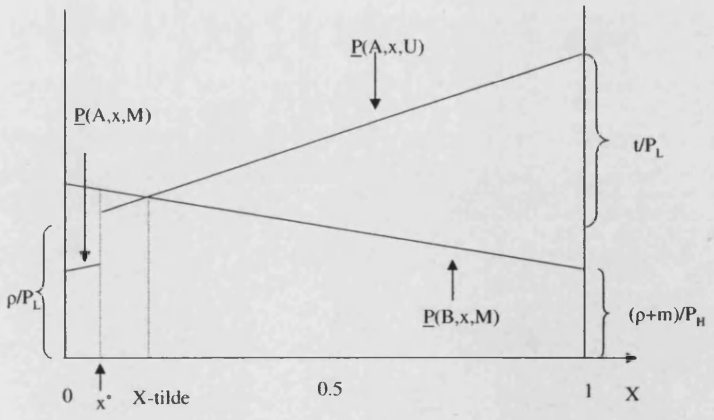
Therefore, if I can show that for the project located at $x = \widetilde{X}(t)$ both types of monitoring yield the same profit, I will have shown that the effort-minimizer always begins monitoring existing borrowers at $x = 0$ and only steals market share from B if, when monitoring all existing borrowers (i.e., all those with locations $0 \leq x \leq \widetilde{X}(t)$), the profit target has not yet been attained. Demonstrating that both types of monitoring yield the same profit for the borrower located at $x = \widetilde{X}(t)$ requires showing that

$$p_H F(1-2x) - p_L\left[I - G - (F+H)x\right] = p_H F(1-2x)$$

$$\Leftrightarrow -p_L\left[I - G - (F+H)x\right] = 0$$

$$\Leftrightarrow \frac{I-G}{(F+H)} = x$$

However, the last equality must be true by the definition of $\widetilde{X}(t)$. That proves the claim. I have shown that A monitors in order of increasing x, beginning at $x = 0$. I illustrate regime 2 in the following diagram.



Regime 2: The effort-minimizer has to monitor some, but not all of its borrowers.

The next task is to obtain an explicit expression for t^*. For a certain value of t which I call t^* in the proposition, A will attain its profit

target by monitoring exactly those projects that would borrow from A anyway. In that case, $x^* = \widetilde{X}(t^*)$. The expression x^* when $t \geq t^*$ is different from the one when $t < t^*$, this is why x^* needs to be defined in a piecewise way. I first derive t^*, then x^* $(t^* < t \leq \underline{t}(A = U; B = M))$ and x^* $\left(\frac{C}{p_L} < t \leq t^*\right)$. In order to obtain an explicit solution for t^* I proceed as follows. I know $A's$ profit when lending to and monitoring the borrowers in the interval $[0, \widetilde{X}(t)]$. I set that profit equal to the profit target which gives me a quadratic equation in t. I obtain the two values of t that solve the equation, only one of them should fulfil the requirement that it be greater than $\frac{C}{p_L}$ and smaller than $\underline{t}(A = U; B = M)$. I show that the " $-$ " root is less than $\frac{C}{p_L}$ and therefore cannot be the value of t that I am looking for. Last, I show that the " $+$ " root is greater than $\frac{C}{p_L}$ and less than $\underline{t}(A = U; B = M)$ and constitutes the solution. The actual calculations have been moved to the appendix to simplify the exposition.

Ultimately, the variable of interest is the interval in which A monitors borrowers, i.e. $[0, x^*(t)]$. Defining t^* is important because the expression for $x^*(t)$ is different for $t > t^*$ and $t \leq t^*$. I first define $x^*(t^* < t < \underline{t}(A = U; B = M))$ and, in the next sub-section, consider $x^*(\frac{C}{p_L} < t \leq t^*)$ (which applies to regime 3 in which A has to monitor all its borrowers in order to attain the profit target). When $t > t^*$, $x^*(t^* < t < \underline{t}(A = U; B = M))$ (which I abbreviate as x^* in the following) is defined by the equation:

$$x^* \frac{1}{p_H} \left[\left(1 - \frac{x^*}{2}\right)(p_H - p_L)t + C \right] + \frac{1}{p_H D} \frac{(tp_L - C)^2}{t} = \underline{\pi} \qquad (5.11)$$

$x^* \frac{1}{p_H} \left[\left(1 - \frac{x^*}{2}\right)(p_H - p_L)t + C \right]$ is the additional profit that A makes by monitoring existing borrowers located in the interval $[0, x^*]$. The additional profit from monitoring an existing borrower located at x is given by $p_H F(1-2x) - p_L \left[I - G - (F+H)x\right] = \frac{1}{p_H} \left[(1-x)(p_H - p_L)t + C\right]$. If A monitors borrowers located in the interval $[0, x^*]$, its additional profit from doing so is given by the product of the number of borrowers monitored (x^*) and the average per-borrower profit which is the profit on the borrower located at $x = \frac{x^*}{2}$. $\frac{1}{p_H D} \frac{(tp_L - C)^2}{t}$ is $A's$ profit in the absence of any monitoring. The sum of these two expressions has to be equal to the profit target $\underline{\pi}$. I derive the x^* thus defined in the appendix where I also show that the following three properties hold:

1. $x^*(t = \underline{t}) = 0$
2. $x^*(t = t^*) = \widetilde{X}(t^*) = \left(\frac{t^* p_L - C}{t^*(p_H + p_L)}\right)$
3. x^* is monotonically decreasing in t and continuous

5.4.4 Regime 3: The effort-minimizer monitors all projects that it lends to

When t falls to t^* or beyond, A needs to monitor all the projects in its loan portfolio. Restating the result above, for $\frac{C}{p_L} < t \leq t^*$, $x^*(\frac{C}{p_L} < t < \underline{t}(A = U; B = M))$ is given by $x^*(\frac{C}{p_L} < t \leq t^*)$ with the property that $\widetilde{X}(t^*) \leq x^*(\frac{C}{p_L} < t \leq t^*) < \frac{1}{2}$, i.e., the effort-minimizer lends to projects located in the interval $[0, x^*(...)]$ and monitors all of them. $x^*(\frac{C}{p_L} < t \leq t^*)$ is defined by

$$x^* = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}$$

To see that this expression is correct, note that I had earlier shown that a bank located at $x = 0$ lending to and monitoring borrowers in the interval $[0, x]$, $x \leq \frac{1}{2}$, makes a profit of $p_H F(1 - 2x) = (1 - 2x)t$ on a borrower located at x. Therefore the total profit from lending to and monitoring borrowers in the interval $[0, x]$, $x \leq \frac{1}{2}$, is given by $x(1-2\frac{x}{2})t = x(1-x)t$ as the product of the number of borrowers and the average per-borrower profit (which is the profit on the borrower located at $\frac{x}{2}$). Since by assumption $t < t^*$, this profit incorporates the effortless profit which does not now need to be considered separately. Then the x^* up to which an effort-minimizer monitors projects is defined by:

$$x^*(1 - x^*)t = \underline{\pi} = \frac{1}{p_H D} \frac{(\underline{t} p_L - C)^2}{\underline{t}} \tag{5.12}$$

This is a quadratic equation in x^* which can be solved as follows:

$$x^*(1 - x^*)t = \underline{\pi}$$

$$\Leftrightarrow x^* t - (x^*)^2 t - \underline{\pi} = 0$$

$$\Leftrightarrow -x^* t + (x^*)^2 t + \underline{\pi} = 0$$

$$\Leftrightarrow (x^*)^2 - x^* + \frac{\underline{\pi}}{t} = 0$$

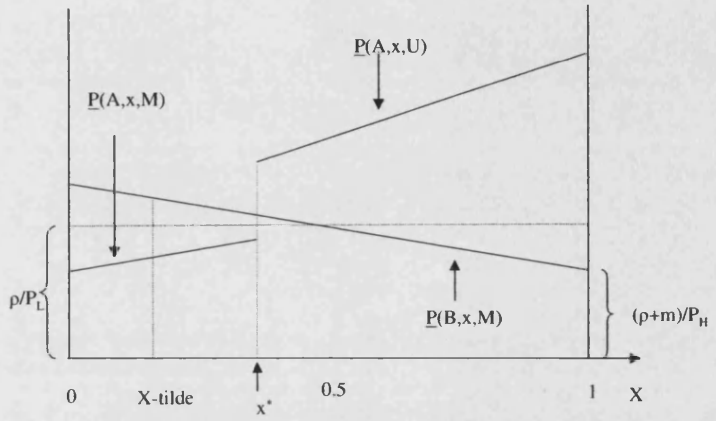
The candidate solutions are

$$(x^*)_{1/2} : \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}$$

The $\{plus\}$ root can be ruled out immediately, since x^* has to be less than or equal to one-half. So for $\frac{C}{p_L} < t \leq t^*$, A monitors borrowers up to x^* where x^* is given by

$$x^* = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}} \tag{5.13}$$

I illustrate this in the following diagram:



Regime 3: The effort-minimizer monitors all the borrowers in the loan portfolio

Note that $x^*\left(\frac{C}{p_L} < t \leq t^*\right)$ is monotonically decreasing in t and continuous for $t > 4\underline{\pi}$:

$$\frac{\partial}{\partial t}\left(\frac{1}{4} - \frac{\underline{\pi}}{t}\right) = \frac{\underline{\pi}}{t^2}$$

Then $\dfrac{\partial x^*\left(\frac{C}{p_L} < t < t^*\right)}{\partial t}$ is given by:

$$\frac{\partial x^*}{\partial t} = \frac{\partial}{\partial t}\left(\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}\right) = (-1)\frac{1}{2}\frac{1}{\sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}}\frac{\underline{\pi}}{t^2} < 0$$

So $x^*\left(\frac{C}{p_L} < t \leq t^*\right)$ is monotonically decreasing in t. Continuity follows from inspection.

These results complete the discussion of the scenario in which the effort-minimizer A makes a strictly positive effortless profit, but not enough to attain the profit target; i.e., $\frac{C}{p_L} < t < \underline{t}(A = U; B = M)$ (composed of the two regimes $\frac{C}{p_L} < t \leq t^*$ and $t^* < t < \underline{t}(A = U; B = M)$). This forces the effort-minimizer to monitor some (regime 2) or all (regime 3) of the borrowers in its portfolio. I show that there is a critical threshold of the lack-of-competition parameter t which I call t^* such that when t is strictly greater than t^*, A lends to those borrowers that it can win effortlessly and monitors only some of them. When t is strictly less than t^*,

A not only lends to its effortless borrowers but also steals market share from B; in that case (as well as when $t = t^*$) A monitors all projects that it lends to. I now briefly turn to the discussion of the case in which A has no effortless profit since it does not win any borrowers without monitoring, i.e. $4\underline{\pi} \leq t \leq \frac{C}{p_L}$.

$4\underline{\pi} \leq t \leq \frac{C}{p_L}$

Whenever $t \leq t^*$, the effort-minimizer monitors all the borrowers that it lends to. Whether t is greater than $\frac{C}{p_L}$ (and the effortless profit thus strictly positive) or not does not matter: The expression defining x^* does not change: $x^*\left(\frac{C}{p_L} < t < t^*\right) = x^*\left(4\underline{\pi} \leq t \leq \frac{C}{p_L}\right)$. Since $x^* = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\pi}{t}}$ is a continuous function of t for $t \geq 4\underline{\pi}$, it is also continuous at $\frac{C}{p_L}$[9]. No particular significance attaches to the value that x^* takes when $t = \frac{C}{p_L}$. This is different for $x^*(t = \underline{t})$ which is zero by definition and $x^*(t = t^*)$ which is equal to $\widetilde{X}(t^*)$, t^* being the highest value of t such that A monitors all borrowers in its lending portfolio. In summary, I can simply restate the previous result and note that $x^*\left(4\underline{\pi} \leq t \leq \frac{C}{p_L}\right)$ is given by

$$x^* = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\pi}{t}}$$

The above result on x^* being a decreasing function of t continues to apply, of course. Note that when $t = 4\underline{\pi}$, the square root term is knocked out and A monitors projects up to $x = \frac{1}{2}$.

5.4.5 Surplus in regimes 2 and 3

The ultimate variable of interest is surplus and how it changes with t. In light of the preceding sub-sections, I note that I need to consider two regimes in which different expressions for x^* apply: Surplus for t greater than t^* ($t^* < t < \underline{t}(A = U; B = M)$) and surplus for t less than t^* ($4\underline{\pi} \leq t \leq t^*$). However, the central result applies across both regimes:

Proposition 17 *For $4\underline{\pi} \leq t < \underline{t}(A = U; B = M)$, surplus is continuously decreasing in t. Surplus is also continuous at $t = \underline{t}(A = U; B = M)$.*

To unburden the exposition I demonstrate this result and analyze the various effects that play a role in the appendix and only give an overview

[9] As noted in the discussion of the two-profit-maximizers case, there is a lower bound for t such that for t below that lower bound, the effort-minimizer A is not be able to attain its profit target regardless of monitoring intensity (i.e., neither would a profit-maximizing bank). That lower bound for t was shown to be $4\underline{\pi}$. I implicitly assume $\frac{C}{p_L}$ to be large enough relative to $4\underline{\pi}$ for the present discussion to make sense.

here: I first consider regime 2 and show that for $t^* < t < \underline{t}(A = U; B = M)$ surplus before processing costs is decreasing in t whereas processing costs (which reduce surplus) are increasing in t. Surplus before processing costs is decreasing in t because lower t leads A to monitor more of its borrowers which increases surplus. Lower t also reduces $A's$ effortless market share, thus increasing the market share of B which monitors all of its borrowers. Processing costs reduce surplus and are increasing in t. The direct effect of higher t assuming a constant effortless market share for A is to increase processing costs; this direct effect dominates the indirect effect which lowers processing costs (because higher t increases $A's$ effortless market share so that more projects borrow from A at lower processing cost than if they borrowed from B). Since processing costs enter surplus with a negative sign, an increase in t increases processing costs and lowers surplus. Overall, then, surplus is decreasing in t.

In regime 3 ($4\underline{\pi} \leq t \leq t^*$), surplus before processing costs is constant; all borrowers are monitored. However, surplus still increases as t falls all the way down to $4\underline{\pi}$. The source of this surplus gain is a reduction in processing costs. As t falls, x^* increases, so more projects borrow from A at lower processing cost than if they borrowed from B. In regime 3, therefore, the direct and indirect effect of t on processing costs operate in the same direction. So surplus is decreasing in t even though surplus before borrowing costs is constant. Interestingly, this property is not entirely assumed into the model. The direct effect of t on processing costs certainly is an artefact of the model setup in which the intensity of competition is captured by the difference in costs. However, the indirect effect (via the impact of t on $A's$ market share x^*) runs in the same direction as the direct effect, so that even if one ignored the direct effect, surplus would still be decreasing in t.

The final task is to show the continuity of surplus. The continuity of surplus follows from standard results except at the two threshold values $t = t^*$ and $t = \underline{t}$. In the appendix, I show that surplus is continuous at these values of t and thus at all values of $t \geq 4\underline{\pi}$.

5.4.6 Summary

In this sub-section, I let a profit-maximizing bank compete against an effort-minimizer. I describe how the effort-minimizer monitors borrowers as a function of the intensity of competition and I show that there are two threshold values of t (which parameterizes the intensity of competition). For $t \geq \underline{t}(A = U; B = M)$ (regime 1) defined in (5.10), the effort-minimizer does not monitor at all. For $t^* < t < \underline{t}(A = U; B = M)$

(regime 2) [with t^* defined by (5.14)], the effort-minimizing bank will monitor some but not all of its borrowers. Finally, for $t \leq t^*$ (regime 3), the effort-minimizer monitors all borrowers in its portfolio. Building on these results, I show that surplus is continuously increasing in the intensity of competition (with lower values of t corresponding to more intense competition) and that, in particular, surplus is continuous at the threshold values of t. I also decomposed changes in surplus and analyzed the different effects that play a role.

5.5 Possible extensions

While I am confident that the present contribution captures a first-order effect that has been overlooked in earlier literature, it will not be the last word in the debate on the optimum intensity of bank competition. Which extensions appear particularly interesting? Within the context of the present model, I do not think that one would gain much additional insight by letting two effort-minimizers compete; that scenario is rather close to collusion. It might be more interesting to model the different technologies available to the borrower in a more graduated way. I have kept this element as simple as possible in the above; however, it would be worthwhile to see whether the results are robust to a formulation in which efficient monitoring choices by the profit-maximizing bank are not a given.

Would the results change if the borrower (entrepreneur) were also an effort-minimizer and would work harder (and generate more social surplus) if she had to pay higher interest rates? Quite possibly, yes. However, I do not believe that scenario to be very plausible. I think of banks' small- and medium-sized business borrowers as rather entrepreneurial (profit-maximizing). Only the evidence can tell whether this conjecture is correct. While there is work investigating the effect of financial market pressures on large companies (finding that debt can be an effective discipline device), I am not aware of a dataset that establishes a disciplinary role of debt for smaller, more entrepreneurial companies.

In the above, strategic interaction in the sense of cooperation between effort-minimizer and profit-maximizer has been ignored. However, I think that there might be more scope for collusion when one of the player minimizes effort rather than maximizes profit. This aspect is not captured here, but may be worth looking into.

Adding a screening (creditworthiness test) stage would be interesting far beyond the model developed here. While there are general contract

theory problems combining moral hazard and adverse selection, the interaction has, to the best of my knowledge, not been considered in the bank competition literature (i.e., the assumptions that the cashflow distribution is invariant and common knowledge are not relaxed simultaneously.) Schnitzer (1999a) is a partial exception which discusses the implications of allowing for borrower moral hazard in a screening model, but does not feature monitoring. Allowing for both screening and monitoring would give rise to interesting questions. For example, would screening intensity and monitoring effort be substitutes or complements? In concrete real-world terms, would a bank with lax credit standards simply be a more active monitor in such a way as to equalize the marginal cost and benefits for both activities? Or are there links that make it impossible to make one decision independently of the other? The very definition of what a positive NPV project is may depend on whether it is monitored and with what intensity it is monitored. I conjecture that much will depend on the assumed cashflow characteristics and properties of the technologies.

Lastly, I think that more robust modeling of information acquisition with private-value components that reduce the value of free-riding on rivals' creditworthiness tests that is such a pervasive real-world phenomenon would be a promising route.

5.6 Conclusion

In the model developed in this chapter I explored a mechanism that reconciles the existing theoretical work on bank competition with the empirical literature. As I showed in the literature review, the theoretical literature is sceptical with respect to the effects of more intense bank competition. The empirical literature, however, was shown to have found large gains from more intense bank competition. I hope that the conjecture explored in the present chapter, namely that banks (and agents generally) may not maximize profits and that it is worthwhile to study the implications this has, will prove fruitful and be developed further by others.

5.7 Appendix

5.7.1 Regime 1: Profits are increasing in t

Since increases in t indicate a reduction in the intensity of competition, one would expect profit to be increasing in t. For sensible parameter values that turns out to be the case indeed. Profit is the product of market share (demand) and average per-borrower profit and, in the case of A, both have been shown to be increasing in t. Things are less straightforward in the case of B. $B's$ average per-borrower profit is increasing in t, but its market share (demand) is decreasing in t. In the following I show formally that both profit expressions are increasing in t.

Beginning with π_A, it is instructive to write the expression as follows:

$$\pi_A = \frac{1}{p_H D}\frac{(tp_L - C)^2}{t} =$$

$$\frac{1}{p_H D}\left[\frac{(tp_L)^2}{t} - \frac{2tp_L C}{t} + \frac{C^2}{t}\right] =$$

$$\frac{1}{p_H D}\left[t\,(p_L)^2 - 2p_L C + \frac{C^2}{t}\right]$$

This expression describes a hyperbola; π_A goes toward minus infinity for $t \to 0$ from below and toward plus infinity for $t \to 0$ from above. Differentiating π_A with respect to t yields

$$\frac{\partial \pi_A}{\partial t} = \frac{1}{p_H D}\left[(p_L)^2 - \frac{C^2}{t^2}\right]$$

The first-order condition suggests that π_A might take extreme values for the values of t defined by

$$\left[(p_L)^2 - \frac{C^2}{t^2}\right] = 0$$

$$\Leftrightarrow (p_L)^2 = \frac{C^2}{t^2}$$

$$\Leftrightarrow t^2 = \frac{C^2}{(p_L)^2}$$

The solutions are $t_1 = \frac{C}{p_L}$ and $t_2 = \frac{(-1)C}{p_L}$. To check the second-order condition, I differentiate π_A with respect to t a second time:

$$\frac{\partial^2 \pi_A}{\partial t^2} = \left(\frac{1}{p_H D}\right)\frac{2C^2}{t^3}$$

Therefore π_A takes a minimum for $t_1 = \frac{C}{p_L}$ and a maximum for $t_2 = \frac{(-1)C}{p_L}$. However, a look at the economics shows that only a subset of values of t make economic sense as opposed to being mathematically valid. First, t clearly cannot be negative. Second, I had already restricted t to be greater than $\frac{C}{p_L}$ so that A has positive market share (demand); $\frac{C}{p_L}$ is exactly the value of t for which π_A reaches a minimum (of zero). In sum, for sensible values of t, i.e., $t > \frac{C}{p_L}$, π_A is monotonically increasing as the earlier discussion suggested.

The discussion of the effects of t on π_B proceeds along analogous lines:

$$\pi_B = \frac{1}{p_L D}\frac{(tp_H + C)^2}{t} =$$

$$\frac{1}{p_L D}\left(\frac{t^2 (p_H)^2}{t} + \frac{2tp_H C}{t} + \frac{C^2}{t}\right) =$$

$$\frac{1}{p_L D}\left(t (p_H)^2 + 2p_H C + \frac{C^2}{t}\right)$$

The shape of $B's$ profit function is quite similar to that of π_A. Differentiating π_B with respect to t yields

$$\frac{\partial \pi_B}{\partial t} = \frac{1}{p_L D}\left((p_H)^2 - \frac{C^2}{t^2}\right)$$

The necessary condition for π_B to take extreme values is therefore given by

$$\left((p_H)^2 - \frac{C^2}{t^2}\right) = 0$$

$$\Leftrightarrow (p_H)^2 = \frac{C^2}{t^2}$$

$$\Leftrightarrow t^2 = \frac{C^2}{(p_H)^2}$$

The solutions are $t_1 = \frac{C}{p_H}$ and $t_2 = \frac{(-1)C}{p_H}$. To check the second-order condition, differentiate π_B with respect to t a second time:

$$\frac{\partial^2 \pi_B}{\partial t^2} = \left(\frac{1}{p_L D}\right)\frac{2C^2}{t^3}$$

So π_B takes a minimum for $t_1 = \frac{C}{p_H}$ and a maximum for $t_2 = \frac{(-1)C}{p_H}$. No attention needs to be paid to negative values of t, but could it be that there are reasonable values of t such that π_B is not monotonically increasing in t (as the earlier discussion suggested might happen)? It turns out that the answer is no; for sensible values of t the positive effect of t on average per-borrower profit will dominate the negative effect of

t on $B's$ market share (demand) and π_B will be strictly increasing in t. To see this, note that *sensible values of* t can be translated as $t > \frac{C}{p_L}$. π_B must be increasing in t for $t > \frac{C}{p_H}$. Since $\frac{C}{p_L} > \frac{C}{p_H}$ and by assumption $t > \frac{C}{p_L}$, t must also be greater than $\frac{C}{p_H}$.

5.7.2 Regime 1: Derivation of an explicit expression for $\underline{t}(A = U; B = M)$

Let $\underline{\pi} > 0$ denote the profit target. A will be able to attain the target without monitoring if and only if

$$\pi_A = \frac{1}{p_H D}\frac{(tp_L - C)^2}{t} \geq \underline{\pi}$$

$$\Leftrightarrow t^2 (p_L)^2 - 2tp_L C + C^2 \geq p_H D\underline{\pi}t$$

The values of t that will cause this relation to hold with equality (and only one of which, in the light of the earlier discussion, will be economically meaningful) are given by the quadratic equation

$$t^2 (p_L)^2 - 2p_L Ct - p_H D\underline{\pi}t + C^2 = 0$$

$$\Leftrightarrow t^2 - \frac{2p_L C + p_H D\underline{\pi}}{(p_L)^2}t + \frac{C^2}{(p_L)^2} = 0$$

Mathematically, the solutions are given by

$$t_{1/2} = \frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} \pm \sqrt{\left(\frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2}\right)^2 - \frac{C^2}{(p_L)^2}}$$

$$= \frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} \pm \sqrt{\left(\frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2}\right)^2 - \frac{2^2 (p_L)^2}{2^2 (p_L)^2}\frac{C^2}{(p_L)^2}}$$

$$= \frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2}$$

$$\pm \frac{1}{2(p_L)^2}\sqrt{2^2 (p_L)^2 C^2 + 2 * 2p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2 - 2^2 (p_L)^2 C^2}$$

$$= \frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} \pm \frac{1}{2(p_L)^2}\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2}$$

From the preceding discussion, I know that the correct solution must be greater than $\frac{C}{p_L}$. The "$-$" root does not fit the bill:

$$\frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} - \frac{1}{2(p_L)^2}\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2} < \frac{2p_L}{2p_L}\frac{C}{p_L}$$

$$\Leftrightarrow 2p_LC + p_H D\underline{\pi} - \sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2} < 2p_LC$$

$$\Leftrightarrow p_H D\underline{\pi} - \sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2} < 0$$

$$\Leftrightarrow p_H D\underline{\pi} < \sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2}$$

$$\Leftrightarrow (p_H D\underline{\pi})^2 < 4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2$$

$$\Leftrightarrow 4p_LCp_H D\underline{\pi} > 0$$

which is clearly true. However, the "+" solution is greater than $\frac{C}{p_L}$ and therefore constitutes the critical value of t that I am looking for:

$$\frac{2p_LC + p_H D\underline{\pi}}{2(p_L)^2} + \frac{1}{2(p_L)^2}\sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2} > \frac{2p_L}{2p_L}\frac{C}{p_L}$$

$$\Leftrightarrow 2p_LC + p_H D\underline{\pi} + \sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2} > 2p_LC$$

$$\Leftrightarrow p_H D\underline{\pi} + \sqrt{4p_LCp_H D\underline{\pi} + (p_H D\underline{\pi})^2} > 0$$

which must be true.

5.7.3 Regime 1: Surplus is decreasing in t

I first obtain an expression for surplus. All applicants will obtain loans, but only a fraction $(1-\widetilde{X}(t))$ will be monitored. Surplus (before processing costs) on a monitored borrower is $(p_H R - \rho - m)$ whereas surplus (before processing costs) on an unmonitored borrower is $(p_L R - \rho)$. The average processing cost incurred by A (as a function of average borrower distance from A) is $\frac{\widetilde{X}(t)}{2}t$; the average processing cost incurred by B is $\frac{(1-\widetilde{X}(t))}{2}t$. Aggregate surplus S is then given by $S = \widetilde{X}(t)(p_L R - \rho - \frac{\widetilde{X}(t)}{2}t) + (1-\widetilde{X}(t))\left[p_H R - \rho - m - \frac{(1-\widetilde{X}(t))}{2}t\right]$.

I first rearrange this expression slightly, as it will be instructive to consider surplus before processing costs and (distance-dependent) processing costs separately.

$$S = \widetilde{X}(t)(p_L R - \rho) - \frac{\left(\widetilde{X}(t)\right)^2}{2}t + (1-\widetilde{X}(t))(p_H R - \rho - m) - \frac{(1-\widetilde{X}(t))^2}{2}t$$

$$= \widetilde{X}(t)(p_L R - \rho) + (1-\widetilde{X}(t))(p_H R - \rho - m) - \frac{t}{2}\left[\left(\widetilde{X}(t)\right)^2 + (1-\widetilde{X}(t))^2\right]$$

$$= (p_H R - \rho - m) - \widetilde{X}(t)(p_H R - \rho - m - p_L R + \rho)$$
$$-\frac{t}{2}\left[\left(\widetilde{X}(t)\right)^2 + 1 - 2\widetilde{X}(t) + \left(\widetilde{X}(t)\right)^2\right]$$

$$= \underbrace{(p_H R - \rho - m) - \widetilde{X}(t)\left[(p_H - p_L)R - m\right]}_{\text{surplus before processing costs}} - \underbrace{\frac{t}{2}\left[1 - 2\widetilde{X}(t) + 2\left(\widetilde{X}(t)\right)^2\right]}_{\text{processing costs}}$$

Since I need to differentiate S with respect to t, it will be convenient to have an expression for $\frac{\partial \widetilde{X}(t)}{\partial t}$. Recall that $\widetilde{X}(t) = \frac{tp_L - C}{t(p_H + p_L)} = \frac{p_L}{(p_H + p_L)} - \frac{C}{t(p_H + p_L)}$. Therefore, $\frac{\partial \widetilde{X}(t)}{\partial t} = \frac{-(-1)C(p_H + p_L)}{t^2(p_H + p_L)^2} = \frac{C}{t^2(p_H + p_L)}$. As the preceding rearrangement shows, $\frac{\partial S}{\partial t}$ can be written as the difference between $\frac{\partial(\text{surplus before proc. costs})}{\partial t}$ and $\frac{\partial(\text{proc. costs})}{\partial t}$. I next obtain expressions for these terms and then establish the overall sign of $\frac{\partial S}{\partial t}$.

I begin with surplus before processing costs:

$$\frac{\partial\left(\text{surplus before proc. costs}\right)}{\partial t} =$$

$$\frac{\partial\left((p_H R - \rho - m) - \widetilde{X}(t)\left[(p_H - p_L)R - m\right]\right)}{\partial t} =$$

$$(-1)\frac{\partial \widetilde{X}(t)}{\partial t}\left[(p_H - p_L)R - m\right] =$$

$$(-1)\frac{C}{t^2(p_H + p_L)}\left[(p_H - p_L)R - m\right] < 0$$

Later on, it will be convenient to write this as

$$(-1)\frac{2(p_H + p_L)C}{2t^2(p_H + p_L)^2}\left[(p_H - p_L)R - m\right]$$

Inspection suggests that surplus before processing costs is decreasing in t.

Processing costs change with t as follows (note that I omit the negative sign; I will re-insert it when I take the difference):

$$\frac{\partial\left(\text{proc. costs}\right)}{\partial t} = \frac{\partial \frac{t}{2}\left[1 - 2\widetilde{X}(t) + 2\left(\widetilde{X}(t)\right)^2\right]}{\partial t} =$$

$$\frac{1}{2}\left[1 - 2\widetilde{X}(t) + 2\left(\widetilde{X}(t)\right)^2\right] + \frac{t}{2}\left[-2\frac{\partial \widetilde{X}(t)}{\partial t} + 2 * 2\widetilde{X}(t)\frac{\partial \widetilde{X}(t)}{\partial t}\right] =$$

$$\frac{1}{2} - \widetilde{X}(t) + \left(\widetilde{X}(t)\right)^2 + t\left[(-1)\frac{\partial \widetilde{X}(t)}{\partial t} + 2\widetilde{X}(t)\frac{\partial \widetilde{X}(t)}{\partial t}\right] =$$

$$\underbrace{\frac{t^2(p_H + p_L)^2}{t^2(p_H + p_L)^2}\frac{1}{2} - \frac{2t(p_H + p_L)}{2t(p_H + p_L)}\frac{(tp_L - C)}{t(p_H + p_L)} + \frac{2\left(tp_L - C\right)^2}{2t^2(p_H + p_L)^2}}_{} + t\frac{\partial \widetilde{X}(t)}{\partial t}(2\widetilde{X}(t) - 1)$$

Note that the first three terms grouped together here capture what one might call the direct effect of higher t on surplus; i.e., the effect of changes

in t on processing costs assuming an unchanged split of the market $\widetilde{X}(t)$. These terms now have the common denominator $2t^2(p_H + p_L)^2$. This allows me to consolidate terms in the numerator as follows:

$$t^2(p_H+p_L)^2 - 2t(p_H+p_L)\left(tp_L - C\right) + 2\left(tp_L - C\right)^2 =$$

$$= t^2\left[(p_H)^2 + 2p_Hp_L + (p_L)^2\right] - 2t(p_H+p_L)tp_L$$

$$+2t(p_H+p_L)C + 2\left(t^2(p_L)^2 - 2tp_LC + C^2\right) =$$

$$= t^2(p_H)^2 + 2p_Hp_Lt^2 + t^2(p_L)^2 - 2t^2p_Hp_L - 2t^2(p_L)^2$$

$$+2tp_HC + 2tp_LC + 2t^2(p_L)^2 - 4tp_LC + 2C^2$$

Cancelling terms, I am left with

$$t^2(p_H)^2 + t^2(p_L)^2 + 2tp_HC - 2tp_LC + 2C^2 =$$

$$t^2(p_H)^2 + t^2(p_L)^2 + 2Ct(p_H - p_L) + 2C^2$$

Given that the denominator $2t^2(p_H + p_L)^2$ clearly must be positive, this establishes that the direct effect of t on surplus is the intuitive one: Higher t leads to higher processing costs which (when multiplied with the negative sign that I omitted here) lead to lower surplus. Turning now to the term $+t\frac{\partial\widetilde{X}(t)}{\partial t}(2\widetilde{X}(t) - 1)$, I note that this can be written as:

$$+t\frac{\partial\widetilde{X}(t)}{\partial t}(2\widetilde{X}(t) - 1) =$$

$$+t\frac{2}{2}\frac{C}{t^2(p_H+p_L)}\left[\frac{2\left(tp_L - C\right)}{t(p_H+p_L)} - \frac{t(p_H+p_L)}{t(p_H+p_L)}\right] =$$

$$+\frac{2C}{2t(p_H+p_L)}\left[\frac{2tp_L - 2C - tp_H - tp_L}{t(p_H+p_L)}\right] =$$

$$+\frac{2C}{2t^2(p_H+p_L)^2}\left(tp_L - 2C - tp_H\right) =$$

$$+\frac{(-1)2C}{2t^2(p_H+p_L)^2}\left[2C + t(p_H - p_L)\right] =$$

$$+\frac{-4C^2 - 2Ct(p_H - p_L)}{2t^2(p_H+p_L)^2}$$

The preceding expression captures the indirect effect of t via changes in $\widetilde{X}(t)$ induced by changes in the intensity of competition. The indirect effect clearly is negative. A decrease in the intensity of competition leads to a redistribution of market share from B to A and reduces the average distance that a borrower has to travel. Keeping everything else equal,

this reduces processing costs and increases surplus. I can now use the results of these computations to obtain an expression for $\frac{\partial(\text{proc. costs})}{\partial t}$ that consolidates the direct and indirect effects:

$$\frac{\partial\,(\text{proc. costs})}{\partial t} =$$

$$= \frac{t^2\,(p_H)^2 + t^2\,(p_L)^2 + 2Ct(p_H - p_L) + 2C^2}{2t^2(p_H + p_L)^2} + \frac{-4C^2 - 2Ct(p_H - p_L)}{2t^2(p_H + p_L)^2}$$

Since these fractions have a common denominator, I can add them to get:

$$\frac{t^2\,(p_H)^2 + t^2\,(p_L)^2 + 2Ct(p_H - p_L) + 2C^2 - 4C^2 - 2Ct(p_H - p_L)}{2t^2(p_H + p_L)^2}$$

Cancelling terms, I am left with

$$\frac{t^2\,(p_H)^2 + t^2\,(p_L)^2 - 2C^2}{2t^2(p_H + p_L)^2}$$

If I can show that this expression is positive, I will have shown that overall processing costs are increasing in t. The denominator must be positive, so I need to show that $t^2\,(p_H)^2 + t^2\,(p_L)^2 - 2C^2 > 0$. Recall that by assumption $t > \frac{C}{p_L}$. Then it must be the case that

$$t^2\,(p_H)^2 + t^2\,(p_L)^2 - 2C^2 > \left(\frac{C}{p_L}\right)^2 (p_H)^2 + \left(\frac{C}{p_L}\right)^2 (p_L)^2 - 2C^2 > 0$$

$$\Leftrightarrow \left(\frac{C}{p_L}\right)^2 (p_H)^2 - C^2 > 0$$

$$\Leftrightarrow C^2\,(p_H)^2 > C^2\,(p_L)^2$$

which clearly is true. So processing costs are increasing in t. When multiplied with a negative sign (as the term for processing costs will be in the surplus expression), the effect of increases in t on surplus via processing costs is negative: Higher t leads to higher processing costs and lower surplus. The proof is now substantially complete. I have shown that surplus before processing costs is decreasing in t and that processing costs (which enter surplus with a negative sign) are increasing in t. For completeness I combine these two partial results:

$$\frac{\partial S}{\partial t} = (-1)\frac{2(p_H + p_L)C}{2t^2(p_H + p_L)^2}\left[(p_H - p_L)\,R - m\right] - \frac{t^2\,(p_H)^2 + t^2\,(p_L)^2 - 2C^2}{2t^2(p_H + p_L)^2} < 0$$

$$\Leftrightarrow (-1)2(p_H + p_L)C\left[(p_H - p_L)\,R - m\right] - \left[t^2\,(p_H)^2 + t^2\,(p_L)^2 - 2C^2\right] < 0$$

I now use the result obtained above demonstrating that $t^2 (p_H)^2 + t^2 (p_L)^2 - 2C^2 > \left(\frac{C}{p_L}\right)^2 (p_H)^2 + \left(\frac{C}{p_L}\right)^2 (p_L)^2 - 2C^2 > 0$. This means that

$$(-1)2(p_H + p_L)C\left[(p_H - p_L)R - m\right] - \left[t^2 (p_H)^2 + t^2 (p_L)^2 - 2C^2\right]$$

$$< (-1)2(p_H + p_L)C\left[(p_H - p_L)R - m\right]$$
$$- \left[\left(\frac{C}{p_L}\right)^2 (p_H)^2 + \left(\frac{C}{p_L}\right)^2 (p_L)^2 - 2C^2\right]$$
$$< (-1)2(p_H + p_L)C\left[(p_H - p_L)R - m\right] < 0$$

which must be true and establishes the claim made in the proposition.

5.7.4 Derivation of t^*

Consider a bank located at $x = 0$ lending to and monitoring projects in the interval $[0, x]$, $x \leq \frac{1}{2}$. This is exactly the behavior of a profit-maximizing bank. The profit that a profit-maximizing bank at $x = 0$ competing against another profit-maximizing bank makes on a borrower located at $x \leq \frac{1}{2}$ had already been obtained as

$$p_H[\underline{P}(B, x, M) - \underline{P}(A, x, M)] = (1 - 2x)t$$

If this bank lends to projects located in $[0, x]$ its total profit will be given by the product of x times the average per-borrower profit which in turn is the profit on the borrower located at $\frac{x}{2}$. Therefore, the profit expression that I am looking for is

$$x(1 - 2\frac{x}{2})t = x(1 - x)t$$

I can now substitute $\widetilde{X}(t)$ in for x and set the product equal to the target profit $\underline{\pi}$. Since $\widetilde{X}(t)$ is $A's$ market share when not monitoring, the expression $\widetilde{X}(t)\left[1 - \widetilde{X}(t)\right]t$ is equal to the profit that A makes when lending to these $\widetilde{X}(t)$ borrowers and also monitoring them.

$$\widetilde{X}(t)\left[1 - \widetilde{X}(t)\right]t = \underline{\pi}$$

$\widetilde{X}(t)$ had earlier been shown to be equal to $\left[\frac{tp_L - C}{t(p_H + p_L)}\right]$, $1 - \widetilde{X}(t)$ equals $\left[\frac{tp_H + C}{t(p_H + p_L)}\right]$ so the above equation can be stated as

$$\left[\frac{tp_L - C}{t(p_H + p_L)}\right]\left[\frac{tp_H + C}{t(p_H + p_L)}\right]t = \underline{\pi}$$

Cancelling the t in both denominator and numerator on the l.h.s. and multiplying both sides by $(p_H + p_L)^2 t$ yields

$$(tp_L - C)(tp_H + C) = \underline{\pi}(p_H + p_L)^2 t$$

$$\Leftrightarrow t^2 p_L p_H - Ctp_H + Ctp_L - C^2 - \underline{\pi}(p_H + p_L)^2 t = 0$$

$$\Leftrightarrow t^2 p_L p_H - C(p_H - p_L)t - \underline{\pi}(p_H + p_L)^2 t - C^2 = 0$$

$$\Leftrightarrow t^2 - \frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{p_L p_H} t - \frac{C^2}{p_L p_H} = 0$$

The two candidate solutions are thus:

$$t_{1/2} : \frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H}$$

$$\pm\sqrt{\left(\frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H}\right)^2 + \frac{C^2}{p_L p_H}\frac{4p_L p_H}{4p_L p_H}} =$$

$$= \frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H}$$

$$\pm\frac{1}{2p_L p_H}\sqrt{(C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2)^2 + 4C^2 p_L p_H}$$

Note that $(C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2)^2 = C^2(p_H - p_L)^2 + 2C(p_H - p_L)\underline{\pi}(p_H + p_L)^2 + \underline{\pi}^2\left((p_H + p_L)^2\right)^2$ and that $C^2(p_H - p_L)^2$ and $4C^2 p_L p_H$ can be combined as follows: $C^2(p_H - p_L)^2 + 4C^2 p_L p_H = C^2[(p_H)^2 - 2p_L p_H + (p_L)^2 + 4p_L p_H] = C^2(p_H + p_L)^2$. Taking $(p_H + p_L)^2$ out from under the square root allows the candidate solutions to be written as

$$t_{1/2} : \frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H}$$

$$\pm\frac{(p_H + p_L)}{2p_L p_H}\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

By assumption, t^* has to fulfil the requirement that $\frac{C}{p_L} < t^* < \underline{t}(A = U; B = M)$. I will next show that only one of the two roots satisfies these restrictions. In fact, the " $-$ " root is less than $\frac{C}{p_L}$. To see this, consider the inequality

$$\frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H}$$

$$-\frac{(p_H + p_L)}{2p_L p_H}\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2} < \frac{C}{p_L}$$

Multiplying through by $2p_L p_H$ and moving $C(p_H - p_L)$ to the r.h.s. yields

$$\underline{\pi}(p_H + p_L)^2 - (p_H + p_L)\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

$$< 2Cp_H - C(p_H - p_L) = C(p_H + p_L)$$

Dividing by $(p_H + p_L)$ and isolating the square root term on the r.h.s., I obtain

$$\underline{\pi}(p_H + p_L) - C < \sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

Squaring both sides leads to an inequality that must be correct, validating the initial claim that the " $-$ " root can be ruled out:

$$\underline{\pi}^2(p_H + p_L)^2 - 2\underline{\pi}(p_H + p_L)C + C^2 < C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2$$

To see that the " $+$ " root is bigger than $\frac{C}{p_L}$, I repeat the preceding steps (with the " $<$ " replaced by " $>$ ") for the " $+$ " root to obtain

$$C - \underline{\pi}(p_H + p_L) < \sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

Squaring both sides leads to the same inequality already shown to be correct:

$$C^2 - 2\underline{\pi}(p_H + p_L)C + \underline{\pi}^2(p_H + p_L)^2 < C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2$$

The " $+$ " root also has to be smaller than $\underline{t}(A = U; B = M)$:

$$\frac{C(p_H - p_L) + \underline{\pi}(p_H + p_L)^2}{2p_L p_H} + \frac{(p_H + p_L)}{2p_L p_H}\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

$$< \underline{t}(A = U; B = M)$$

where $\underline{t}(A = U; B = M)$ is given by

$$\underline{t}(A = U; B = M) = \frac{2p_L C + p_H D\underline{\pi}}{2(p_L)^2} + \frac{1}{2(p_L)^2}\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2}$$

Multiplying through by $2(p_L)^2 p_H$ to get rid of the fractions, this is

$$C(p_H - p_L)p_L + \underline{\pi}(p_H + p_L)^2 p_L$$

$$+(p_H + p_L)p_L\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

$$< 2p_L C p_H + (p_H)^2 D\underline{\pi} + p_H\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2}$$

I now proceed as follows: I consolidate the non-square root terms on the r.h.s of the inequality. I then replace $\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$ with a bigger expression and I replace $\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2}$ with a smaller expression. If I can show that the inequality still holds, I will

have demonstrated that the original inequality must have been correct, too. Bringing all the non-square root terms to the r.h.s., replacing D with $2(p_H + p_L)$ and arranging terms in the order numeral, C, $\underline{\pi}$, p_L, p_H, I have

$$2p_L C p_H + (p_H)^2 D\underline{\pi} - C(p_H - p_L)p_L - \underline{\pi}(p_H + p_L)^2 p_L$$

$$= 2Cp_L p_H + 2\underline{\pi}(p_H)^2 (p_H + p_L) - Cp_L p_H + C(p_L)^2 - \underline{\pi}p_L(p_H + p_L)^2$$

Cancelling terms and grouping, this is

$$Cp_L p_H + C(p_L)^2 + \underline{\pi}(p_H + p_L)[2(p_H)^2 - p_L(p_H + p_L)]$$

$$= Cp_L(p_H + p_L) + \underline{\pi}(p_H + p_L)[2(p_H)^2 - p_L p_H - (p_L)^2]$$

$$= Cp_L(p_H + p_L) + \underline{\pi}(p_H + p_L)[(p_H)^2 - (p_L)^2 + p_H(p_H - p_L)]$$

Now note that

$$\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2 + 4Cp_L\underline{\pi}}$$

$$> \sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

The former square root can be transformed as follows:

$$\sqrt{C^2 + 2C(p_H - p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2 + 4Cp_L\underline{\pi}}$$

$$= \sqrt{C^2 + 2C(p_H + p_L)\underline{\pi} + \underline{\pi}^2(p_H + p_L)^2}$$

$$= C + \underline{\pi}(p_H + p_L)$$

Multiplying this with $(p_H + p_L)p_L$, the original inequality whose validity I want to prove can be made more restrictive:

$$p_L(p_H + p_L)[C + \underline{\pi}(p_H + p_L)]$$

$$< Cp_L(p_H + p_L) + \underline{\pi}(p_H + p_L)[(p_H)^2 - (p_L)^2$$

$$+p_H(p_H - p_L)] + p_H\sqrt{4p_L C p_H D\underline{\pi} + (p_H D\underline{\pi})^2}$$

Replacing $\sqrt{4p_LCp_HD\underline{\pi}+(p_HD\underline{\pi})^2}$ with a term smaller than $\sqrt{4p_LCp_HD\underline{\pi}+(p_HD\underline{\pi})^2}$ will make the inequality even more restrictive. The term I use is $2p_H\underline{\pi}(p_H+p_L) = p_HD\underline{\pi} = \sqrt{(p_HD\underline{\pi})^2} < \sqrt{4p_LCp_HD\underline{\pi}+(p_HD\underline{\pi})^2}$. Cancelling $Cp_L(p_H+p_L)$ on both sides, the inequality then becomes

$$p_L(p_H+p_L)\underline{\pi}(p_H+p_L)$$

$$< \underline{\pi}(p_H+p_L)[(p_H)^2-(p_L)^2+p_H(p_H-p_L)]+2(p_H)^2\underline{\pi}(p_H+p_L)$$

Moving all terms to the r.h.s gives

$$0 < \underline{\pi}(p_H+p_L)[(p_H)^2-(p_L)^2+p_H(p_H-p_L)+2(p_H)^2-p_L(p_H+p_L)]$$

which when written as

$$0 < \underline{\pi}(p_H+p_L)[(p_H)^2-(p_L)^2+p_H(p_H-p_L)+(p_H)^2-(p_L)^2+p_H(p_H-p_L)]$$

must be true. This shows that t^* is equal to

$$\frac{C(p_H-p_L)+\underline{\pi}(p_H+p_L)^2}{2p_Lp_H}+\frac{(p_H+p_L)}{2p_Lp_H}\sqrt{C^2+2C(p_H-p_L)\underline{\pi}+\underline{\pi}^2(p_H+p_L)^2} \tag{5.14}$$

When t takes the value t^*, the effort-minimizer is lending to and monitoring exactly its effortless borrowers; i.e., projects located in the interval $[0,\widetilde{X}(t^*)] = \left[0,\left(\frac{t^*p_L-C}{t^*(p_H+p_L)}\right)\right]$ When t takes a value greater than t^* $(t^* < t \leq \underline{t}(A=U;B=M))$ A will monitor existing borrowers in the interval $[0,x^*(t^* < t \leq \underline{t}(A=U;B=M))]$ with $x^*(t^* < t \leq \underline{t}(A=U;B=M)) < \widetilde{X}(t^* < t \leq \underline{t}(A=U;B=M))$ (i.e., lend to some projects without monitoring them) and, conversely, when t takes a value smaller than t^* $(\frac{C}{p_L} < t < t^*)$ A monitors borrowers in the interval $[0,x^*(\frac{C}{p_L} < t < t^*)]$ with $x^*(\frac{C}{p_L} < t < t^*) > \widetilde{X}(\frac{C}{p_L} < t < t^*)$. In other words, in the latter case the additional profit from monitoring $A's$ effortless borrowers is not sufficient to attain the profit target; A also needs to steal market share from B.

5.7.5 Derivation of $x^*(t^* < t < \underline{t}(A=U;B=M))$

It will be convenient to use (5.11) and to replace $\underline{\pi}$ with $\frac{1}{p_HD}\frac{(tp_L-C)^2}{t}$ which it must be equal to by definition of $\underline{t}(A=U;B=M)$ (here abbreviated

as $\underline{t}$). The difference between $\underline{\pi}$ and $\frac{1}{p_H D}\frac{(tp_L-C)^2}{t}$ is then given by

$$\underline{\pi} - \frac{1}{p_H D}\frac{(tp_L - C)^2}{t} =$$

$$\frac{1}{p_H D}\frac{(\underline{t}p_L - C)^2}{\underline{t}} - \frac{1}{p_H D}\frac{(tp_L - C)^2}{t} =$$

$$\frac{1}{p_H D}\frac{1}{t\underline{t}}\left[t\left(\underline{t}p_L - C\right)^2 - \underline{t}\left(tp_L - C\right)^2\right] =$$

$$\frac{1}{p_H D}\frac{1}{t\underline{t}}\left[t\left(\underline{t}p_L\right)^2 - 2t\underline{t}p_L C + tC^2 - \underline{t}\left(tp_L\right)^2 + 2\underline{t}tp_L C - \underline{t}C^2\right]$$

Cancelling $2\underline{t}tp_L C$ and grouping terms, I am left with

$$\frac{1}{p_H D}\frac{1}{t\underline{t}}\left[t\underline{t}\left(p_L\right)^2\left(\underline{t} - t\right) - \left(\underline{t} - t\right)C^2\right] =$$

$$\frac{1}{p_H D}\frac{\left(\underline{t} - t\right)}{t\underline{t}}\left[t\underline{t}\left(p_L\right)^2 - C^2\right] =$$

$$\frac{1}{2p_H(p_H + p_L)}\frac{\left(\underline{t} - t\right)}{t\underline{t}}\left[t\underline{t}\left(p_L\right)^2 - C^2\right]$$

The original equation can then be written as

$$x^*\frac{1}{p_H}\left[\left(\frac{2 - x^*}{2}\right)\left(p_H - p_L\right)t + \frac{2}{2}C\right]$$

$$= \frac{1}{2p_H(p_H + p_L)}\frac{\left(\underline{t} - t\right)}{t\underline{t}}\left[t\underline{t}\left(p_L\right)^2 - C^2\right]$$

Multiplying through by $-2p_H$ gives

$$(-1)x^*\left[\left(2 - x^*\right)\left(p_H - p_L\right)t + 2C\right]$$

$$= (-1)\frac{1}{(p_H + p_L)}\frac{\left(\underline{t} - t\right)}{t\underline{t}}\left[t\underline{t}\left(p_L\right)^2 - C^2\right]$$

Rearranging the l.h.s:

$$(-1)x^*\left[\left(2 - x^*\right)\left(p_H - p_L\right)t + 2C\right] =$$

$$-x^*2(p_H - p_L)t + \left(x^*\right)^2\left(p_H - p_L\right)t - x^*2C =$$

$$\left(x^*\right)^2\left(p_H - p_L\right)t - 2\left[\left(p_H - p_L\right)t + C\right]x^*$$

Moving all terms to the l.h.s., I have

$$\left(x^*\right)^2\left(p_H - p_L\right)t - 2\left[\left(p_H - p_L\right)t + C\right]x^*$$

$$+\frac{1}{(p_H+p_L)}\frac{(\underline{t}-t)}{t\underline{t}}\left[t\underline{t}\,(p_L)^2-C^2\right]=0$$

Dividing through by $(p_H-p_L)t$ yields the quadratic equation

$$(x^*)^2-\frac{2\left[(p_H-p_L)t+C\right]}{(p_H-p_L)t}x^*$$

$$+\frac{1}{\left((p_H)^2-(p_L)^2\right)}\frac{(\underline{t}-t)}{(t)^2\,\underline{t}}\left[t\underline{t}\,(p_L)^2-C^2\right]=0$$

The candidate solutions are therefore

$$x^*_{1/2}:\frac{\left[(p_H-p_L)t+C\right]}{(p_H-p_L)t}$$

$$\pm\sqrt{\left(\frac{\left[(p_H-p_L)t+C\right]}{(p_H-p_L)t}\right)^2-\frac{1}{\left((p_H)^2-(p_L)^2\right)}\frac{(\underline{t}-t)}{(t)^2\,\underline{t}}\left[t\underline{t}\,(p_L)^2-C^2\right]}$$

Inspection shows that the " $+$ " root can be ruled out immediately as the resulting x^* would be greater than 1, whereas the solution needs to be less than $\frac{1}{2}$ (A will not lend to borrowers located further away than $\frac{1}{2}$). The result can be simplified a little by pulling $\frac{1}{(p_H-p_L)t}$ out from under the square root:

$$\frac{[(p_H-p_L)t+C]}{(p_H-p_L)t}-\sqrt{\left(\frac{[(p_H-p_L)t+C]}{(p_H-p_L)t}\right)^2-\frac{(p_H-p_L)}{(p_H-p_L)}\frac{1}{((p_H)^2-(p_L)^2)}\frac{(\underline{t}-t)}{(t)^2\underline{t}}\left[t\underline{t}\,(p_L)^2-C^2\right]}$$

$$=\frac{[(p_H-p_L)t+C]}{(p_H-p_L)t}-\frac{1}{(p_H-p_L)t}\sqrt{[(p_H-p_L)t+C]^2-\frac{(p_H-p_L)(\underline{t}-t)\left[t\underline{t}(p_L)^2-C^2\right]}{(p_H+p_L)\underline{t}}} \tag{5.15}$$

I wish to show that this expression for x^* satisfies the following three requirements:

1. $x^*(t=\underline{t})=0$
2. $x^*(t=t^*)=\widetilde{X}(t^*)=\left(\frac{t^*p_L-C}{t^*(p_H+p_L)}\right)$
3. x^* is monotonically decreasing in t and continuous (continuity is established by invoking standard results on the continuity of a function following from the continuity of its components and the observation that t by assumption is restricted to be in the interval $\frac{C}{p_L}<t<\underline{t}(A=U;B=M)$)

Property 1 can be demonstrated by inspection. Setting $t=\underline{t}$ knocks out the second term under the square root and leads x^* to vanish.

Property 2 is more complicated. It involves showing that

$$\frac{[(p_H-p_L)t^*+C]}{(p_H-p_L)t^*}-\frac{1}{(p_H-p_L)t^*}\sqrt{[(p_H-p_L)t^*+C]^2-\frac{(p_H-p_L)(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{(p_H+p_L)\underline{t}}}$$

$$= \left(\frac{t^*p_L - C}{t^*(p_H+p_L)}\right)$$

I first simplify the above equation as much as possible. The preceding expression is equivalent to

$$1 + \frac{C}{(p_H-p_L)t^*} - \frac{1}{(p_H-p_L)t^*}\sqrt{[(p_H-p_L)t^*+C]^2 - \frac{(p_H-p_L)}{(p_H+p_L)}\frac{(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{\underline{t}}}$$

$$- \left(\frac{p_L}{(p_H+p_L)} - \frac{C}{t^*(p_H+p_L)}\right) = 0$$

which in turn can be written as

$$\frac{p_H}{(p_H+p_L)} + \frac{C}{(p_H-p_L)t^*} + \frac{C}{t^*(p_H+p_L)}$$

$$-\frac{1}{(p_H-p_L)t^*}\sqrt{[(p_H-p_L)t^*+C]^2 - \frac{(p_H-p_L)}{(p_H+p_L)}\frac{(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{\underline{t}}} = 0$$

Multiplying through by t^* yields

$$\frac{p_H t^*}{(p_H+p_L)} + \frac{C}{(p_H-p_L)} + \frac{C}{(p_H+p_L)}$$

$$-\frac{1}{(p_H-p_L)}\sqrt{[(p_H-p_L)t^*+C]^2 - \frac{(p_H-p_L)}{(p_H+p_L)}\frac{(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{\underline{t}}} = 0$$

Note that $\frac{C}{(p_H-p_L)} + \frac{C}{(p_H+p_L)} = \frac{C(p_H+p_L)+C(p_H-p_L)}{(p_H-p_L)(p_H+p_L)} = \frac{2Cp_H}{(p_H-p_L)(p_H+p_L)}$. Multiplying the equation by $(p_H - p_L)$ and moving the square root term to the r.h.s. I obtain

$$\frac{(p_H-p_L)p_H t^*}{(p_H+p_L)} + \frac{2Cp_H}{(p_H+p_L)} = \sqrt{[(p_H-p_L)t^*+C]^2 - \frac{(p_H-p_L)}{(p_H+p_L)}\frac{(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{\underline{t}}}$$

Taking the factor $\frac{p_H}{(p_H+p_L)}$ out of both expressions on the l.h.s. allows me to write $\frac{(p_H-p_L)p_H t^*}{(p_H+p_L)} + \frac{2Cp_H}{(p_H+p_L)}$ as $\frac{p_H}{(p_H+p_L)}[(p_H - p_L)t^* + 2C]$, so squaring both sides gives

$$\left(\frac{p_H}{(p_H+p_L)}\right)^2[(p_H-p_L)t^*+2C]^2 = [(p_H-p_L)t^*+C]^2 - \frac{(p_H-p_L)}{(p_H+p_L)}\frac{(\underline{t}-t^*)\left[t^*\underline{t}(p_L)^2-C^2\right]}{\underline{t}}$$

While it looks as if there might be scope to simplify this expression further, proving the validity of the preceding claim at some point involves substituting the expressions obtained for t^* and $\underline{t}$ into the equation and that is computationally involved indeed. Instead of doing this manually, I have employed Maple, the maths engine that is part of the Scientific Workplace software package, to ascertain the veracity of the claim by using the {check equality} command. The statement is indeed true[10].

[10]Using Maple requires certain changes to the notation that are not documented in the manual / online help. Maple will only work with expressions that contain exclusively standard parentheses, i.e., (and), and also will not accept underlined or

As for property 3, there are two routes to showing that x^* is monotonically decreasing in t. The first involves simple differentiation of x^*. The second route proceeds by noting that the equation

$$x^* \frac{1}{p_H}\left[\left(1-\frac{x^*}{2}\right)(p_H-p_L)t+C\right]+\frac{1}{p_H D}\frac{(tp_L-C)^2}{t}=\underline{\pi}$$

implicitly defines x^* as a function of t and that the requirements for applying the implicit function theorem to compute $\frac{\partial x^*}{\partial t}$ are fulfilled. I then use Maple to establish the equivalence of the solutions obtained following these two routes. I begin by restating

$$x^* = \frac{[(p_H-p_L)t+C]}{(p_H-p_L)t} - \frac{1}{(p_H-p_L)t}\sqrt{[(p_H-p_L)t+C]^2 - \frac{(p_H-p_L)(\underline{t}-t)\left[t\underline{t}(p_L)^2-C^2\right]}{(p_H+p_L)\underline{t}}}$$

and rewriting this as

$$x^* = 1+\frac{C}{(p_H-p_L)t}-\frac{1}{(p_H-p_L)t}\sqrt{Y}$$

$$= 1+\frac{1}{(p_H-p_L)t}[C-\sqrt{Y}]$$

where

$$Y=[(p_H-p_L)t+C]^2-\frac{(p_H-p_L)(\underline{t}-t)\left[t\underline{t}(p_L)^2-C^2\right]}{(p_H+p_L)\underline{t}}$$

Differentiating x^* with respect to t gives

$$\frac{\partial x^*}{\partial t}=\frac{\partial}{\partial t}\left[\frac{1}{(p_H-p_L)t}\right]*[C-\sqrt{Y}]+\frac{1}{(p_H-p_L)t}\frac{\partial}{\partial t}[C-\sqrt{Y}]$$

where $\frac{\partial}{\partial t}[C-\sqrt{Y}]$ is given by

$$\frac{\partial}{\partial t}[C-\sqrt{Y}]=(-1)\frac{\partial}{\partial t}\sqrt{Y}=(-1)\frac{1}{2}\frac{1}{\sqrt{Y}}\frac{\partial Y}{\partial t}$$

Note that

$$\frac{\partial}{\partial t}\left[\frac{1}{(p_H-p_L)t}\right]=\frac{-(p_H-p_L)}{(p_H-p_L)^2t^2}=\frac{(-1)}{(p_H-p_L)t^2}$$

otherwise decorated characters, so for example $\underline{\pi}$ is replaced by P. I have collected all computations in a SWP file, maple-computations.tex, which I would be pleased to make available for checking upon request.

In differentiating Y with respect to t, I rewrite Y as

$$Y = [(p_H - p_L)t + C]^2 - \frac{(p_H - p_L)}{(p_H + p_L)\underline{t}}(\underline{t} - t)\left[t\underline{t}\,(p_L)^2 - C^2\right]$$

$$\frac{\partial Y}{\partial t} = 2\left[(p_H - p_L)t + C\right](p_H - p_L)$$

$$-\frac{(p_H - p_L)}{(p_H + p_L)\underline{t}}\left[\frac{\partial}{\partial t}(\underline{t} - t) * \left[t\underline{t}\,(p_L)^2 - C^2\right] + (\underline{t} - t)\frac{\partial}{\partial t}\left[t\underline{t}\,(p_L)^2 - C^2\right]\right]$$

where

$$\left[\frac{\partial}{\partial t}(\underline{t} - t) * \left[t\underline{t}\,(p_L)^2 - C^2\right] + (\underline{t} - t)\frac{\partial}{\partial t}\left[t\underline{t}\,(p_L)^2 - C^2\right]\right]$$

$$= \left[(-1)\left[t\underline{t}\,(p_L)^2 - C^2\right] + (\underline{t} - t)\underline{t}\,(p_L)^2\right]$$

$$= \left[-t\underline{t}\,(p_L)^2 + C^2 + (\underline{t})^2\,(p_L)^2 - t\underline{t}\,(p_L)^2)\right]$$

$$= \left[C^2 + (\underline{t})^2\,(p_L)^2 - 2t\underline{t}\,(p_L)^2)\right]$$

Thus

$$\frac{\partial Y}{\partial t} = 2\left[(p_H - p_L)t + C\right](p_H - p_L) - \frac{(p_H - p_L)}{(p_H + p_L)\underline{t}}\left[C^2 + (\underline{t})^2\,(p_L)^2 - 2t\underline{t}\,(p_L)^2)\right]$$

This can be simplified further:

$$\frac{\partial Y}{\partial t} = \frac{1}{(p_H + p_L)\underline{t}}\{2(p_H - p_L)t(p_H - p_L)(p_H + p_L)\underline{t}$$

$$+2C(p_H - p_L)(p_H + p_L)\underline{t} - (p_H - p_L)C^2$$

$$-(p_H - p_L)\,(\underline{t})^2\,(p_L)^2 + (p_H - p_L)2t\underline{t}\,(p_L)^2\}$$

$$= \frac{1}{(p_H + p_L)\underline{t}}\{2t\underline{t}(p_H - p_L)\left[(p_H)^2 - (p_L)^2\right] + 2C\underline{t}(p_H + p_L)(p_H - p_L)$$

$$-C^2(p_H - p_L) - (\underline{t})^2\,(p_L)^2\,(p_H - p_L) + 2t\underline{t}\,(p_L)^2\,(p_H - p_L)\}$$

Cancelling $2t\underline{t}\,(p_L)^2\,(p_H - p_L)$ and pulling $(p_H - p_L)$ out of the expression, I am left with

$$\frac{\partial Y}{\partial t} = \frac{(p_H - p_L)}{(p_H + p_L)\underline{t}}\left(2t\underline{t}\,(p_H)^2 + 2C\underline{t}(p_H + p_L) - C^2 - (\underline{t})^2\,(p_L)^2\right)$$

Cancelling $(p_H - p_L)$ and substituting in the other elements, I obtain

$$\frac{\partial x^*}{\partial t} = \frac{(-1)}{(p_H - p_L)t^2} * [C - \sqrt{Y}]$$

$$-\frac{1}{2t}\frac{1}{\sqrt{Y}}\frac{1}{(p_H+p_L)\underline{t}}\left(2t\underline{t}\,(p_H)^2+2C\underline{t}(p_H+p_L)-C^2-(\underline{t})^2\,(p_L)^2\right)$$

Clearly, it is not straightforward to sign this expression. Following the second route and applying the implicit function theorem yields a far more manageable result. Call the l.h.s. of (5.11) $F(...)$:

$$\underbrace{x^*\frac{1}{p_H}\left[\left(1-\frac{x^*}{2}\right)(p_H-p_L)t+C\right]+\frac{1}{p_HD}\frac{(tp_L-C)^2}{t}}_{F(...)}=\underline{\pi}$$

Then $\frac{\partial x^*}{\partial t}=(-1)\frac{\frac{\partial F(...)}{\partial t}}{\frac{\partial F(...)}{\partial x^*}}$. Note that $D=2(p_H+p_L)$; $\frac{(tp_L-C)^2}{t}=\frac{(tp_L)^2}{t}-\frac{2tp_LC}{t}+\frac{C^2}{t}=t\,(p_L)^2-2p_LC+\frac{C^2}{t}$ and that $\frac{\partial}{\partial t}\frac{C^2}{t}=\frac{-C^2}{t^2}$.

$$\frac{\partial F(...)}{\partial t}=$$

$$=x^*\frac{1}{p_H}\left(\frac{2-x^*}{2}\right)(p_H-p_L)+\frac{1}{2p_H(p_H+p_L)}\left[(p_L)^2-\frac{C^2}{t^2}\right]$$

$$=\frac{1}{2p_H}\left(x^*\,(2-x^*)\,(p_H-p_L)+\frac{1}{(p_H+p_L)}\left[\frac{t^2\,(p_L)^2-C^2}{t^2}\right]\right)$$

$$\frac{\partial F(...)}{\partial x^*}=$$

$$=\frac{1}{p_H}\left[\left(1-\frac{x^*}{2}\right)(p_H-p_L)t+C\right]+x^*\frac{1}{p_H}\left(\frac{-1}{2}\right)(p_H-p_L)t$$

$$=\frac{1}{2p_H}\left([(2-x^*)\,(p_H-p_L)t+2C]-x^*(p_H-p_L)t\right)$$

$$=\frac{1}{2p_H}2\left[(1-x^*)\,(p_H-p_L)t+C\right]$$

Putting these results together and cancelling $\frac{1}{2p_H}$ in the process:

$$\frac{\partial x^*}{\partial t}=(-1)\frac{x^*\,(2-x^*)\,(p_H-p_L)+\frac{1}{(p_H+p_L)}\left[\frac{t^2\,(p_L)^2-C^2}{t^2}\right]}{2\left[(1-x^*)\,(p_H-p_L)t+C\right]}$$

Leaving aside the (-1) preceding the fraction and recalling that by assumption $t>\frac{C}{p_L}$, the numerator of the expression must be positive. Then the overall sign of $\frac{\partial x^*}{\partial t}$ will be negative if and only if the denominator is positive. The condition for this to be the case is

$$2\left[(1-x^*)\,(p_H-p_L)t+C\right]>0$$

$$\Leftrightarrow (p_H - p_L)t + C > x^*(p_H - p_L)t$$

$$\Leftrightarrow \frac{(p_H - p_L)t + C}{(p_H - p_L)t} > x^*$$

which must be true since it has been shown that A would not monitor borrowers located beyond $x = \frac{1}{2}$. This result confirms that x^* is decreasing in t. Using Maple, I have confirmed that the results obtained via the two different routes are equivalent[11].

5.7.6 $x^*\left(\frac{C}{p_L} < t \leq t^*\right) = \widetilde{X}(t^*)$ for $t = t^*$

To show that $x^*\left(\frac{C}{p_L} < t \leq t^*\right) = \widetilde{X}(t^*)$ for $t = t^*$, I begin by writing the term under the square root, $\frac{1}{4} - \frac{\pi}{t}$, as follows:

$$\frac{1}{4} - \frac{\pi}{t} = \frac{1}{4} - \frac{1}{t}\frac{1}{p_H D}\frac{(tp_L - C)^2}{t}$$

With D replaced by $2(p_H + p_L)$ and both fractions converted to the same denominator, this can be written as

$$\frac{1}{4} - \frac{\pi}{t} = \frac{1}{4}\frac{tt p_H (p_H + p_L)}{tt p_H (p_H + p_L)} - \frac{1}{t}\frac{2}{2}\frac{1}{p_H 2(p_H + p_L)}\frac{(tp_L - C)^2}{t} =$$

$$= \frac{1}{4}\frac{1}{tt p_H (p_H + p_L)}\left(tt p_H (p_H + p_L) - 2(tp_L - C)^2\right)$$

Noting that $(tp_L - C)^2 = (t)^2 (p_L)^2 - 2tp_L C + C^2$, the expression is

$$= \frac{1}{4}\frac{1}{tt p_H (p_H + p_L)}\left(tt p_H (p_H + p_L) - 2(t)^2 (p_L)^2 + 4tp_L C - 2C^2\right)$$

Pulling the $\frac{1}{4}$ out from under the square root, x^* is given by

$$x^* = \frac{1}{2} - \frac{1}{2}\sqrt{\underbrace{\frac{1}{tt p_H (p_H + p_L)}\left(tt p_H (p_H + p_L) - 2(t)^2 (p_L)^2 + 4tp_L C - 2C^2\right)}_{Y}}$$

The equation that is to be shown to hold is

$$\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\pi}{t^*}} = \widetilde{X}(t^*) = \left[\frac{t^* p_L - C}{t^*(p_H + p_L)}\right]$$

[11] As noted above, I have recorded all computations step-by-step in a SWP file, maple-computations.tex, which I would be pleased to make available for checking upon request.

This can be rearranged as

$$\frac{1}{2} - \left[\frac{t^* p_L - C}{t^*(p_H + p_L)}\right] = \sqrt{\frac{1}{4} - \frac{\pi}{t^*}}$$

Multiplying both sides by 2 and moving the terms on the l.h.s. onto the same denominator

$$1 - \frac{2\left(t^* p_L - C\right)}{t^*(p_H + p_L)} = \frac{t^*(p_H + p_L) - 2\left(t^* p_L - C\right)}{t^*(p_H + p_L)}$$

$$\Leftrightarrow \frac{t^*(p_H - p_L) + 2C}{t^*(p_H + p_L)} \stackrel{?}{=} 2\sqrt{\frac{1}{4} - \frac{\pi}{t^*}} = \sqrt{Y}$$

Squaring both sides

$$\left(\frac{1}{t^*(p_H + p_L)}\right)^2 \left[t^*(p_H - p_L) + 2C\right]^2 =$$

$$= \frac{1}{t^* \underline{t} p_H \left(p_H + p_L\right)} \left(t^* \underline{t} p_H \left(p_H + p_L\right) - 2\left(\underline{t}\right)^2 \left(p_L\right)^2 + 4\underline{t} p_L C - 2C^2\right)$$

Canceling $t^*(p_H + p_L)$ on both sides

$$\frac{1}{t^*(p_H + p_L)} \left[t^*(p_H - p_L) + 2C\right]^2 =$$

$$= \frac{1}{\underline{t} p_H} \left(t^* \underline{t} p_H \left(p_H + p_L\right) - 2\left(\underline{t}\right)^2 \left(p_L\right)^2 + 4\underline{t} p_L C - 2C^2\right)$$

Noting that $\left[t^*(p_H - p_L) + 2C\right]^2 = \left(t^*\right)^2 \left(p_H - p_L\right)^2 + 4t^*(p_H - p_L)C + 4C^2$ and getting rid of the fractions

$$\left(\left(t^*\right)^2 \left(p_H - p_L\right)^2 + 4t^*(p_H - p_L)C + 4C^2\right) \underline{t} p_H =$$

$$= \left(t^* \underline{t} p_H \left(p_H + p_L\right) - 2\left(\underline{t}\right)^2 \left(p_L\right)^2 + 4\underline{t} p_L C - 2C^2\right) t^*(p_H + p_L)$$

Multiplying out gives $\left(t^*\right)^2 \left(p_H - p_L\right)^2 \underline{t} p_H + 4t^*(p_H - p_L) C \underline{t} p_H + 4C^2 \underline{t} p_H =$ $= t^* \underline{t} p_H \left(p_H + p_L\right) t^*(p_H + p_L) - 2\left(\underline{t}\right)^2 \left(p_L\right)^2 t^*(p_H + p_L) + 4\underline{t} p_L C t^*(p_H + p_L) - 2C^2 t^*(p_H + p_L)$ which can be rearranged in the order numeral, C, t^*, $\underline{t}$, p_L, p_H, $(p_H \pm p_L)$: $\left(t^*\right)^2 \underline{t} p_H \left(p_H - p_L\right)^2 + 4Ct^* \underline{t} p_H (p_H - p_L) + 4C^2 \underline{t} p_H = \left(t^*\right)^2 \underline{t} p_H \left(p_H + p_L\right)^2 - 2t^* \left(\underline{t}\right)^2 \left(p_L\right)^2 (p_H + p_L) + 4Ct^* \underline{t} p_L (p_H + p_L) - 2C^2 t^*(p_H + p_L)$. Moving everything to the l.h.s. and combining terms: $\left(t^*\right)^2 \underline{t} p_H \left[\left(p_H\right)^2 - 2p_H p_L + \left(p_L\right)^2 - \left(p_H\right)^2 - 2p_H p_L - \left(p_L\right)^2\right] + 4Ct^* \underline{t} \left[\left(p_H\right)^2 - p_H p_L - p_H p_L - \left(p_L\right)^2\right] + 4C^2 \underline{t} p_H + 2t^* \left(\underline{t}\right)^2 \left(p_L\right)^2 \left(p_H + p_L\right) + 2C^2 t^*(p_H + p_L)$. Cancelling terms, the equality can be rewritten as

$$-4\left(t^*\right)^2 \underline{t} p_L \left(p_H\right)^2 + 4Ct^* \underline{t} \left(\left(p_H\right)^2 - 2p_H p_L - \left(p_L\right)^2\right)$$

$$+4C^2\underline{t}p_H + 2t^*\left(\underline{t}\right)^2\left(p_L\right)^2\left(p_H + p_L\right) + 2C^2t^*(p_H + p_L) = 0$$

This is as much as I have been able to simplify the original equality manually without substituting in for t^* and $\underline{t}$. I have confirmed that the equality holds using the {check equality} command in Maple. As noted above, the SWP file with the calculations is available upon request.

5.7.7 Surplus for $4\underline{\pi} \leq t < \underline{t}(A = U; B = M)$ decreasing in t

I consider the two cases in turn, $t^* < t < \underline{t}(A = U; B = M)$ first. Surplus (expected surplus) before processing costs on a monitored borrower is given by $p_H R - \rho - m$, (expected) surplus before processing costs on an unmonitored borrower is $p_L R - \rho$. All projects obtain a loan, but not all of them are monitored: Projects located in $[0, x^*]$ borrow from A and are monitored; projects located in $(x^*, \widetilde{X}(t)]$ borrow from A, but are not monitored, and projects located at $x > \widetilde{X}(t)$ borrow from B and are monitored again. As above, I will consider surplus before processing costs and processing costs separately. A total of $x^* + (1 - \widetilde{X}(t))$ borrowers are monitored, the remaining $\widetilde{X}(t) - x^*$ are not monitored. Thus total surplus before processing costs is given by

$$\underbrace{\left(x^* + \left(1 - \widetilde{X}(t)\right)\right)\left(p_H R - \rho - m\right) + \left(\widetilde{X}(t) - x^*\right)\left(p_L R - \rho\right)}_{\text{Surplus before processing costs}}$$

$$= \left(1 - \left(\widetilde{X}(t) - x^*\right)\right)\left(p_H R - \rho - m\right) + \left(\widetilde{X}(t) - x^*\right)\left(p_L R - \rho\right)$$

$$= \left(p_H R - \rho - m\right) - \left(\widetilde{X}(t) - x^*\right)\left[\left(p_H R - \rho - m\right) - \left(p_L R - \rho\right)\right]$$

$$= \left(p_H R - \rho - m\right) - \left(\widetilde{X}(t) - x^*\right)\left[\left(p_H - p_L\right) R - m\right]$$

Clearly, the crucial term in this is $\left(\widetilde{X}(t) - x^*\right)$ which captures the number of borrowers that are not monitored. If this can be shown to be increasing in t, surplus before processing costs must be decreasing in t. [This formulation takes into account the fact that $\left(\widetilde{X}(t) - x^*\right)$ enters surplus with a negative sign.] I have shown $\widetilde{X}(t)$ to be increasing in t and I have shown x^*, which has a negative sign, to be decreasing in t. So surplus before processing costs is decreasing in t.

As for processing costs, I state these as the product of the number of borrowers and the average processing cost. In the case of A, the number of borrowers is $\widetilde{X}(t)$ and the average processing cost is thus $\frac{\widetilde{X}(t)}{2}t$. Then $A's$ total processing costs are $\left(\widetilde{X}(t)\right)^2 \frac{t}{2}$. In the case of B, the number

of borrowers is $\left(1-\widetilde{X}(t)\right)$ and the average processing cost is therefore $\frac{1-\widetilde{X}(t)}{2}t$, leading to total processing costs for B of $\frac{t}{2}\left(1-\widetilde{X}(t)\right)^2$. In sum

$$\frac{t}{2}\left(\widetilde{X}(t)\right)^2+\frac{t}{2}\left(1-\widetilde{X}(t)\right)^2=$$

$$=\underbrace{\frac{t}{2}\left(1-2\widetilde{X}(t)+2\left(\widetilde{X}(t)\right)^2\right)}_{\text{processing costs}}$$

All expressions relating to processing costs are exactly unchanged from the scenario where A did not need to monitor at all; results and interpretations carry over as well: Processing costs are increasing in t. The direct effect of higher (lower) t assuming an unchanged division of the market $\widetilde{X}(t)$ is to increase (decrease) processing costs; this direct effect dominates the indirect effect that lowers (increases) processing costs, because an increase (decrease) in t leads to an increase (decrease) in $\widetilde{X}(t)$ - more (fewer) projects borrowing from A at lower processing cost than if they borrowed from B. Since processing costs enter surplus with a negative sign, an increase in t increases processing costs and lowers surplus.

I have now shown for $t^* \leq t \leq \underline{t}(A=U;B=M)$ what I set out to show: Surplus (the sum of two components) is unambiguously decreasing in t. The second component of surplus, processing costs, is increasing in t for the same reasons as in regime 1. However, surplus before processing costs (the first component) is now affected by an additional factor: An increase in t also leads to a reduction in x^* (i.e., A monitoring fewer borrowers). Since $\widetilde{X}(t)$ enters surplus with a negative sign, both of these effects link increased values of t with lower surplus. I could try to consolidate all four effects formally as I did above, but in view of the unwieldiness of the expressions involved (notably that for x^*) this is likely to be very tedious and unlikely to provide any additional insights. Note that surplus is also a continuous function of t.

I next consider the case of $4\underline{\pi} \leq t \leq t^*$ (regime 3). For $4\underline{\pi} \leq t \leq t^*$, A lends to and monitors borrowers located up to x^* (where the definition of x^* is different from the one for regime 2) with B winning the remainder of the market $(1-x^*)$. All projects are being monitored, so is surplus invariant to changes in t? Not quite. Surplus increases as t falls all the way down to $t=4\underline{\pi}$. The source of this surplus gain is a reduction in processing costs. As t falls, x^* increases. More projects borrow from A at lower processing costs than if they borrowed from B. So surplus is decreasing in t even though surplus before processing costs is constant for $4\underline{\pi} \leq t \leq \frac{C}{p_L}$.

Formally, A and B monitor all borrowers in their loan portfolios of size x^* and $(1-x^*)$, respectively. Then surplus before processing costs is

$$x^*\left(p_H R-\rho-m\right)+\left(1-x^*\right)\left(p_H R-\rho-m\right)$$
$$=\left(p_H R-\rho-m\right)$$

which does not vary with t. As for processing costs, in the case of A the number of borrowers is x^* and the average processing cost is thus $\frac{x^*}{2}t$. Then $A's$ total processing costs are $\left(x^*\right)^2\frac{t}{2}$. In the case of B, the number of borrowers is $(1-\ x^*)$ and the average processing cost is $\frac{1-x^*}{2}t$, leading to total processing costs for B of $\frac{t}{2}\left(1-x^*\right)^2$. Formally, the sum of these terms is

$$\frac{t}{2}\left(x^*\right)^2+\frac{t}{2}\left(1-x^*\right)^2=$$
$$=\underbrace{\frac{t}{2}\left(1-2x^*+2\left(x^*\right)^2\right)}_{\text{processing costs}}$$

Differentiating this with respect to t gives

$$\frac{1}{2}\left(1-2x^*+2\left(x^*\right)^2\right)+\frac{t}{2}\left(-2\frac{\partial x^*}{\partial t}+4\left(x^*\right)\frac{\partial x^*}{\partial t}\right)=$$

$$=\underbrace{\frac{1}{2}-x^*+\left(x^*\right)^2}_{\text{change in costs for given } x^*}+\underbrace{t\left(2x^*\frac{\partial x^*}{\partial t}-\frac{\partial x^*}{\partial t}\right)}_{\text{change in costs for given } t}$$

Once again, it makes sense to consider the direct and indirect effects of t separately. The direct effect relates to the effect of t on processing costs keeping x^* fixed. One would expect the sign of the direct effect in the above equation to be positive. Note that

$$\frac{1}{2}-x^*+\left(x^*\right)^2=\frac{1}{2}-x^*(1-x^*)$$

The relevant x^* is given by

$$x^*=\frac{1}{2}-\sqrt{\frac{1}{4}-\frac{\pi}{t}}$$

Therefore

$$\frac{1}{2}-x^*(1-x^*)=$$
$$\frac{1}{2}-\left(\frac{1}{2}-\sqrt{\frac{1}{4}-\frac{\pi}{t}}\right)\left(1-\frac{1}{2}+\sqrt{\frac{1}{4}-\frac{\pi}{t}}\right)=$$

$$\frac{1}{2} - \left(\left(\frac{1}{2}\right)^2 - \left(\sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}\right)^2 \right) =$$

$$\frac{1}{2} - \left(\frac{1}{4} - \left(\frac{1}{4} - \frac{\underline{\pi}}{t} \right) \right) =$$

$$\frac{1}{2} - \frac{1}{4} + \left(\frac{1}{4} - \frac{\underline{\pi}}{t} \right) = \frac{1}{2} - \frac{\underline{\pi}}{t}$$

So if I can show that $\left(\frac{1}{2} - \frac{\underline{\pi}}{t}\right) > 0$, I will have shown that the direct effect of t on processing costs and therefore surplus is the intuitive one: Higher t means higher costs and lower surplus.

$$\left(\frac{1}{2} - \frac{\underline{\pi}}{t} \right) > 0 \Leftrightarrow$$

$$\frac{1}{2} > \frac{\underline{\pi}}{t} \Leftrightarrow$$

$$t > 2\underline{\pi}$$

But since t by assumption is greater than $4\underline{\pi}$, the preceding inequality must hold.

The indirect effect of t on processing costs relates to the change in costs via the change in x^* induced by changes in t. The relevant component of costs is given by

$$+t \left(2x^* \frac{\partial x^*}{\partial t} - \frac{\partial x^*}{\partial t} \right)$$

$$= t \left(2x^* - 1\right) \frac{\partial x^*}{\partial t}$$

$$= (-1)t \left(1 - 2x^*\right) \frac{\partial x^*}{\partial t}$$

I had shown $\dfrac{\partial x^*}{\partial t}$ to be given by

$$\frac{\partial x^*}{\partial t} = \frac{\partial}{\partial t} \left(\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}} \right) = (-1)\frac{1}{2} \frac{1}{\sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}}} \frac{\underline{\pi}}{t^2}$$

Note that $\sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t}} = \frac{1}{2} - x^* = \frac{1 - 2x^*}{2}$. This means that $\dfrac{\partial x^*}{\partial t}$ can be written as

$$\frac{\partial x^*}{\partial t} = (-1)\frac{1}{2} \frac{1}{\frac{1 - 2x^*}{2}} \frac{\underline{\pi}}{t^2} = (-1) \frac{1}{(1 - 2x^*)} \frac{\underline{\pi}}{t^2}$$

Substituting this above yields

$$(-1)t\,(1-2x^*)\,\frac{\partial x^*}{\partial t} =$$

$$= (-1)t\,(1-2x^*)\,(-1)\frac{1}{(1-2x^*)}\frac{\pi}{t^2}$$

$$= \frac{\pi}{t}$$

This term has a positive sign in the sum capturing processing costs. Therefore in the present case the indirect effect of t (in contrast to regime 2) also links increases in t with increases in processing costs (and, therefore, decreases in surplus). What are the economics behind this? When $t > t^*$, an increase in t leads to an increase in $A's$ effortless market share $\widetilde{X}(t)$. That is bad for surplus because A does not monitor the most distant borrowers in that scenario (even though a higher market share for A means lower processing costs). When $t < t^*$, the effect of an increase in t is to *decrease* x^* because it allows A to monitor (and therefore lend to) *fewer* projects which means that B takes a greater market share (but has higher processing costs).

It is now but a formality to add up and consolidate the direct and indirect effects. Both effects are going in the same direction, so it is not surprising that the overall effect of higher t on processing costs (surplus) is positive (negative):

$$\frac{\partial\ (\text{processing costs})}{\partial t} =$$

$$= \underbrace{\frac{1}{2} - x^* + (x^*)^2}_{\text{change in costs for given } x^*} \quad \underbrace{+t\left(2x^*\frac{\partial x^*}{\partial t} - \frac{\partial x^*}{\partial t}\right)}_{\text{change in costs for given } t}$$

$$= \left(\frac{1}{2} - \frac{\pi}{t}\right) + \frac{\pi}{t} = \frac{1}{2} > 0$$

In summary, in the case where $4\underline{\pi} \leq t \leq t^*$ surplus is also decreasing in t, even though all projects are monitored. Interestingly, this property is not entirely assumed into the model. The direct effect of t on processing costs certainly is an artefact of the model setup in which the intensity of competition is captured by the difference in costs. However, the indirect effect (via the impact of t on $A's$ market share x^*) runs in the same direction as the direct effect, so that even if one ignored the direct effect, surplus would still be decreasing in t. Note also that surplus is a continuous function of t.

Surplus is given by different expressions in the three regimes $t \geq$

$\underline{t}(A = U; B = M)$, $t^* < t < \underline{t}(A = U; B = M)$, and $4\underline{\pi} \leq t \leq t^*$. What remains to be done is to show that surplus (and not just $x^*/\widetilde{X}(t)$) is continuous at the threshold values for t that separate the three regimes, namely $t = \underline{t}(A = U; B = M)$ and $t = t^*$. Formally, I need to show that

$$\lim_{t \to \underline{t}} S(t) = S(\underline{t})$$

$$\lim_{t \to t^*} S(t) = S(t^*)$$

This boils down to demonstrating that two equalities set out below hold. These state that surplus at the threshold values of t is the same regardless of whether one uses the surplus expression relating to the interval above or below the threshold value. The demonstration as presented here avoids formalities and examines whether surplus is smoothly decreasing in t everywhere. I have shown this for almost all values of t for which surplus is defined, namely $t \geq 4\underline{\pi}$. All that remains to be shown is that surplus is continuous at the threshold values.

I first consider is $t = \underline{t}(A = U; B = M)$. The l.h.s. of the equation uses the surplus expression as defined for $t^* < t < \underline{t}(A = U; B = M)$; the r.h.s. uses the surplus expression relating to $t \geq \underline{t}(A = U; B = M)$:

$$(p_H R - \rho - m)$$

$$- \left(\widetilde{X}(\underline{t}) - x^*(\underline{t})\right) \left[(p_H - p_L) R - m\right]$$

$$-\frac{\underline{t}}{2}\left(1 - 2\widetilde{X}(\underline{t}) + 2\left(\widetilde{X}(\underline{t})\right)^2\right)$$

$$\stackrel{?}{=} (p_H R - \rho - m)$$

$$-\widetilde{X}(\underline{t}) \left[(p_H - p_L) R - m\right]$$

$$-\frac{\underline{t}}{2}\left[1 - 2\widetilde{X}(\underline{t}) + 2\left(\widetilde{X}(\underline{t})\right)^2\right]$$

Inspection shows that $x^*(\underline{t}) = 0$ is a necessary and sufficient condition for the equality to hold. But that is precisely the property that I had earlier demonstrated. So surplus is indeed continuous at $t = \underline{t}(A = U; B = M)$.

The other value I consider is $t = t^*$. The r.h.s. of the equation uses the surplus expression as defined for $t^* < t < \underline{t}(A = U; B = M)$; the

l.h.s. uses the surplus expression relating to $4\underline{\pi} \leq t \leq t^*$. Note that the x^* that relate to these two regimes are different, although I initially abbreviate both of them with the same symbol x^* to save on notation:

$$(p_H R - \rho - m)$$

$$-\frac{t^*}{2}\left(1 - 2x^*(t^*) + 2\left(x^*(t^*)\right)^2\right)$$

$$\stackrel{?}{=} (p_H R - \rho - m)$$

$$-\left(\widetilde{X}(t^*) - x^*(t^*)\right)\left[(p_H - p_L)R - m\right]$$

$$-\frac{t^*}{2}\left(1 - 2\widetilde{X}(t^*) + 2\left(\widetilde{X}(t^*)\right)^2\right)$$

I can rely on earlier work to show that this equality holds. First, note that cancelling $(p_H R - \rho - m)$ on both sides and multiplying through by (-1) yields the following, slightly simpler expression:

$$\frac{t^*}{2}\left(1 - 2x^*(t^*) + 2\left(x^*(t^*)\right)^2\right)$$

$$= \left(\widetilde{X}(t^*) - x^*(t^*)\right)\left[(p_H - p_L)R - m\right]$$

$$+\frac{t^*}{2}\left(1 - 2\widetilde{X}(t^*) + 2\left(\widetilde{X}(t^*)\right)^2\right)$$

I had earlier demonstrated that $\widetilde{X}(t^*) = x^*(t^*)$. This knocks out the first term on the r.h.s. to yield

$$\frac{t^*}{2}\left(1 - 2x^*(t^*) + 2\left(x^*(t^*)\right)^2\right)$$

$$= \frac{t^*}{2}\left(1 - 2\widetilde{X}(t^*) + 2\left(\widetilde{X}(t^*)\right)^2\right)$$

Clearly, the equality holds if and only if $x^*(t^*)$, which here refers to

$$x^*(t^*) = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\underline{\pi}}{t^*}}$$

is equal to $\widetilde{X}(t^*)$. That is what I have shown in the preceding sub-sub-section. Therefore, surplus is continuous at $t = t^*$.

Chapter 6

Lending-related incentive schemes as a tool of bank regulation

6.1 Overview and Summary

I begin this chapter with an introduction (6.2) that motivates the research reported here. In particular, I defend the non-standard methodology employed, namely the use of interviews to address the question 'What scope is there for devising effective lending-related incentive arrangements in practice? Is it possible to construct incentive schemes that align the interests of a regulator with those of the banks?' I approached this subject by interviewing approximately 120 practitioners in a sample of German banks. Reporting the results of qualitative interviewing is an inherently 'wordy' affair and I only have space to discuss the main issues and findings in the body of the thesis. A separate note with supplementary material [Eggenberger (2006c)] is available upon request. In that note I discuss how the regulation of lending-related incentive schemes fits in with the current framework of bank regulation. I provide a summary in section 6.3 below. In the following section (6.4), I set out the case for regulating incentive compensation in banks [also in the form of an extended summary of a more detailed discussion in Eggenberger (2006c)]. Three issues are examined: I provide evidence that lending-related incentive arrangements should be an area of concern for bank regulation and I document that this view is shared by a number of prominent academics. Next, I argue that there are good reasons for being sceptical as to whether these incentives are appropriately taken into account at present. Third, an extensive literature review shows that we know very little about actual incentive arrangements and that what we do know is fairly worrying.

The last section of chapter 6 discusses the questions of interest and scope of the study in more detail (6.5). I explain that I concentrate on non-tradable credit risk in combination with long or indeterminate maturities where, in addition, the borrower's cashflow distribution is influenced by the bank's monitoring and considerable discretion on the part of loan officers means that human effort and judgment matter. If – unlike in the case of small-business lending – these preconditions are not met, constructing an incentive scheme would be a relatively straightforward matter. Note that the focus on non-tradable risks means that credit portfolio management is beyond the scope of my study. Next, I explain that I concentrate on line employees from both the front- and back-offices as well as work-out specialists dealing with problem loans and I argue that these employees have considerable discretion. I outline how human effort matters and how its exercise (and therefore the scope for the use of incentive schemes) is constrained by the sophistication of the available IT and software architecture. I set out the general requirements that a lending-related incentive scheme should fulfill from the point of view of a regulator: It should be able to 1) deal with staff turnover; 2) make possible a regular (at least annual) performance evaluation prior to the maturity dates of loans; 3) evaluate performance objectively and link remuneration to performance in an explicit fashion; and 4) provide incentives, in particular, for post-disbursement monitoring.

In chapter 7 I report my findings in detail. Following an introduction (7.1) I explain and defend the methodology I used in my fieldwork (7.2). The cooperation with the Employers' Federation of German Cooperative Banks which sponsored the study is described in Eggenberger (2006c). The sophistication of risk-rating and pricing tools is a potentially critical constraint for the design of incentive schemes. Therefore, in section 7.3 I outline the credit risk rating and pricing methodology in use in the sample banks and I show in which ways there is scope for human influence. I also discuss the implications for incentive schemes. In the following section (7.4) I describe possible objectives (performance measures). I outline desirable features of incentive schemes and give some institutional and organizational background. The discussion of possible objectives for front-office employees (loan officers) begins with comments on the use of profit targets, followed by the analysis of several other candidate objectives. The risk objective must be at the heart of any lending-related incentive scheme and is discussed in detail. The section ends with a discussion of possible objectives for credit analysts and the turn-around/liquidation unit. Two important caveats are tackled on the

basis of the existing literature rather than my fieldwork: In section 7.5 I review potential shortcomings of the risk measurement tools currently in use and in Eggenberger (2006c) I outline fundamental doubts with respect to explicit contingent-pay schemes.

What is the bottom line of this fieldwork? While it would be intellectually dishonest to claim to be able to give a definitive answer to the above question on the basis of the fieldwork reported here, my findings indicate that it may be possible to construct sensible schemes. I illustrate this claim here with two potential objectives:

1. Linking a front-office employee's remuneration with the profit he or she generates for the bank is conceivably a useful objective from the point of view of both bank and regulator. However, many practical issues would need to be resolved. Some of these (such as the heterogeneity of business customers that makes it difficult to design comparable portfolios for different relationship managers and may, therefore, undermine the fairness requirement) are mainly a headache for the bank. Other problems (such as possibly inaccurate profitability estimates) are more directly a concern from the regulator's point of view. On the basis of my exploratory interviews it is not possible to gauge the extent to which such difficulties make it inadvisable to use a profit objective.

2. By comparing actual and expected loan losses on a portfolio, it may be possible to measure the quality of a loan officer's risk-related work. While such an objective may expose employees to considerable risk, simple work-arounds may be able to defuse this problem and sufficiently isolate employee performance from exogenous fluctuations. It is not possible to say with confidence that such an arrangement would be effective in practice, but my interviews suggest that it may well work.

I discuss these and additional candidate objectives in detail and present obstacles and potential problems discovered in the interviews. In summary, my findings do not provide conclusive proof that the regulation of incentive compensation in banks should be a tool of financial regulation. However, the results of my fieldwork constitute a first stab at the question of whether there is scope for devising effective lending-related incentive arrangements that align the interests of a regulator with those of the banks. My hope is that regulators will build on these findings and use their influence to further explore this sensitive but critically important area.

6.2 Introduction and Motivation

This chapter and the next aim to establish whether there is scope for devising effective lending-related incentive arrangements in practice. While the question is straightforward, the methodology employed in this chapter, interview-based fieldwork, is sufficiently unusual for an economics doctoral thesis to warrant a brief explanation. I came across the subject of lending-related incentive arrangements via the theoretical work reported in the preceding chapters. I wondered how the elements of mainstream theoretical work correspond to reality and I started looking for literature on incentive structures in banks. The dearth of available material piqued my interest and I decided to pursue the topic further.

One angle from which to study real-world incentive structures is as an extension to theoretical work akin to a robustness check. In the theoretical models of bank competition reviewed earlier in this thesis, there is much discussion of the effort exerted by 'the manager'. In those models the link between the intensity of competition and effort, while often subtle, is always reasonably transparent. A theoretical model hopes to gain insight into a problem by isolating and capturing some aspect of reality. One prerequisite for that approach is to consider banks as black boxes such that a single agent represents hundreds or even thousands of employees. In reality, not surprisingly, the mechanism through which competitive pressures are transmitted to and within the black-box that is the bank is far more complex and complicated. Therefore, as far as the real-world applicability of theoretical work is concerned, one crucial question is how the pressures that the bank is facing at the institution level are transmitted throughout the organization. As an economist, one is trained to look for and analyze incentives as a key element of this transmission mechanism. In particular, organizational slack might be linked with the absence of appropriately designed incentive schemes. The work presented here can thus be read as a complement to the theoretical work in the preceding chapters.

While the perspective adopted in the earlier parts of the thesis is that of a policy-maker keen to maximize surplus, the concerns outlined above also matter from the point of view of financial stability: A lot of things can go wrong even if incentives are cleverly designed at the institution level. In this sense the perspectives of a policy-maker whose priority is surplus and that of a financial regulator who seeks to prevent systemic crises (and, to some extent, individual bank failures) are different, but overlap. In both cases the basic issue is how to influence behavior in desirable ways.

Given that in the rest of the thesis to all intents and purposes I ignore the possibility of bank failures, I decided to approach the subject of lending-related incentive arrangements from a regulator's point of view and ask whether banks could be made safer through cleverly-designed rules on permissible or mandatory lending-related incentive arrangements where incentive arrangements are understood in the narrow sense of an explicit link between objective performance measures and remuneration. (That is what a regulator could impose in contrast to, say, intrinsic motivation on the part of bankers to take risk in a disciplined manner.) Such rules on compensation arrangements would limit the autonomy of owners and top management to run their banks as they see fit and thus constitute an intrusion, but the principle of intrusion surely is not contested if there is to be regulation at all.

That particular intrusion might also be a rather useful one. I make this claim on the basis of an extensive literature review [included in Eggenberger (2006c) and summarized below] in which I show that banking failures and banking crises are systematically associated with dysfunctional incentive arrangements. There is little information on incentive structures in the public domain and what information there is is fairly worrying. All this suggests that far from being a peripheral issue, incentive structures should become an important concern for regulators. To the best of my knowledge, other researchers have not ventured beyond making that assertion. My contribution is to investigate what actually could be done about dysfunctional incentive arrangements. In my fact-finding I look at issues that may appear pedestrian to a theorist, but in analyzing the scope for sensible lending-related incentive structures it surely would be most inappropriate to assume away the concerns of practitioners.

This work is built on the premise that a desire to avoid systemic crises can justify restrictions on permissible incentive contracts or might even entail prescriptive regulation. The evidence I have seen suggests that existing incentive arrangements are problematic from a financial stability point of view. In many cases such incentive structures may not even be the result of a conscious tradeoff between different objectives (e.g., senior management consciously inducing loan officers to take risk in ways that appear imprudent from the regulator's point of view), but may reflect simple carelessness. I realize that this conjecture is difficult to square with our usual idea of rational actors, so given the lack of hard evidence I do not wish to overstress this point.

By way of analogy, I note that prior to the Basel II discussion many

banks argued that it was not necessary to differentiate loans very finely by risk and that sophisticated statistical rating tools were a waste of time because such systems would not achieve anything that the judgement of an experienced banker could not accomplish without all the formality. A few years later people thought about this very differently. So (abstracting from unintended flaws in the New Capital Accord) this may even be a case of regulation in which all sides have gained. It is possible that judicious regulation of incentive arrangements may have similar beneficial effects.

It is one thing to assert that existing incentive structures are dysfunctional; it is quite another to actually analyze whether one could do better. That involves looking at many nitty-gritty issues, as the only way to find out what we need to know is to talk to practitioners. As I document below, among academics, regulators, and even practitioners, the issue of lending-related incentive arrangements is largely uncharted territory.

In terms of practicalities, I focused on working with German banks. This was partly for logistical convenience, but mainly because the German banking sector is characterized by a high degree of fragmentation and variety in organizational forms. I eventually worked with cooperative banks, but at the outset it seemed reasonable to try for a sample as varied as possible. Having said that, I was also in contact with banks from other countries and German for-profit as well as public-sector banks.

I first attempted to convince the Bundesbank to support my project. The board member in charge of banking regulation confirmed that the question was of great interest, but pointed to legal restrictions and noted that the Bundesbank by law has to keep confidential all information obtained from the banks it supervises (meaning that I would not be able to use it in my thesis). There was also concern that banks which were at the time (spring of 2002) assisting the Bundesbank prepare for the revision of the Capital Accord through various quantitative impact studies and surveys should not be burdened with further requests for cooperation. I therefore contacted banks directly to gain their support on my own. This was an exercise as gratifying as it was frustrating. It was gratifying because the interest my proposal generated on the part of senior practitioners proved that I was on to something. In particular, the interest among banks far larger than the banks with which I eventually worked suggests that the subject has not been dealt with in a convincing way outside my sample. Anecdotal evidence gathered in conversations with practitioners and regulators supports that view. With respect to the topic of interest, there may actually be little difference between banks

(including banks outside Germany).

The experience was also frustrating as I invested a considerable amount of time and effort into getting banks to support my research only for the project to be ground up by organizational politics. With the benefit of hindsight, it has become clear to me that compensation arrangements may be among the most sensitive issues in the life of an organization[1]. By contrast, bankers were often surprisingly comfortable talking about their risk rating technology and revealing information that I would expect to be far more commercially sensitive than information on remuneration structures. In the case of almost all banks that I contacted, I initially approached the person in charge of credit risk methodology as I intended to look at (actual and potential) incentive structures from a risk management point of view. In a number of cases the negotiations reached an advanced stage (progressing from telephone conversations to more detailed proposals to meetings with senior managers) before organizational politics of some kind (one unit being interested, but other units not cooperating; unease about the reaction that participating in such a study would provoke on the part of the union) brought discussions to an end or resulted in offers of cooperation too restricted for me to be able to tackle the question of interest. In addition to disclosing its compensation policies and allowing access to line managers, a bank also would have needed to explain its risk management infrastructure and run the risk that potential shortcomings would be discovered and spill over into the public or semi-public[2] domain.

The bottom line is that it was very difficult to gain the kind of access to banks required to tackle the research question. However, working directly with banks is the only way to obtain the information needed and I did eventually manage to find an organization to work with. There are few cases that I am aware of in which a researcher has managed to gain that kind of cooperation. Comparable work either involves established academics [e.g., Bartel (2004), James (1996), Santomero (1997)] or regulators [e.g., Treacy and Carey (1998), English and Nelson (1998)] who have substantial powers of persuasion short of invoking any statutory

[1] By way of illustration, in a discrimination lawsuit recently brought against Morgan Stanley which was widely reported in the financial press, the bank – despite apparently having a strong case in purely legal terms – settled out-of-court for an amount of USD 54 million rather than having to divulge details on its compensation policies. The Economist (2004f) noted that it was not hard to see why a settlement suited the bank: 'Morgan Stanley dreaded seeing tales of its traders' boorishness in the tabloids and on daytime television. (...) Just as scary, perhaps, a trial would have revealed details of people's pay, to the glee of the firm's competitors and the ire of some of its employees. And all this even if Morgan Stanley had won.'

[2] As part of the confidentiality agreement that I signed, I am allowed to use my findings in the thesis, but I am not allowed to use them in any other way.

rights to information.

I am confident that my findings are a substantial step forward and shed light on the problems inherent in designing and implementing lending-related incentives. I hope that the analysis of both existing and desirable incentive arrangements will be taken further by others. As will become clear, the only way to advance our knowledge is to talk to practitioners.

6.3 Incentives in the context of bank regulation

In section 1 of Eggenberger (2006c) I demonstrate that regulation of incentive arrangements might be a useful addition to the current bank regulatory framework – a framework that has steadily evolved from a focus on quantitative regulation to a more qualitative approach. Sub-section 1.A outlines why banks may be thought to be 'special'. Sub-section 1.B sketches out the current regime of bank regulation centered on capital requirements. Sub-section 1.C looks at regulatory strategies beyond capital requirements and, in this context, briefly examines the case for the regulation of incentive arrangements in banks.

6.4 The Case for Regulating Incentive Compensation

This section – like the preceding one – constitutes background material and is included as section 2 in Eggenberger (2006c). In the following, I provide a summary.

In 2.A I outline the evidence on the role of incentives (and, therefore, their regulation) in bank failures and banking system distress. The main findings are the following:

- Bank failures have multiple and diverse causes.
- Human error and other (in principle) avoidable problems dwarf the role of exogenous shocks.
- Bad loans are the leading proximate cause of failure; credit risk management in particular is an area of concern.
- Dysfunctional lending-related incentives are a major contributing factor to bank distress.

Sub-section 2.B asks why regulators should bother to regulate compensation when this is really a task for owners. A number of mainly theoretical arguments are considered. However, it is argued that only a look

at the evidence can reveal whether the issues outlined ought to be considered material real-world concerns or of merely theoretical relevance. Specifically, one would want to know which incentive arrangements are in place, whether they cause problems and whether they can be improved.

These questions are addressed in the following sub-section (2.C) as part of an extensive literature review. More specifically, I examine

- Descriptive material on human resource management policies in banks, in particular compensation practices;
- Academic studies of top-management compensation;
- Other academic work that touches on human resource management practices, including compensation, in banking;
- Publications by and/or for practitioners, banking textbooks, and academic publications that are prescriptive in nature;
- Credit-rating agencies' material; and
- Pertinent Government Regulations (Germany, U.S., and BCBS)

The upshot of this literature review is that little can be learnt about actual incentive schemes in real-world banks from public sources; there is hardly any literature on the subject, even when a generously broad search protocol is used. However, I show that what little we do know is fairly worrying. None of the sources presents a blueprint for a sensible incentive scheme nor describes such an incentive scheme in operation. Scattered references in practitioner material, the literature on the causes of bank failures and the anecdotal evidence I gathered in numerous discussions with practitioners and regulators suggest to me that the paucity of publicly available information on lending-related incentive arrangements cannot be taken as an indication that these do not matter. Dysfunctional incentive arrangements have been identified as a key driver of bank failures and a cause for concern even for banks that are in good financial health. Thus, the most plausible explanation for the apparent gap in the literature is that the lack of material reflects a dearth of appropriately designed incentive schemes. The conclusion that emerges from this literature review is that there is no substitute for talking to practitioners to examine in more depth those schemes in use and the arrangements that could be effective and feasible. In particular, one question hanging over the entire issue is whether it is at all possible to construct sensible incentive arrangements. [An appendix (2.D) discusses the sources used.]

In the following, I present selected quotations that illustrate the concerns of a number of prominent academics with respect to lending-related incentives.

Llewellyn (2002) asserts that

'incentive structures and moral hazards faced by decision-makers (bank owners and managers, lenders to banks, borrowers and central banks) are central to an understanding of bank crises.' (p. 160)

Expanding further on the incentive structures within banks, he notes that the following issues are relevant:

'the extent to which reward structures are based on the volume of business undertaken; the extent to which the risk characteristics of decisions are incorporated into the reward structures; the nature of the internal control systems within banks; internal monitoring of the decision-making of loan officers; the nature of profit-sharing schemes and the extent to which decision-makers also share in losses, etc. High staff turnover and the speed with which officers are moved within the bank, may also create incentives for excessive risk-taking. (...) [B]ank managers have frequently been rewarded on the basis of the volume of loans made. Many cases of bank distress have been associated with inappropriate incentive structures creating a bias in favour of balance sheet growth (...).' (p. 162)

John, Saunders and Senbet (2000) argue that

'bank regulation, through capital requirements and asset restrictions, has limited effectiveness, given the high leverage ratios of banks.' (p. 96)

They advocate a

'prominent role for management compensation structures in bank regulation'

and point out:

'Unlike capital and asset regulations, which have at best indirect effects on managerial incentives and thus on managerial decisions, altering top-management compensation is a direct and effective way of influencing managerial return and risk-taking incentives.' (p. 97)

Santomero (1997) points out the desirability of appropriate incentive schemes, but like the other authors cited here has clearly not investigated in detail how such arrangements could be made to work:

'To the extent that management can enter incentive compatible contracts with line managers and make compensation related to the risks borne by these individuals, then the need for elaborate and costly controls is lessened. However, such incentive contracts require accurate position valuation and proper internal control systems. (FN 5 omitted) Such tools, which include position posting, risk analysis, the allocation of costs, and setting of required returns to various parts of the organization, are not trivial. Notwithstanding the difficulty, well-designed systems align the goals of managers with other stakeholders in a most

desirable way. (FN 6 omitted) In fact, most financial debacles can be traced to the absence of incentive compatibility (...).' (p. 87–88)

Goodhart (1996) also addresses the problem of incentive arrangements from the point of view of bank regulation, but is more skeptical with respect to its political feasibility. He observes:

'The influence of capital adequacy on risk-seeking behaviour is understood by most people, and the consequential rationale for external regulators to enforce (graduated) additional controls over financial institutions with insufficient capital is generally accepted. The iso-morphism with the effect on risk-seeking behaviour of the pay structure is generally not well understood (FN 8 omitted), and there is little willingness or acceptance by internal management for allowing external regulators to have any say in the matter.[3]' (p. 11)

Llewellyn (2000) writes that

'while external regulation has a role in fostering a safe and sound banking system, this role is limited. Equally and increasingly important are the incentive structures faced by private banking agents and the efficiency of the necessary monitoring and supervision of banks by official agencies and the market.' (pp. 69-70)

He argues that

'two structural shifts are needed within the regulatory regime: (1) external regulation needs to become less prescriptive, more flexible, and more differentiated among different institutions, and (2) more emphasis needs to be given to incentive structures and the contribution that regulation can make to creating appropriate incentive structures.' (p. 70)

More specifically,

'The key is that appropriate internal incentives need to be developed for management to behave in appropriate ways and that the regulator has a role in ensuring that internal incentives are compatible with the objectives of regulation. Combining appropriate incentives for owners and managers contributes to a robust financial system, and, in principle, the market would evolve such incentives. However, experience indicates that, in many areas, and most especially when the competitive environment is changing and the regulatory regime is being adjusted, it is hazardous to rely on the market evolving appropriate incentives.' (p. 95)

On the basis of the evidence, it would appear that banks fail because of avoidable problems and that, whatever the ultimate causes may be,

[3]The measure that Goodhart recommends is to 'require the internal audit committee (...) to signify that they have considered the implications for the risk preferences of key personnel of their pay structures.' (p. 17)

credit losses are the primary proximate cause. Dysfunctional incentive arrangements are a major contributing factor. Clearly, banks that do not manage to offset their loan losses with adequate interest margins on non-defaulting borrowers will eventually go bankrupt themselves. If bankers personally bore (some of) the cost of poor decisions, is it not conceivable that they would exert greater care in making lending decisions?

Before one can think about how supervisors might regulate incentive arrangements in this area, an important question arises that the preceding quotation hints at: Why should regulators bother to regulate compensation when this is really a task for owners? I address this question in more detail in Eggenberger (2006c). One possible answer is that regulators' objectives differ from those of the owners. Another possible justification is that corporate governance problems in banks make it difficult for owners to implement sensible schemes, whereas this would be relatively straightforward for a regulator. One possibility that one should not dismiss out of hand is that simple thoughtlessness leads to dysfunctional incentive arrangements being adopted. Goodhart (1996), for example, observes that

'the internal committees and groups that examine risk, and its control, within firms virtually never consider internal pay structures. Equivalently, the (personnel) committees and groups within companies that decide on remuneration virtually never consider, or discuss, the implications of what they are doing for risk-seeking behaviour.' (p. 11)

Evidence that thoughtlessness might be playing more of a role in this area than the concept of homo oeconomicus allows for is also provided by scattered references in the practitioner literature. For example, Nadler (2000a), in the American Banker, writes:

'Consider policies that reward lending officers for the volume of loans they place on the books rather than on successful repayment. Such policies create a great temptation to make marginal loans. And if such a loan eventually goes bad, the lenders may be long gone from the organization. Indeed, the whole concept of performance-based compensation needs thorough examination, to make sure that individuals do not make decisions that reward themselves but place the bank at risk.' (p. 5)

In another article, Nadler (2000b) puts it somewhat more drastically:

'As for rewarding loan officers for placing new loans on the book, this is like buying a deck chair on the Titanic. It is easy to make a loan; the job is to get your money back on time and with interest. How many banks have paid rewards to hotshot loan officers who leave long before the bank finds out that it has a sour credit on the books?' (p. 6)

Bird (2002) reports that

'Another bank's program rewarded lenders only for loan volume growth. The loan portfolio grew rapidly, but the lenders were not held accountable for documentation errors, rapid loan deterioration, special mention loans, etc. As a result, during the recent recession the new loans did not hold up as well as previous loans.' (p. 9)

Nadler (2003) claims that

'It is much easier to put a loan on the books than to get it paid off. Yet most incentive programs for lenders base the rewards on the former.' (p. 5)

As for the German practitioner literature, an article by Bölz and Thorsch (2002) is very insightful. These two authors wonder why – given banks' enormous expenditures on risk management systems and information technology during recent years – German banks were hit just as hard by the most recent recession as by previous recessions. They argue that financial institutions have indeed built up sophisticated risk management systems, but that

'their role in the decision-making process is questionable. Personal bonus payments of line management decision-makers, which frequently amount to more than half of total annual pay, still depend on volumes and gross returns. Risk-adjusted quantities that appropriately take into account risk at the moment the loan decision is made are almost never used in incentive systems.'[4]

Rathmann (2002) quotes an industry consultant as saying that

'Banks' focus is on volume growth. People [in the back office] frequently do not dare to voice their concerns if the risk appears too large.' (p. 17)[5]

At any rate, as I demonstrate in Eggenberger (2006c) we know little about existing incentive structures and even less about which arrangements would be both effective and feasible. In the following section, I give an overview of the questions that we would want to have answers to; these questions outline the scope of my fieldwork.

[4] 'Fraglich ist jedoch deren Bedeutung im Entscheidungsprozess. Noch immer hängen die persönlichen Zielboni der operativen Entscheider, die nicht selten weit mehr als die Hälfte des gesamten Jahressalärs betragen, von Volumina und Bruttoerträgen ab. Risikoadjustierte Kennzahlen, die bereits im Zeitpunkt der Kreditentscheidung das Risiko adäquat berücksichtigen, fehlen nahezu gänzlich in den Anreizsystemen.'

[5] 'Die Betonung in den Kreditinstituten liegt auf dem Volumenwachstum. Da wagt dann oft keiner mehr, gegen ein Kreditengagement zu sprechen, wenn das Risiko ihm zu hoch erscheint.'

6.5 Questions of interest

In this section, I will discuss the four issues that have determined the scope of my investigation of possible lending-related incentive arrangements. My only claim here is that I have covered the essential questions that any such investigation would want to address. One could certainly approach the issue in different ways.

1. Which of the products that banks sell are relevant for the research question? Which business areas should be considered?

2. Which groups of employees are relevant for the research question?

3. What are the roles played by human effort and the technical infrastructure, respectively?

4. Which general requirements should lending-related incentive schemes fulfill?

6.5.1 Which of the products that banks sell are relevant for the research question? Which business areas should be considered?

Throughout the paper, I will concentrate on lending to small- and medium-sized companies by commercial banks[6]. Smaller companies do not usually have access to other sources of funding; how well banks assess and manage risk in this segment will therefore have important implications for the wider economy. The second characteristic that makes smaller-company debt particularly interesting is the fact that the bank's own monitoring of the borrower post-disbursement can be expected to influence the probability of default. I shall discuss this point in detail below. Third, for all practical purposes, smaller company debt is non-tradable, and there are, therefore, no market prices. Smaller companies do not usually have a credit rating from a rating agency and even the debt of those that do will likely remain highly illiquid for the foreseeable future. Fourth, lending to smaller companies tends to be highly human-capital intensive, both while the loan is negotiated and after it has been disbursed. This is because

[6]The exact definition of what constitutes a loan to a smaller or medium-sized business in the sample will be discussed when I describe the rating models used for different customer segments. For 'official' definitions, see the discussion on p. 53 of Hommel and Schneider (2003). They outline both the EU definition and the one typically used in Germany which is proposed by the Institut für Mittelstandsforschung. Conditions for a company to be small- or medium-size relate to the number of employees (EU: < 250, IfM: < 500) and turnover (EU:<EUR 40mn, IfM:<EUR 50mn). The EU definition incorporates additional requirements with respect to total assets and the share of capital that is held by a large company. The Hommel and Schneider paper also provides a good introduction to the subject of SME financing in Germany.

smaller companies are particularly informationally opaque. The important role of human judgment and effort imply that effective incentive arrangements could be particularly valuable in this business segment.

The particular focus chosen means that I do not have much to say about incentive schemes in other types of financial intermediaries. Investment banks, for example, do not perform traditional commercial banks' transformation function (turning short-term savings into long-term loans). The risks that these investment banks take on are mostly traded in markets. As a consequence, there are usually market prices and risk management systems available that make, for example, performance evaluation for security traders fairly straightforward. While spectacular losses due to 'rogue traders' make for good headlines, trading losses have tended not to sink banks in industrialized countries (a well-known, but systemically irrelevant exception being the case of Barings).[7] Market risk in general is much better understood than credit risk, and risk management techniques are correspondingly more sophisticated[8]. (However, there are worries about current market risk management models, which I discuss below.)

Tradable credit risk in its various manifestations (corporate bonds for which liquid markets have existed for many years, outright loan trading, securitized assets of all kinds, credit derivatives) raises its own set of questions, but touches on issues very different from those of non-tradable credit risk that are considered here. Where market prices for credit risk are available, devising an incentive scheme for employees that trade this risk becomes fairly straightforward, at least at a superficial level.

As my research focus is on non-tradable credit risk in individual transactions (expected loss), credit portfolio risk (unexpected loss) management warrants only a brief discussion. Every transaction through which a bank takes on credit risk changes the bank's portfolio. Risk management on the credit portfolio level can only yield its full benefit (for example, by reducing concentrations) when credit risk is tradable. When credit risk is

[7]The only episode that I am aware of in which errors in market risk management might potentially have had systemic implications is the rather special case of the hedge fund, Long-Term Capital Management. The fall of LTCM which, in the words of The Economist (2000a) 'came about through a combination of greed and arrogance', is a sobering example of how excessive confidence in historical relationships and correlations applying in the future compounded by other errors can lead even an apparently very sophisticated risk management strategy astray. For a very readable account of the demise of LTCM, see Edwards (1999). Jorion (2000) provides a far more in-depth analysis. The official report by the President's Working Group on Financial Markets (1999) is usefully supplemented by testimony from senior officials: Greenspan (1998), McDonough (1998), Gensler (1999), and Parkinson (1999).

[8]A clear distinction between market risk and credit risk is, of course, only a hypothetical construct. In reality the two kinds of risk are clearly related, even if they are not usually managed in an integrated way yet.

non-tradable, credit portfolio management (which will typically exploit correlations between the values of assets with different characteristics) is restricted to affecting new business, typically via crude devices such as quantitative restrictions (limits) for certain industries, jurisdictions, risk classes etc; the use of credit portfolio models for pricing is still in its infancy. Alternatively, a credit portfolio manager would have to use financial instruments for hedging that expose the bank to considerable basis risk.

The specific difficulty in creating incentives for good lending-related performance in the smaller company segment thus arises from the fact that no market prices for loans are available. The true value of a loan is unknown and can at best be estimated. For loans of short maturities, that might not be much of a concern. However, a maturity of ten or more years was not unusual in my sample and that is a long time during which to be uncertain as to whether the borrower will repay. It is also a time horizon over which staff turnover will be a major consideration in almost any financial institution. The role of human effort in lending to smaller companies stands in sharp contrast with loans for which the banker has virtually no discretion and a computer takes all the decisions. The credit scoring techniques that are used in segments such as auto loans, mortgages, and credit card loans are described in Mester (1997). She notes that

'[Scoring] has not been widely applied in business lending (...). One reason for the delay is that business loans typically differ substantially across borrowers, making it harder to develop an accurate method of scoring.' (p. 3)[9]

While one might consider Mester's paper superseded in view of rapid technical progress in the credit risk area, more recent publications speak of the continuing difficulty of automating the loan process for small- and medium-sized business borrowers. For example, the Federal Reserve notes with respect to business loans,

'Although credit scoring is becoming more universal, considerable het-

[9]The U.S. Small Business Administration's Office of Advocacy defines a small business as one with less than 500 employees for research purposes [Small Business Administration (2004)]. Different standards apply as far as the official designation as a small business is concerned. The designation is useful, for example, in government contracting. Mester does not give a definition, but the loan sizes that she quotes for illustration clearly indicate that she is referring to businesses at the lower end of the size distribution. In Federal Reserve Call Reports, a small business loan is defined as a loan of less than one million dollars whereas a loan of less than USD 100,000 is termed a micro-loan [see p. 15 of Board of Governors of the Federal Reserve System (2002)]. The definition used in flow-of-funds analyses is different again, leading to discrepancies [see FN 49, p. 61, Board of Governors of the Federal Reserve System (2002)]

erogeneity remains in evaluating loans across banks'[10] [p. 12, Board of Governors of the Federal Reserve System and U.S. Securities and Exchange Commission (2000)].

The only systematic review of the use of credit scoring that I am aware of dates back to a survey carried out by the Federal Reserve Bank of Atlanta in January 1998 [see pp. 816 – 817 in Frame, Srinivasan and Woosley (2001) for a description].[11] The scarcity of data makes it difficult to gauge to what extent scoring techniques have replaced human judgment. In sum, while progress with statistical techniques to assist or replace human judgment in loan decisions has undoubtedly been made, the (small) business loan segment does not lend itself easily to the application of such techniques. Additionally (and contrary to standard practice for credit card borrowers or homeowners taking out a mortgage, where the bank merely makes sure that scheduled payments are made) the monitoring of smaller- and medium-sized businesses post-disbursement usually involves considerable human input.

Extending credit is but one of the three classic banking functions, the other two being deposit-taking and the provision of payment services, and it is quite clear that the lending business is but one and not necessarily the major source of revenue for many banks[12]. Depending on the applicable legislation, banks may be providing many more services. The banks in the sample, for example, also sell insurance, mutual funds and other investment products of all kinds and some offer even more peripheral services, such as real-estate brokerage. However, none of these other lines of business pose the particular challenges for devising a sensible incentive scheme as credit risk and none of them have a similar potential to generate losses. Since I am looking at incentive arrangements from a regulatory point of view, the focus on lending to smaller and medium-sized borrowers seems appropriate.

[10]The section on 'Credit Scoring' in a more recent report by the Board of Governors [pp. 53 – 57 in Board of Governors of the Federal Reserve System (2002)] primarily refers back to Mester's paper in its analysis.

[11]However, the differences between scoring models and internal rating systems are not very well defined. Scoring models are sometimes used to fully automate the lending decision. As part of an internal rating system a scoring tool typically does not make but assists in making a decision.

[12]Radecki (1999), using a wide but not implausible definition of payments services, estimates that in 1996 payments-related revenue accounted for between 36% and 42.2% of operating revenue for the 25 largest U.S. bank holding companies.

6.5.2 Which groups of employees are relevant for the research question?

To explain which employees should be considered in a study of incentive arrangements, I will outline the organizational structure of commercial banks and the lending process. Interestingly, it is not entirely up to the banks themselves how they structure either their lending processes or their internal organization. The Basel Committee on Banking Supervision (2000a) released the 'Principles for the Management of Credit Risk' which have been transformed in Germany into national regulation by the financial regulator BaFin (Bundesanstalt für Finanzdienstleistungsaufsicht) as Mindestanforderungen an das Kreditgeschäft der Kreditinstitute or 'Minimum Requirements for the Lending Business of Financial Institutions' [BaFin (Bundesanstalt für Finanzdienstleistungsaufsicht) (2002)]. This regulation stipulates that a financial institution's board of management has to develop a credit risk strategy (para. 9) that outlines the planned development of lending activities by industry, geographical breakdown, etc., (para. 11) and review it yearly (para. 10). This task cannot be delegated (para. 9). Front-office and back-office have to be clearly functionally separated; the back-office, independent of the front-office, has a vote in any lending decision (para. 25). Both have to agree if a loan is to be approved (para. 31). If they cannot find a common point of view, the loan application must either be rejected or the decision is to be made by more senior representatives of each function (escalation procedure, para. 34). Independent monitoring of risks at the portfolio level and independent reporting are tasks that must not be assigned to the front-office (para. 26). Front-office and back-office have to have separate reporting lines up to and including the board of management (para. 27). Certain kinds of collateral are to be assessed by a unit other than the front-office (para. 28). There are provisions on 'intensive care' for borrowers in difficulty (para. 56), as well as a work-out unit (para. 58). Work-outs are to be monitored by a unit other than the front-office (para. 58). The regulation also restates a statutory requirement for regular monitoring through internal audits (para. 93).

While other countries have different regulations dealing with these issues, the above extracts give a good idea of the different functions in a bank that are involved in the lending process. Top management will typically formulate strategic objectives. The actual work of acquiring new borrowers or extending further credit in existing client relationships falls to the front-office. The back-office is charged with the processing of

the loan applications and has to review applications under risk aspects[13]. Specialists for assessing the value of collateral may be involved. If difficulties arise in the course of lending that appear manageable, these are to be handled by giving the borrower 'intensive care'. If things take a turn for the worse, a work-out unit will either attempt a turnaround or try to salvage as much as possible should the borrower go bankrupt. There is to be monitoring of credit risk at the portfolio level and a continual review by internal auditors. Not mentioned here but of great importance are those employees that develop the bank's risk management architecture such as the internal rating system and portfolio models.

While ideally the incentives of all employees who have a role in the lending process should be analyzed, practical considerations required concentrating on those groups of employees most interesting with respect to the question I am trying to answer: What scope is there for devising effective lending-related incentive arrangements in practice? Is it possible to construct incentive schemes that align the interests of a regulator with those of the banks? The focus of my study, therefore, is firmly on the line (as opposed to staff) functions: I shall be looking at the front-office employees (whom I will also call relationship managers or loan officers) and the back-office employees with whom they work (also called credit analysts), as well as the work-out specialists. In other words, I am taking the entire infrastructure (most notably, the risk rating systems) as given, and I will not be looking at incentive schemes for top management, credit portfolio managers, credit risk modelers and other functions outlined above.

As for the reasons for the focus chosen, note that credit portfolio management, for example, is still in its early stages. In the case of the borrowers that I will concentrate on, credit risk is not tradable, which severely hampers any portfolio management. Attempts to securitize this type of debt have been slow to get off the ground [see section 4 in Eggenberger (2006c)]. In addition, how well credit portfolio managers do their job will depend in large part on how good the models are with which they work. But evaluating the quality of risk rating, portfolio models and similar tools raises mostly technical questions entirely different from the ones I wish to address. In the light of the literature, there is reason to think that these models have shortcomings; particularly worrying is the treatment of what is alternatively called aggregate or systematic

[13]In the banks that I worked with, back-office employees as a general rule are not supposed to have any contact with clients so that their judgment is as objective as possible and not clouded by personal likes and dislikes. This is also standard practice in a number of other banks that I was in contact with, although I have not found any reference to this being a statutory requirement.

risk: changes in aggregate default premia (as priced in the markets in which credit risk is traded) and changes in aggregate actual default rates over time driven primarily by business cycle fluctuations. It is not, thus far, a risk that is well understood and the analytical treatment of this risk component is even less advanced than that of other aspects of credit portfolio risk (such as correlations between different industries). Consequently, serious doubts exist as to whether systematic risk is priced appropriately, either in capital markets or by banks' internal models. I review these concerns below with a view to exploring the implications for the performance measures I propose.

As for top management, there is a large literature – both empirical and 'prescriptive' – on executive compensation. As I have argued, the existing literature on banks and the incentives they face tends to view financial institutions as monolithic 'black boxes' and does not consider the organizational issues that arise from the fact that decisions are not made by a single individual – either 'owner' or 'manager'. This is especially true for business-lending, where bank employees have a relatively large degree of discretion. This paper will fill this gap by looking at incentive schemes for the rank-and-file whose decisions, in the aggregate, determine the future health of individual financial institutions, as well as the wider financial system, just like those taken by the chief executive and senior management. Whether one looks at these issues through a corporate governance lens or from a regulator's perspective, the fundamental question is always whether and, if yes, how the behavior of these front-line bankers can be constrained or influenced and how incentives get broken down from larger units to individuals. Analyzing incentives for owners and top managers is a very indirect way of investigating the actual decisions on the approval and pricing of loans taken by all (including more junior) employees. This is not to say that giving owners and top management the right incentives is not important, but it is an area reasonably well covered by the existing literature.

6.5.3 What are the roles played by human effort and the technical infrastructure, respectively?

The precondition for any incentive scheme is some discretion on the part of the employee concerned: His or her effort, diligence, creativity or some other unobservable skill must matter. In view of the rapid advances in credit risk measurement and management technology during recent years, one might wonder how much human influence is left. Above I suggested (1) that the bank's own monitoring of the borrower post-disbursement

can be expected to influence the probability of default; (2) that lending to smaller companies tends to be highly human capital intensive both while the loan is negotiated and after it has been disbursed; and (3) that the role of human effort in lending to smaller companies stands in sharp contrast with other kinds of loans, where the banker has virtually no discretion and the decision-making process is largely or fully automated. In the following I will detail where this human influence comes into the lending process, sketch out the role of technology and outline the relationship between the two. The argument at its essence is that human effort matters, but that its contribution is dependent on technology in the way that, for example, the performance of a cyclist depends on his or her bicycle. Inadequate technology may critically compromise an otherwise effective incentive scheme.

Any incentive scheme has to follow the information technology (IT) and software infrastructure (for accounting, risk rating, pricing, and other elements of the loan process). The sophistication of that infrastructure may therefore constrain sensible schemes. In fact, based on my experience I would claim that incentive arrangements cannot be studied fruitfully without a working knowledge of the technical systems in use. Broadly speaking, human effort matters in three phases of the lending process, which are influenced by technology to varying extents. These are

1. Risk rating;
2. Pricing; and
3. Monitoring

The first task in which human effort plays a role is the initial risk assessment. For example, while credit card loans are scored and the bank's IT system will decide on approval and applicable interest rate, judging the creditworthiness of a business customer is far more complex. Statistical tools are widely used and contribute to the analysis, but the creditworthiness assessment typically still involves a fair amount of judgment and processing of 'soft' information. The assessment normally results in the assignment of the borrower to a risk class (calibrated such that it represents a certain default probability). It comprises the evaluation of collateral and it provides the front-office employee with a benchmark for pricing the loan. With slight exaggeration, one could claim that there is no 'wrong' risk class, only 'wrong' interest rates: Why should a bank not lend to a risky borrower if it is properly compensated for the risk that it is assuming?

The creditworthiness assessment is crucial for the front-office employee in negotiating the interest rate and other terms of the loan. The

market for loans to small- and medium-sized companies is such that the price tends to be only one of several criteria that borrowers use to determine which bank to borrow from[14], so there is typically considerable scope for the relationship manager to negotiate an advantageous rate for the bank. These negotiations on terms and conditions are the second step in the lending process in which human effort matters.

Third, effort and diligence are very important in the ongoing risk-rating, including monitoring of and interaction with borrowers, as well as the work-out and liquidation functions (which become involved in case a borrower gets into difficulties). As mentioned above, a key characteristic of loans to small- and medium-sized companies is that the probability of default is not independent of the monitoring provided by the loan officer assigned to that particular client relationship. This is because to some extent (and typically constrained by legal, i.e. liability considerations), a good banker will act as a financial advisor, pointing out operational weaknesses, discussing different scenarios with the owner or CEO of the client company, which will on average make the borrower financially more robust. Another very important task of the relationship manager is to sound the alarm as early as possible if things go wrong. The earlier a borrower in difficulty is given 'intensive care', the higher the probability that any rescue attempt will be successful (and that the bank will get its money back). If the borrower goes bankrupt, collateral has to be realized and effort will typically determine how much of its losses the bank can recoup. The post-disbursement monitoring is so important that I will return to the issue at some length below in the discussion of what general requirements for lending-related incentive arrangements should be.

Thus, it is clear that the frontline banker is dependent on the bank's technology in a number of ways, most importantly when using the bank's statistical tools in the credit rating process. In fact, the relationship between human effort and judgment and the available technology parallels that between staff and line functions. Some central risk management unit will typically be in charge of developing and refining the bank's technical infrastructure and the front office must rely on the accuracy of the tools provided. As long as portfolio considerations are not transmitted to line

[14]For hard evidence supporting this observation (which I found confirmed in almost all of my interviews), see Greenwich Associates (2002). That study, which also covers non-credit topics, is based on interviews with more than 19,000 corporate financial managers from U.S. companies with between 10 and 500 million USD in annual sales. The chart on p. 1 that specifically deals with the importance of various factors from the point of view of banks' customers is based on information from more than 600 companies. On p.1, one of the authors states that 'Banking for middle market companies is largely a relationship business, and decisions to enter or break a relationship revolve around many more factors than credit pricing.'

staff via the pricing tools, the information on how desirable exposures (to particular countries, industries, or regions) are from a portfolio point of view has to be made available to the front office in some other way. A typical (if crude) arrangement is for a central unit to impose quantitative limits, which take into account the lending strategy formulated by top management. The frontline banker has to rely on these instructions being sensible. In a similar way and related to portfolio concerns, one also would want economy-wide or even global developments to be taken into account. This would involve, in particular, watching out for imbalances, such as exchange rate misalignments, building up of real estate, credit, equity and other bubbles, risks inherent in budget deficits and the current-account position. It would not be reasonable to expect individual bankers to obtain this information on their own initiative, although it is clearly relevant for lending decisions. Again, the front-office must be able to rely on the information provided[15].

In sum, how well the front-office bankers do their jobs is constrained by the technical infrastructure and instructions from headquarters. If, for example, the risk-rating tools were inadequate, the ability to make good lending decisions might be severely hampered. As two practitioners discussing a similar problem put it:

'We have employed risk-sensitive capital tools on an internal basis for some time. We have learned – sometimes the hard way – that a rigid application of these formulas can lead to bad decisions if the human element is forgotten' [Ervin and Seidel (2002)].

Also problematic are strategic or portfolio considerations that turn out to have been misguided. One senior banker told me: 'If management decides that the bank should lend more to the telecoms industry and the telecoms industry subsequently experiences a sharp downturn, it is hardly

[15]Another factor that one might expect to be of importance in this regard are centrally mandated credit standards in the sense of non-price loan terms. Lown, Morgan and Rohatgi (2000) put the reasoning thus: 'By raising loan rates, lenders may drive off all but the least creditworthy applicants or elicit riskier behavior by borrowers. Rather than raising loan rates to curtail the supply of credit, lenders may tighten their standards and cut off credit to the marginal borrowers that do not meet the higher standards. In essence, credit markets may operate like a trendy night club in New York City : you have to clear the velvet rope before you pay the door charge.' (p. 4) The U.S. evidence suggests that these credit standards have historically had considerable predictive power with respect to lending and output [see Lown, Morgan and Rohatgi (2000) who use the responses from the quarterly Federal Reserve Senior Loan Officer Opinion Survey to measure how stringent lending standards are] and one might thus expect that credit standards must have affected frontline bankers. However, the better banks get at pricing, i.e. the more risk-sensitive pricing becomes, the less they will have to rely on non-price terms and the type of rationing alluded to here to overcome any adverse selection problems. It is not clear what the impact of finer pricing on moral hazard is. At any rate, I will not investigate non-price terms in detail.

the individual banker that arranged the loan who is to blame.' In view of the role that the risk-rating infrastructure plays, it must be a part of any discussion of incentive schemes. The focus of this analysis, however, is on the 'human factor'. The bank employees considered in this study have at least some discretion to influence lending decisions and their effort matters enough to warrant thinking about the development of feasible and effective incentive arrangements.

6.5.4 Which general requirements should lending-related incentive schemes fulfill?

Four issues that have a bearing on the scope of the study need to be discussed under this heading:

A. Dealing with staff turnover

B. Performance evaluation prior to maturity dates of loans

C. Objective performance evaluation and explicit incentive schemes

D. Incentives for post-disbursement monitoring

Additional requirements that are important to practitioners are discussed below; these four are merely the key requirements for regulators.

Dealing with staff turnover

As pointed out before, maturities of ten years or more would not be unusual in the banks with which I worked. Over a time horizon of ten years, staff turnover, as well as internal transfers, are a major concern. Not surprisingly, staff turnover also has been identified as a major problem when incentive schemes are not well constructed. Under a scheme, for example, that rewards non-risk-adjusted loan volume, the temptation is to make as many loans as possible without regard to the borrower's creditworthiness and then move on to the next bank before loans turn sour. There is usually a considerable delay between disbursement and default, a phenomenon known as the 'aging' effect and well-documented for publicly rated bonds[16]: For a given rating grade, the default rate is significantly lower during the first couple of years following origination than in later years. I will therefore be looking for incentive schemes that can deal with turnover.

[16]See, for example, FN6, p. 7, in Moody's Investors Service (2004), an annual default study published by that rating agency. The aging effect is here defined as follows: 'The 'aging effect' is based on the empirical observation that bond issuers' risk of default is highest in the third or fourth year after issuance.' The aging effect is somewhat worrying for rating agencies, as it implies that apparently some information is systematically not incorporated in their ratings.

Performance evaluation prior to maturity dates of loans

From a non-practitioner's point of view, the simplest solution to these difficulties would be to make bonus payments dependent on realized profitability (and to pay the bonus only when a loan has been repaid as promised). I floated this idea a number of times in my interviews and was usually laughed at: Getting a bonus for today's work in ten years' time is an idea whose time has not yet come. Based on this anecdotal evidence as well as practical considerations (including job rotations, employees leaving the bank altogether, shared responsibilities), I am not convinced that any such solution would find acceptance – a key requirement for an incentive scheme to work in practice. I will, therefore, be looking for incentive arrangements that allow for performance evaluations on at least an annual basis.

Objective performance evaluation and explicit incentive schemes

My interest throughout will center on explicit incentive schemes based exclusively on objective performance measures. First, such a scheme suited the banks with which I worked. However, more general reasons to focus on explicit schemes include the following:

- Subjective evaluations are almost by definition not verifiable which may make both the employer and the employee view such evaluations with suspicion or lead employees to curry favor with their evaluator;
- Subjective evaluations are exceedingly hard (and very costly in terms of time) to do well, as banker after banker told me in the interviews[17]. If they are not done well they can be highly counter-productive because any perception of unfairness breeds resentment;
- The 'technology' for a subjective evaluation scheme, aside from some general statements one might find in a company handbook, resides in the person of the evaluator and is thus inherently non-transferable within banks or from one bank to another except by a lengthy process of 'learning by doing', getting feedback from more experienced colleagues, or

[17] A well-designed subjective evaluation would have to focus on the criteria that really matter, bearing in mind that one needs to restrict attention to a few items for the evaluation to remain operational. The supervisor would need to gather and record 'evidence' (observations of the employee to be evaluated in different situations) continually and systematically and, in doing so, would need to guard against any personal biases. The evaluation itself might have to involve highly critical feedback without 'burning bridges'; procedural justice is very important in such evaluations. Finally, a supervisor also might need to be able to defend relative rankings in a way that convinces employees of the fairness of the assessment. A lot of research suggests that people are – contrary to what standard economic theory postulates – quite concerned with relative position and not just how they are doing in absolute terms. For a brief introduction, see pp. 102 – 104 in Baron and Kreps (1999).

other informal methods; and

- My primary interest is the use of incentive arrangements as a regulatory tool. A system based on subjective evaluations would be hard to impose by a regulator and certainly would not lend itself to scrutiny in the same way as an explicit incentive scheme.

My focus on explicit schemes does not imply that subjective evaluations are superfluous or unhelpful. In fact, adding subjective elements to an objective evaluation can be an excellent strategy to reduce the gaming inherent in such schemes.

Incentives for post-disbursement monitoring

Whether and, if yes, how employees should be held responsible for the post-disbursement performance of their clients was easily the most contentious issue I encountered during my study. Since this issue is at the heart of what lending-related incentive arrangements ought to be about, I will explain in some detail why I believe that for the specific product that I am looking at, loans to small- and medium-sized companies, any arrangement that does not incorporate incentives for post-disbursement monitoring is likely to be inadequate.

Three alternative approaches need to be considered[18]. These are incentive arrangements using as objectives:

1. non-risk-adjusted loan volume;
2. expected profit without explicit incentives for post-disbursement monitoring; or
3. expected profit plus some measure of actual profitability that incorporates explicit incentives for post-disbursement monitoring.

1. Some authors implicitly endorse arrangements where bonus payments depend on non-risk-adjusted loan volume. Baker (2000), for example, offers the following rationale firmly rooted in standard principal-agent-theory:

'The typical incentive plan for loan officers in a bank involves paying for "originations": the loan officer receives a bonus for lending money. The puzzle about this type of incentive is that it gives the loan officer no incentive to search for and write "good" loans, that is, high-interest-rate loans that are likely to be repaid. Instead, loan officers have incentives to make any loan, and banks typically have credit committees (made up

[18]In theory a lot of other schemes based on different objectives are conceivable. However, I focus on the three schemes discussed here because 1. non-risk-adjusted loan volume seems to be a widely used objective, at least among U.S. banks, and 2. and 3. are objectives discussed in the literature. These are also the only three types of objectives that I have come across that can make a claim to being sensible.

of higher-level bank officers) whose job it is to determine the creditworthiness of the potential debtor and to approve or deny the loan. (FN 1 omitted) An alternative choice for banks would be to pay loan officers on the eventual profitability of the loan, rather than on its origination. (FN 2 omitted) Bonuses based on loan profitability would have the advantage of giving loan officers incentives to search out good credit risks and sell loans with higher expected value. In other words, such a performance measure would provide less distorted incentives to the loan officers. However, such a scheme would also give the loan officers greater risk, since many things can happen to debtors that are unpredictable when a loan is written. In this case, it appears, the trade-off between risk and distortion is made in favor of lower risk and higher distortion.' (p. 417)

While the partly exogenous nature of credit risk was indeed an important point of discussion in my interviews, the consensus was that regardless of the business cycle and other exogenous influences which Baker refers to there is enough scope for relationship managers and credit analysts to 'work on their clients' to hold them (at least partly) responsible for how their borrowers are doing. This suggests that Baker's explanation is 'too clever'; an economic theorist's hypothesis rather than a description of real-world reasoning on the part of practitioners[19].

Baker (2000) also points out that under the arrangement described above

'the loan officer need not be expert in determining creditworthiness' (FN1, p. 417).

The specialization that is possible as a consequence of only a few credit analysts making loan decisions might reduce the banks' costs by more than any resultant increase in loan losses or might even reduce loan losses. While this is a theoretical possibility, of the sources I have reviewed only Wuffli and Hunt (1993) write that some banks claim that this is actually the case:

'In the United States, for example, many banks have concluded that it is no longer feasible to maintain the skills of hundreds of loan officers through their branch systems. They have elected to limit lending authority to a smaller number of regionally based credit experts. These underwriters are highly trained and carefully motivated specialists. Centralizing credit decisions in this way means that only 70 to 100 people make virtually all the lending decisions – except, perhaps, for those related to large corporate clients. Not only does this concentrate credit

[19]Indeed, in personal communication Baker told me that his 'evidence on how loan officers are paid is almost entirely impressionistic' (email exchange with George Baker, 11 November 2003)

decisions in the hands of the most capable individuals; it also creates centers of information about credit markets, making it possible to exploit areas that were previously too diffuse to handle. The banks that believe in this system argue that the loss of direct customer knowledge is more than compensated by better underwriting skills, more concentrated information, and greater objectivity in making decisions.'

Wuffli and Hunt also note that

'Another approach is to embark on a major skill-building program for all lenders, setting strict standards and regularly reviewing all limits. So far this approach has had mixed reviews. Some banks report success in boosting skills, but others have found the large-scale training and review procedures difficult to implement.'

In other words, where relationship managers are rewarded for originations, they might simply not be skilled enough to contribute to the risk assessment and this might be a profit-maximizing arrangement. Of course, in market segments where the lending decision is automated (and post-disbursement monitoring is not an issue) that would be a suitable arrangement. But even in that case, should front-office employees not have an incentive to negotiate as high an interest rate as the borrower will accept rather than as low an interest rate as they can get away with? One standard theoretical argument against this would be that higher interest rates are not necessarily profit-maximizing because they aggravate the twin problems of adverse selection and moral hazard. Even if this argument applies, the converse is not true: Surely profits are not necessarily maximized by charging a very low interest rate. I therefore suspect that implicit in the above description is a lending policy centered on a single interest rate for all accepted loan applications (non-risk-adjusted interest rates were standard until well into the 1990s in many banks) with the 'yes or no' approval decision turning entirely on non-price terms. The specialization argument may have made sense at the time of Wuffli and Hunt's writing in 1993 when credit analysts' judgment played a far more important role in the lending decision than it does today. However, with the implementation of statistical risk-rating tools that allow discrimination by creditworthiness via risk-adjusted pricing, basing bonus payments on loan volume regardless of interest rate appears sub-optimal.

I can think of other possible justifications for using non-risk-adjusted loan volume as an objective but have never seen these stated explicitly. For example, a loan-volume objective is easier to administer for the bank. Also, using a risk-adjusted target might mean that the front-office employee has to spend more time doing calculations and consulting with his

or her credit analyst; i.e., such an approach creates a larger administrative burden for the relationship manager and less time might be spent with clients.

Thus, it is not impossible that banks see the shortcomings of schemes based on non-risk-adjusted objectives, but they rationally choose to use these regardless because the benefits outweigh the costs. However, the bankers that I talked to (whether from my sample or other institutions) rather uniformly thought such schemes flawed, and I also have found very critical voices in the available literature (see above). Such an approach would not provide any explicit incentives for monitoring post-disbursement. In the light of the evidence, I did not consider non-risk-adjusted loan volume in my study. Concentrating on such a system considerably restricts possible incentive arrangements and would, therefore, have limited the scope of the study.

2./3. Using the risk rating and pricing tools that I describe in more detail below, it is possible to calculate the bank's expected profit on a loan at the time of origination. The key factor that determines actual as opposed to expected profitability is whether the borrower repays or not. As part of the risk-rating process, the front-office employee and the credit analyst estimate the expected loss on the loan. As I pointed out above, some authors assert[20] that this is all that one can ask of employees: A relationship manager who has priced in the expected loss has done his or her job; any incentive pay should be based on the expected profit rather than the actual profit. I should note that even expected-profit-based schemes have not been analyzed in any degree of detail in the literature.

In some circumstances the use of an expected-profit objective is clearly appropriate. If compensation depends on the expected profit, the bank becomes the residual claimant with respect to the difference between expected loss and actual loss; the employee no longer has any explicit incentive to be concerned with the borrower's creditworthiness. I would argue that it makes sense to absolve employees from responsibility for a borrower's credit risk post-disbursement if one of the two following conditions applies:

- The bank can hedge the credit risk by buying default protection (e.g., a credit-default swap), i.e., there is a market price that determines whether the employee has negotiated a good deal for the bank. Such credit protection is available in some market segments [mostly for large borrowers, see the recent report on credit risk transfer by the Committee

[20] One cannot say that these authors 'argue', for I have not yet come across any source discussing the pros and cons of the different approaches systematically. See Eggenberger (2006c)

on the Global Financial System (2003) and Rule (2001)]. Insuring against default or using market prices for the evaluation of profitability might thus be a suitable strategy to pursue for some wholesale banks.

- The bank cannot hedge the credit risk, but human effort does not influence the bank's profit (loss) post-disbursement. In some market segments, it is not cost-efficient to 'monitor' borrowers except in so far as the bank uses automated tools to ensure that borrowers make their contractual payments. In market segments where all monitoring is automated, the lending decision will typically also be largely or even completely automated. The relevant risk factors in such settings are not normally assessed by the front-line banker, but by a central risk management unit for the entire bank.

Clearly both conditions are not fulfilled for lending to small- and medium-sized enterprises:

- There is no market price for credit risk / no default protection; and
- effort and diligence affect the probability of default (and other determinants of actual profit).

This would suggest that a scheme also incorporating incentives for post-disbursement monitoring is superior to a scheme that uses expected profit. If there are other factors to be taken into account that tilt the balance in favor of the latter kind of arrangement, I am not aware of them[21]. As for post-disbursement monitoring, two observations suggest that its value is considerable:

1. There was no doubt among the practitioners in my sample banks that monitoring is important, but there is surprisingly little research evidence on this. However, Carey (1998) has looked at data on privately placed (and thus 'monitored') bonds and compared the loss experience with that for publicly held bonds. Crucially, both categories of debt have risk ratings which serve as indicators of ex ante riskiness. Carey finds that 'in the investment grades, private debt performs as well or better than public debt. Private debt performs better in the riskier grades,

[21] Since loan losses reduce profits, one might think that a system built on expected profit only cannot be superior to a well-calibrated system that also takes into account post-disbursement monitoring. From the practitioner's point of view this might not be so clear-cut. As I pointed out above with respect to 'volume only' schemes, looking at expected profit might well entail lower administrative cost than the alternative. It might be felt that incorporating actual profitability could make loan officers too risk-averse; the bank might fear that under such a scheme, the relationship managers would forego too many good deals to avoid a few bad ones. However, if an expected profit based scheme really is superior for these or any other reasons, I am not aware of anybody having made the case for it (let alone having backed this up with data). This applies in particular to RAROC (risk adjusted return on capital) methodologies which measure risk-adjusted expected return and are in use in some banks.

consistent with monitoring having substantial value.' (p. 1385) While Carey uses data on bonds and not bank loans, the results are suggestive.

2. Apart from regular monitoring (which in the case of my sample banks would involve activities such as discussing business plans and scenarios with the borrower, obtaining information but also reacting to it, giving financial advice), the other key element of post-disbursement monitoring is the early-warning function. There was consensus among the bankers in my sample that time is of the essence when borrowers get into difficulty. One would therefore want any significant deterioration in a borrower's condition to be detected as early as possible so that problems can be confronted aggressively. Crucially, while part of the early warning process can be automated, a good relationship manager will often be aware of bad news before it is reflected in the account transactions or financial statements. Again, no statistics on the value of such early warning are available, but apart from the answers of my interviewees, there is qualitative evidence suggesting that this is an important issue. Voigt and Andersch (2003) report the results of a survey based on responses from around 250 German banks that suggests that smaller banks in particular recognize and deal with borrowers' difficulties too late. That article also discusses quantitative and qualitative warning signals; banks generally attribute considerable importance to qualitative warning signs such as changes in the behavior of a borrower's management (p. 193).

In sum, the features of the lending process can be summarized as follows: The relationship manager is in regular contact with the borrower; partly to sell bank products, but also to collect soft information for risk assessment and monitoring purposes. Post-disbursement monitoring is important and is not automated. It makes sense for the same person, the relationship manager, to be assigned the tasks of both selling and monitoring, i.e. to combine the sales and risk aspects. Splitting up these tasks and involving a second person would lead to duplication and possibly confusion. Thus, if there is to be monitoring at all it would appear to make sense to assign it to the relationship manager.

The question then arises whether one could make the relationship manager's remuneration dependent on expected profit only and still have him or her be responsible for post-disbursement monitoring. In the light of Holmstrom and Milgrom (1991), it must appear problematic to reward sales and not provide incentives for keeping an eye on risk. One large bank with which I was in contact does not use explicit incentives for post-disbursement monitoring, but the risk-related work of the relationship managers is evaluated (subjectively) by the credit analysts that

they work with; this evaluation co-determines the bonus. Although that particular institution claimed to have been apprehensive with respect to the lack of explicit incentives, the fear of undesirable consequences has subsided somewhat during the four years that this arrangement had been in operation at the time of writing. Apparently in this case the subjective evaluation provides strong enough incentives to not neglect the risk aspect. Since I decided to look at explicit schemes, this approach is outside the scope of my study. I will be looking at arrangements that also provide explicit evaluation of post-disbursement monitoring.

Apart from the considerations laid out above, two additional reasons for post-disbursement monitoring should be mentioned: 1. The bankers that I talked to, while critical with respect to details, thought that this was the right approach. 2. In view of potential shortcomings of the risk-rating technologies in use, one might not have much faith in 'expected profit' being a very accurate measure of actual profitability. If, for example, the procyclical nature of current rating tools distorts the risk assessment, tying remuneration to some measure of actual, long-term profitability might mean that the loan officer's judgment to some extent counter-balances the shortcomings of technology. The Bank for International Settlements (BIS) (2001), for example, recognizes this implicitly:

'Remuneration arrangements that focus on short-run outcomes and relative, rather than absolute, performance may discourage a long-term perspective and an assessment of aggregate risk.' (p. 133)

Herring (1999) notes that

'The line officers who are in the best position to assess dangers of increasing vulnerability should have incentives that encourage them to take a long-term view of the institution's exposure. This generally means calibrating bonuses to long-term measures of risk-adjusted profitability rather than short-term returns that make adjustment for risk.' (p. 76)

The Federal Reserve System Task Force on Internal Credit Risk Models (1998) in its discussion of the RAROC (risk-adjusted return on capital) approach notes that

'Increasingly, measurements of shareholder value added are being used by banks in gauging managerial or product performance and in determining managerial compensation.' (p. 10)

FN 9 adds that:

'The shareholder value added concept described above measures an activity's "excess cash flows" above and beyond the bank's relevant opportunity costs for a single period. To mitigate potential distortions created by such a myopic measure of performance, banks typically use

multi-year, rolling averages of shareholder value added when evaluating ex post profitability or managerial performance.'

I take this observation as evidence that clearly supports my assertion that incentive arrangements need to incorporate some measure of actual profitability[22].

Summary of the general requirements that lending-related incentive schemes should fulfill

I shall be looking at incentive arrangements for line employees in lending to small- and medium-sized businesses. I acknowledge that there are constraints imposed by technology, but there should be sufficient human influence to make incentive schemes feasible. The schemes that I will be looking for should be able to deal with staff turnover and to allow performance evaluation of employees before the maturity of a loan. The schemes should be explicit and based on objective performance measures, and should incorporate incentives for post-disbursement monitoring.

Since this section deals with the scope of the study, it is a good place to note that my remarks here apply only to non-criminal behavior. Fraud and related phenomena cannot really be discussed in the context of standard incentive contracting. The study also concentrates on monetary incentive arrangements whereas in reality a lot of other incentive devices (such as promotions) matter. As is in the nature of a case study, the banks with which I worked, the credit risk management software they use, and other aspects of the study are not representative of all commercial banks. However, the issues and questions raised within this specific context hold relevance for other financial institutions, as will be repeatedly demonstrated throughout this analysis.

In Eggenberger (2006c) I discuss related mainstream economics literature and provide brief reviews of work on relationship lending (theoretical and empirical work), small-business lending, and the advantages that small banks have in lending to small businesses (section 3). I also address the question whether the kind of non-automated, labor-intensive lending that is the focus of my study is about to disappear because of technological advances (section 4). I conclude that on the balance of the evidence, the importance of relationship lending and, therefore, the rele-

[22]Unfortunately, the study does not contain any more detailed material on this point. The work of the Federal Reserve System Task Force on Internal Credit Risk Models 'involved extensive discussions with twelve [U.S.] banking organizations, two non-bank securities firms, and several credit risk consultancies' [Federal Reserve System Task Force on Internal Credit Risk Models (1998), p. 3, FN 3]. The banks were best-practice organizations at the time the research was conducted (April 1996 to May 1998).

vance of the present study should be expected to be reduced by advances in technology, but that any such erosion is likely to be very gradual. In the following chapter, I set out the methodology and results of my fieldwork in detail.

Chapter 7

Lending-related incentive arrangements: Fieldwork

7.1 Summary

In the preceding chapter, I argued that while little is known about lending-related incentive schemes, these might conceivably become a useful tool of bank regulation. The research question addressed in the present chapter is: What scope is there for devising effective lending-related incentive arrangements in practice? Is it possible to construct incentive schemes that align the interests of a regulator with those of the banks?

I approached this research question by individually interviewing approximately 120 senior practitioners in a sample of about 25 German cooperative banks discussing various aspects of the question (stretching from "soft" topics such as human resources management principles to "hard" topics such as the measurement and pricing of credit risk). The sample banks are relatively small and use a common risk-rating and pricing methodology which the banks were willing to explain to me in great detail. The banks' typical clients, aside from households, are small- and medium-sized enterprises, so the banks provided an ideal environment in which to examine the monitoring that is seen as one of the key functions of banks and the ways in which bankers can be given incentives to perform this function. The interviews in the cooperative banks (which typically lasted about one to one and a half hours) were supplemented with a number of meetings and telephone interviews in larger commercial banks both inside and outside Germany.

In a nutshell, the findings indicate that it is possible to devise incentive schemes that both bank owners and regulators would approve of. In discussions with the interviewees I identified a number of objec-

tives that could be used to evaluate performance. The most important of these are sensibly defined measures of profit as well as quantitative indicators of the quality of risk management. It is sometimes argued that, given the long maturities of many loans, it would not be feasible to assess the performance of the bankers responsible for these and that, at any rate, it would not be possible to sufficiently isolate bankers' individual performance from exogenous risk factors (such as the state of the overall economy) to which they are exposed. The interviews show that these concerns should not be exaggerated in that there are simple work-arounds for dealing with these complications.

These candidate objectives should now be tested in a field experiment. If these performance measures prove their effectiveness in practice, guidance on the design and implementation of such incentive schemes could then be added to banking regulators' rulebooks. The interviews identified a number of "Dos" (e.g., balance objectives carefully so that employees have appropriate incentives not to neglect any important duties, ensure a very tight integration of performance measures and the analytical tools used to assess and price risks) and "Don'ts" (e.g., don't make bonus payments immediately, wait a bit until there is more information on loan performance) that should be respected. Any incentive scheme should be carefully monitored with respect to two concerns that also came up in the course of my research: First, current risk measurement tools may have potentially serious shortcomings; this might make performance measures less reliable than they ought to be. Second, explicit contingent-pay schemes may not fit into the human resources management strategy of each and every bank. Doubts with respect to explicit incentive arrangements have been voiced for a long time and bank management would do well to monitor the effect that such incentive schemes have on employee behavior. For example, it is well known that team-work and cooperation may suffer if individual incentives are too steep.

The chapter is structured as follows: Since the only way to research the question posed was through field-work and interviews – an approach that is standard in other social sciences but rarely used in economics – I provide a discussion of the methodology (7.2). The risk rating and pricing techniques are crucial tools without which sensible incentive arrangements would be difficult if not impossible to construct. I therefore discuss the risk rating system in use in the sample banks (7.3) before presenting my findings in detail by discussing possible elements of an incentive scheme (7.4). This is followed by the above-mentioned discussion of potential flaws in current risk measurement tools (7.5). The conclusion

highlights key results and reviews the implications of the findings for the regulatory use of incentive arrangements (7.6).

7.2 Methodology

The only way to find out which, if any, lending-related incentive arrangements could be expected to be effective and feasible was to interview practitioners. The interviews were not carried out with the purpose of building a dataset that could then be used for econometric analysis, but with the aim of identifying major themes that can further understanding of the problems inherent in devising and implementing incentive schemes. This methodology is standard in a number of other social science disciplines, including anthropology, social psychology, and organizational behavior, as well as fields such as epidemiology and socio-medical sciences. In public health, for example, qualitative research methodology is well-established and makes up a significant proportion of U.S. public health graduate programs with required methodology courses on interviewing techniques, questionnaire design, and qualitative data analysis. In qualitative field work, a sample size of 15 to 25 per 'segment' (a defined subgroup; e.g., one might have segments differentiated by socio-economic status) is considered very good[1].

The use of interviews and qualitative fieldwork as a research tool is slowly gaining ground in economics, but remains sufficiently non-standard to warrant a brief discussion. As outlined in section 6 of Eggenberger (2006c), getting a hands-on understanding of production processes is a key ingredient in 'insider econometrics', where the qualitative work is combined with the collection of data for more mainstream econometric work. Bartel (2004), for example, in her work on human resource management policies in banks points out that

'Previous studies of productivity in the banking industry indicate the importance of getting "inside the black box" [Berger and Mester (1997)], which can only be done through detailed analysis at the plant level, i.e. the branch.' (pp. 181-182)

Maybe the best-known proponent of interview-based work is Truman Bewley. Bewley (1999) deals with the question of (work) morale and outlines the findings of his book-length study of 'Why wages don't fall during a recession.' That paper gives a very good illustration of how

[1]For a critical view of the current research methodology in empirical economics work, see McCloskey and Ziliak (1996) and Ziliak and McCloskey (2004) as well as the numerous other articles on the topic in the same issue.

insights gained by interviewing can inform subsequent theoretical work[2]. The first contribution in mainstream economics making use of interviews that I am aware of is Alan Blinder's study of the reasons for price stickiness. Blinder (1991) contains a very good description and defense of the methodology; also see the discussion by Shiller, Gordon and Grossman (1991) who offer support as well as criticism. Two recent examples of the use of interviews in banking research are De Haas and Naaborg (2005a) and De Haas and Naaborg (2005b). The authors, researchers at De Nederlandsche Bank, use interviews with practitioners to better understand the role of internal capital markets in multinational banks and the impact of foreign bank entry on SME access to credit, respectively.

The two papers that I found most helpful from a methodological point of view were Helper (2000) and Bewley (2002). Helper starts by pointing out that modern economics began with Adam Smith's visit to a pin factory and proceeds to outline some insights from the NBER's field research program. She notes that

'Economists today typically do their research using econometrics and mathematical modeling. These techniques have many strengths but share the weakness of distance from individual economic actors. In contrast, field research allows direct contact with them, yielding several advantages.' (p. 228)

These are:

1. 'Researchers can ask people directly about their objectives and constraints.' (p. 228)
2. 'Fieldwork allows exploration of areas with little preexisting data or theory' (p. 228)
3. 'Fieldwork facilitates use of the right data' (p. 229)
4. 'Fieldwork provides vivid images that promote intuition' (p. 229)

In the second part of the paper, Helper argues that

'many criticisms of field work can be answered with improved methods.' (p. 229)

She deals in particular with techniques for increasing objectivity, replicability, and generalizability. Bewley (2002) is titled 'Interviews as a valid empirical tool in economics' and is more of a how-to guide than Helper's paper, although Bewley discusses methodological issues as well.

My study, like most interview-based work, has certain features that would be problematic if my objectives were different. For example, I signed a confidentiality agreement that means that my study is non-replicable for other researchers. (Bewley, however, points out that others

[2]Also see the discussion by Taylor (1999) in the same volume.

are always free to do a similar study that would confirm or contradict the findings.) The sample also is not random, but systematic – I worked with 25 banks that indicated an interest in participating in the research, that all have the same organizational form (they are organized as cooperatives) and are all relatively small compared to industry giants such as, for example, Deutsche Bank. There is thus a triple sample-selection bias: First, banks other than co-operative banks were not systematically included in the sample (although I did talk to many other bankers from for-profit and public sector banks in the course of my research). Second, the organization sponsoring the study contacted only around 700 (or around half of its) member banks through their electronic distribution system and requested that banks indicate whether they would be willing to participate in the study[3]. (In return for participating, banks received a detailed report on the findings.) Third, from within the self-selected group of all cooperative banks that had indicated an interest in participating (about 80 strong), I chose an impressionistic sample: I strove to obtain the maximum possible variety in terms of size and location while keeping the sample to manageable proportions. The size of the banks ranged from around EUR 150 million in assets to around EUR 2.5 billion[4].

While I did establish a master questionnaire in cooperation with a pilot bank (where employees were particularly generous with their time and patient with my questions), sticking rigidly to a list of questions impairs the flow of the conversation with the interviewees. I therefore used the questionnaire more as a rough guide for myself rather than as a standardized list of questions to tick off. As my understanding of the issues grew, the initial list of questions was quickly superseded anyway. Other reasons for not using a standardized format are:

- The sample is too small to calculate meaningful descriptive statistics (e.g., percent of interviewees agreeing with a statement), let alone do more sophisticated quantitative work. While at first glance a total of 114 interviews in 25 banks (plus six interviews in the pilot bank) does not look too bad, one needs to bear in mind that these interviews fall

[3]The electronic service was used since this saved time and money and, more importantly, made sure that top management and heads of human resources of member banks would be alerted to the study directly (even within banks, access to the service is highly restricted). My contacts at the sponsoring organization hold the view that banks using the electronic service are not systematically different from the others.

[4]While I believe that the significance of my findings is not diminished because of the way the sample of participating banks was obtained, it would have been interesting to include other, explicitly for-profit and public-sector banks. As noted in the introduction, I invested considerable time and effort into convincing other banks to join the study.

into at least five sub-categories. In each bank I typically interviewed staff from the management tier just below top management overseeing human resources, controlling/credit risk management, the retail front-office, the front-office serving business customers, and the back-office and in each of these interviews quite different issues were discussed.

- The study was explicitly exploratory in nature. Not following the leads that presented themselves quite naturally in the interview would have meant foregoing a lot of valuable information.
- The 'psychology' of the interview situation: It was very important for the success of the research to keep my interviewees interested in the discussion and motivated to help me. They are busy people: An hour and a half or more of their working day is a substantial amount of their time. I found that turning the interview from a somewhat lop-sided question-and-answer session into more of a discussion was much appreciated by the interviewees. Demonstrating detailed knowledge of their business and empathizing with their concerns was very helpful for building credibility. Some of the interviewees had thirty-plus years of work experience in their field and it was gratifying to observe how engaging their interest with probing questions very often turned initially somewhat reserved politeness into enthusiasm. To create this sort of interview atmosphere while rigidly sticking to a questionnaire is, I believe, impossible.

I usually spent one full day in each bank and held one-on-one interviews with the heads of human resources, controlling/credit risk management, the retail front office, the front office serving business customers, and the back office that usually lasted about one and a half hours each. I wrote to the sample banks before the interviews to request permission to record the interviews with a digital voice recorder; almost all of my interviewees were comfortable with that. I believe that it is impossible to take detailed notes and be an active interviewer at the same time; recording the interviews and analyzing them in detail later on while taking only essential notes in the interview (points to be followed up on later in the conversation) is a (time-consuming) solution to this problem.

The first objective for the interviews was to understand all elements of the credit process that have a bearing on incentive issues. In particular, this meant looking at risk rating and pricing tools as well as organizational arrangements in great detail. I also discussed individual bankers' discretion (their scope for influencing decisions) and various aspects of possible incentive pay arrangements. The interviews actually covered more ground than I report here; part of the agreement with the organization that sponsored my work was that I would not focus narrowly on

lending-related incentive schemes for small- and medium-sized business loans, but would consider the issue more broadly.

It would have been desirable for my interviewees to review and comment on a draft of the present chapter. However, I was asked not to pursue that route of obtaining feedback. Instead, my sponsors circulated the internal report of my findings (wider in scope than the present chapter, and looking at incentive arrangements from the point of view of senior management rather than a financial regulator) to the participating banks and indicated that there were no fundamental disagreements.

In section 5 of Eggenberger (2006c) I provide a brief description of the banks that I worked with, preceded by some background information on the German banking industry.

7.3 Risk-rating and pricing

A thorough understanding of a bank's risk-rating and pricing tools is essential for any analysis of lending-related incentive schemes. It is particularly fortuitous in this respect that the Bundesverband der deutschen Volksbanken und Raiffeisenbanken (BVR) e.V., the Federal Association of German Cooperative Banks, has published a fair amount of material on the cooperative banks' credit risk rating and pricing methodology, VR-Control, and the software that implements it, IFB-Okular. (Cooperative banks are not strictly speaking obliged to use VR-Control/IFB-Okular, but all the banks that I visited did and so do most of the other cooperative banks.) The Federal Association of German Cooperative Banks developed this methodology for a number of reasons, not least to prepare member banks for the requirements of the New Capital Accord and to create a sufficient degree of homogeneity or even uniformity in credit risk assessment to enable the securitization of assets on a larger scale than hitherto possible. Most other banks consider such information highly proprietary and do not make it available to the public. In the following I outline the pricing methodology used and, since credit risk considerations are an important component of profitability, discuss the risk rating system.

7.3.1 Pricing scheme

The pricing system proceeds via the computation (or, rather, estimation) of a sequence of profit contributions by deducting various cost components from the revenue obtained through a product sale. The standard reference for this approach in the German-language literature is Schieren-

beck (2001) (especially vol. I). The estimation procedure as outlined here is performed by front-line bankers only in the case of negotiated product sales; it is too involved to be used for the sale of standard products where the individual banker has relatively little scope for adjusting terms and conditions as set by the bank.

A key feature of the pricing system is that all figures are expressed in present-value terms. This approach has important implications: It makes all product sales directly comparable with one another, but of necessity relies on its assumptions being accurate. Here I illustrate the estimation procedure (available to bankers as a software application within IFB-Okular) with the example of a standard, fixed-maturity loan.

Hypothetically refinancing the loan on the capital market (i.e., the bank borrows money with the same maturity as the loan, such that the payments it receives from and makes to its capital market counterparty exactly offset those to and from the borrower under the loan agreement; the software has an interface such that it can be continuously fed current market information), the bank is left with a surplus called profit contribution I. The decision on whether to actually refinance the loan (and thus hedge the interest rate risk) or accept the transformation risk rests with the bank's treasurer and does not concern the front-office employee negotiating the loan. He or she does not assume any interest-rate risk. Next, the borrower might default. By estimating the expected loss and deducting this from profit contribution I, profit contribution II is obtained. Note how the availability of market prices for credit risk would make a fundamental difference: Instead of estimating the expected loss in the complicated way that I will outline below, the bank could simply use the market price for credit risk, just as it already does today for interest rate risk. A credit portfolio manager could decide on whether to accept or hedge a credit risk, just as the treasurer manages interest rate risk. Next, the costs of production (primarily staff time) need to be taken into account, which leads to profit contribution III. The bank also needs to hold regulatory and economic capital against the loan to protect it against unexpected losses. Deducting the cost of holding this capital, one arrives at profit contribution IV. More finely differentiated estimation techniques that take into account, for example, differences between regulatory and economic capital could be considered, but would not add much to the present discussion. The key target variable for cooperative banks at this point is profit contribution III; a shift to profit contribution IV is being phased in. Schierenbeck (2001) covers the above estimation scheme in great detail; Kroon and Pool (2002) provide a concise discus-

sion tailored to the cooperative banks' methodology.

7.3.2 Risk rating

Before looking at the specifics of VR-Control, the discussion of some general background material is helpful. A number of surveys on actual practice are by now outdated, but provide an excellent introduction to the issues of interest. In his survey of commercial bank risk management (incorporating field work and visits to banks), Santomero (1997) describes risk management practices and shortcomings in a number of banks. The banks surveyed were 'best practice' banks at the time and so the analysis was not representative of the U.S. banking system. However, in view of the pace of progress in the field, the wider banking system should not only be expected to have caught up; the paper is most likely largely superseded by now, as Santomero (2003) himself points out (in another excellent article). For example, facility (and not just borrower) ratings are probably standard nowadays, and the recommendations at the end of the earlier paper have largely been taken on board by banks. Operational risk is not mentioned at all in the 1997 article, but had become a 'hot' topic by the time of writing (late 2005). Still, the paper provides a very interesting introduction and also covers a number of issues related to portfolio concerns that I am not looking at in much depth. The discussion of the need for an aggregation of various risks (market, credit, and others) is prescient; little progress has been made in this area. A number of papers have grown out of research on the credit risk management practices of U.S. banks by the Federal Reserve in the late 1990s in anticipation of changes to the Capital Accord. Federal Reserve System Task Force on Internal Credit Risk Models (1998) provides an overview of the features of then 'state of the art' credit risk models in use in the most sophisticated U.S. banks at the time. The study was prepared

'to assess potential uses of banks' internal credit risk and capital allocation models within the supervisory process' (p. 3)

as now proposed by the New Capital Accord. The paper actually takes a top-down approach and begins with a discussion of the distribution of portfolio credit losses before moving on to the risk rating of individual customers. Several of the techniques discussed in the paper (e.g., the 'subjective' approach to assessing credit risk) are likely not in use anymore, such that parts of the discussion are of largely historical interest. However, the report still deals with a number of conceptual questions in a (mostly) non-technical manner and makes for a good introduction. The actual construction of risk rating models is not discussed.

The articles by Treacy and Carey (1998), Treacy and Carey (2000), and English and Nelson (1998) concentrate on stand-alone rather than portfolio risk and cover a larger sample [practices at the fifty largest U.S. banks and the banks that are included in the Survey of Terms of Business Lending, respectively]. Treacy and Carey (1998) touch on incentive issues and conflicts of interest in the rating process; their paper is also a nice illustration of the use of qualitative interviewing. They discuss organizational issues as well as the use of judgment vs. the use of statistically based models, the risk factors considered, data and other pertinent issues. Treacy and Carey (2000) cover much the same ground as the earlier paper. English and Nelson (1998) use data obtained via the Federal Reserve's Survey of Terms of Business Lending and telephone interviews with the banks included in the Survey. They expand the scope of the analysis to include smaller and therefore less sophisticated banks. They also look at the relationship between interest rates and assessed loan risk. Another source of information on actual risk rating practices is the Basel Committee on Banking Supervision (BCBS). Basel Committee on Banking Supervision (1999c) contains an excellent overview of actual credit risk modeling practice (focusing on technical issues) from a regulator's point of view; a later report [Basel Committee on Banking Supervision (2000c)] contains a summary of comments received on the earlier one. This report is particularly instructive as the Basel Committee's experts have clearly highlighted what were the key unsolved questions at the time. A less technical report on internal rating models [Basel Committee on Banking Supervision (2000d)] is based on discussions with about thirty sophisticated banks. Krahnen and Weber (2001) provide an overview of 'generally accepted rating principles'.

The sources cited above are partly outdated by now, but still make for a very good introduction to the conceptual issues. A good source with respect to current 'best practice' (a relative term, as the field moves rapidly) are the requirements for rating systems as stipulated by the New Capital Accord [see Part 2, section III.H., Minimum Requirements for IRB Approach, in Basel Committee on Banking Supervision (2004), pp. 81 – 112].

The textbooks discussed in section 2.C of Eggenberger (2006c) [Caouette, Altman and Narayanan (1998), Crouhy, Galai and Mark (2001), Ong (1999), Saunders (1997), and Saunders (1999)] are also useful, but what really helped to become acquainted with the concerns of practitioners were a series of articles in the journal Bankinformation: Erxleben and Krob (2002) provide an overview of the entire VR-Control architec-

ture, Kirmße, Ehlerding and Putzer (2001) describe the credit process and responsibilities of the various parties involved; the risk rating tools are covered in a series of papers by Kirmße and Jansen (2001), Kramer (2001), and Nowak (2002). Kroon and Pool (2002) and Döhring and Hromadka (2002) look at pricing. Portfolio issues are dealt with in two articles by Kirmße and Schweizer (2001) and Ridder (2002).

VR-Control comprises a total of sixteen different rating tools for different customer segments [Erxleben and Krob (2002), p. 31; the 'not-for-profit-organizations' segment was added later], the two tools relevant for the purposes of this study are the 'smaller companies' rating system (up to EUR 5 mn in sales) and the 'medium-sized companies' rating system[5] (between EUR 5mn and EUR 1 bn in sales). The two rating tools are sufficiently similar to be discussed together.

A rating is primarily an assignment to one of twenty-five risk classes, each of which is calibrated to correspond to a probability of default [see p. 27 in Nowak (2002); note that the last five risk classes all denote default states using different definitions of default]. The rating has two components[6]; a quantitative one (weight of 60% in 'smaller companies' rating and 70% in 'medium-sized companies' rating) and a qualitative one.

The quantitative component is obtained by using a scoring tool on the last two years of financial statements. The financial statements are centrally entered into a database (Geno-FBS), so that there is no discretion for employees and no scope for manual interference with the quantitative score – the computer calculates it. The question of which financial variables to use and how to measure their significance (and especially joint significance) is discussed in the articles cited above. Most borrowers by law have twelve months from the end of the financial year to prepare their financial statements. This means that in theory a rating produced in November or December 2005 might be based on the financial data from 2002 and 2003. In the meantime, the financial situation of the borrower may, of course, have changed very substantially. Most borrowers seem to submit quarterly or monthly interim financial statements (following a template developed by DATEV, an association of accountants and tax advisors, known as Betriebswirtschaftliche Auswertung or BWA)

[5] Additional criteria that have to be met are: The company has to be a separate legal entity that has been around for more than five years, it has to be 'for-profit', and the bulk of the company's business must be outside of agriculture, financial services, and real-estate. In each case where one of the criteria is not met, a separate more specialized rating tool is available.

[6] There is actually scope for including the private income and assets of an entrepreneur via the use of a sub-module in the 'smaller companies' tool. I will not discuss this in more detail as it adds little information of interest.

that do not provide information on inventories and other stock figures, but are generally seen as helpful by credit analysts. Analyzing these BWAs is somewhat time-consuming, as – in contrast to the annual statements – automated analysis tools are not available[7]. Judging from my interviews, banks are not yet systematically pushing their borrowers to submit financial statements as early as possible, but this would clearly be a useful possible objective for relationship managers.

The qualitative component is computed based on the answers to a range of 'hard' questions. A 'hard' question is one that leaves no room for subjective judgment so that any two relationship managers would answer the question in the same way. (The relationship managers are in charge of obtaining the answers to the qualitative questions, because these require fairly in-depth discussions with the borrower.) The articles cited above discuss the selection of questions. While the qualitative questions are intended to be objective in this way, it is not clear that this aim is always attained in practice. For example, when does a company have a succession arrangement for its senior management in place? Assuming there were some scope for subjective assessments to influence answers, one aspect that might be worrisome with respect to the construction of the rating tool would be if during data collection bankers answering the candidate questions (about the defaulted and non-defaulted sample borrowers on the basis of whom the rating is constructed) knew whether the borrower had defaulted or not. This might influence their answers and, as a consequence, the out-of-sample predictive power of the qualitative questions might then be reduced.

The discussion of qualitative information (which is time-consuming and, therefore, costly to obtain) brings us back to the earlier discussion of the value of these data. While I am not at liberty to discuss specifics, the qualitative information obtained in the sample on which the two VR-Control rating tools discussed here are based has been shown to be valuable in statistical terms. Note that the designer of the scheme might consider it advisable to include questions that have no statistical value for essentially psychological reasons (because the acceptance and credibility of the rating tool might otherwise be undermined[8]).

The combined quantitative and qualitative scores lead to the preliminary assignment to a risk class. The relationship manager and the credit analyst each have a veto right with respect to the risk rating assess-

[7]The annual financial statements can be submitted electronically to the bank by the borrower's accountant. A similar interface is not available for the interim statements.

[8]I am grateful to Erlend Berg for pointing this out.

ment and so some discussion might ensue before the assignment becomes definitive. There is scope (within limits) for overriding the assignment proposed by the computer. Any kind of information may be used to justify a downgrade, while only 'hard' (verifiable and verified; e.g., a borrower receiving a prepaid large order) information can serve to justify an upgrade. The target share of overrides is 10 percent. The interviewees suggested that while there were often discussions with respect to the 'correct' assignment of a borrower, it was rare for the escalation procedure to be used (in which the respective superiors of the relationship manager and the credit analyst would be asked to assign the rating). The rating also incorporates 'warning signs' leading to a downgrade, but these are signs indicative of a liquidity crisis (such as garnishments) and therefore likely to be received when the borrower is already in serious financial difficulties. The risk management system does not currently include the automatic monitoring of the account(s) data[9] and nothing in the way of the internet-based monitoring tool discussed in section 4 of Eggenberger (2006c). The responsibility for monitoring manually is shared between the credit analyst (who will typically be in charge of the financial data, following developments in the industry, and developing suggestions on the basis of the financial information) and the relationship manager [whose job it is to keep in touch with the borrower and obtain pertinent information (including 'soft' information) through personal contact, make suggestions, and request additional information].

The risk rating is, in principle, independent of a decision on a loan application. In particular, the above rating procedure will be carried out at least yearly for all borrowers, regardless of whether they are applying for additional credit. Collateral valuations will be updated at a similar rhythm. The rating may, of course, be produced as part of the bank's deciding on a loan application.

7.3.3 Human influence

There are really two questions to be answered here: 1. How does human effort affect actual profitability? and 2. How does human effort affect estimated/expected profitability as established by the pricing tool? The second question matters because of the role that estimated/expected profit is likely to have in any explicit incentive scheme. If an employee

[9]The BWAs contain some information on movements in accounts held at other banks. Requesting that borrowers regularly submit up-to-date statements for the accounts held at other banks has not, so far, been considered in the banks that I visited. I am not aware of any systematic studies of the value of such data as early warning indicators.

can influence the measured profitability of his or her work him- or herself, this has certain undesirable implications that need to be borne in mind in the construction of an incentive scheme.

It is quite clear where and how human effort comes into the lending process: For one thing, the rating system relies on the relationship manager to obtain accurate answers to the qualitative questions. The possibility of overrides leaves considerable scope for the relationship manager and the credit analyst to correct an inappropriate risk class assignment. Monitoring and especially the early detection of financial difficulties involve careful analysis and, on the part of the relationship manager, negotiating skills if he or she would like the borrower to implement suggested changes. The role of human discretion in collateral valuation deserves particular mention. The banks currently rely on a set of rules issued by regional associations on which 'haircuts' are to be applied to which kinds of collateral (so-called Beleihungsrichtlinien). These 'haircuts' are fairly schematic and do not really answer the question: 'What would we get if we had to sell this (...) tomorrow?' Few banks back-test collateral valuations at present; some of my interviewees indicated that they found significant deviations of actual sale prices from estimated valuations. In the future, as part of the VR-Control project the post-'haircut' estimates will be corrected once again to obtain 'loss given default' (LGD) estimates based on actual experience. An experienced credit analyst should be expected to quickly spot cases where the general rules are not applicable and where even the tool that LGD estimates are based on might lead to biased estimates. Because changes in estimated collateral values feed through directly into expected profitability, this issue is of considerable importance for the subject of incentive contracts.

To all intents and purposes there is no human influence in obtaining profit contribution I. The expected loss (the difference between profit contributions I and II) is obtained as the product of probability of default (estimated via the risk rating procedure outlined above), exposure at default ('How much of a line of credit is drawn at the time of default?') and loss given default. Therefore, the human influence in estimating expected loss is material. In addition to the mostly standardized statistical rating, the credit analyst also will analyze an applicant's debt service capacity in detail. This is another element of the credit process in which diligence and accuracy matter greatly. The estimated cost of production (which leads from profit contribution II to III) is typically obtained as an average of how long it takes an employee to carry out the various tasks required and, given the characteristics of the transaction, cannot

be modified. Administrative production costs ought to be estimated by each bank on its own, but in practice this is considered too costly and disruptive and so a set of general guidelines is applied. The upshot of this is that the estimated costs may be very different from the actual costs if different processes are used. This problem is magnified in the case of business loans which are highly heterogeneous although only a single cost estimate can be used per product. According to my interviewees, it is not appropriate to use such an undifferentiated cost estimate for all the different loan sales that one might encounter. However, they indicated that it was possible to disaggregate the loan approval process into its different components for which standardized times and costs could then be estimated. Something like a sub-mask in which cost components may be added by mouse click might therefore be a useful addition to the software. Human effort enters via the actual staff time (and other resources) used. However, the relatively undifferentiated nature of estimated administrative costs is actually not such a great concern in business-lending, because administrative costs tend to be small relative to the expected loss. The last cost component, the (economic and/or regulatory) capital charge, is sufficiently important to warrant a detailed discussion.

7.3.4 Economic capital charges

The Federal Reserve System Task Force on Internal Credit Risk Models (1998) notes that

'Conceptually, the economic capital allocated to a particular activity should be measured as that activity's marginal contribution to the portfolio's overall economic capital requirement, taking into consideration diversification effects between that activity and the rest of the bank. Operationally, this marginal contribution would be calculated as the economic capital allocation for the entire portfolio less the hypothetical capital allocation for the portfolio excluding the activity of interest.' (p. 8)

The calculation of that marginal contribution is computationally highly involved. The larger the portfolio, the more complicated the computation and the less – all things equal – the gain from further diversifying the portfolio. Issues of database infrastructure may make it impossible to compute a capital charge without actually booking the exposure. Also, an offer is typically extended to the prospective borrower and kept open for some time. During that period the portfolio composition might shift and the estimated capital charge may no longer be sufficient. The shortfall may then either have to be absorbed by the bank/relationship manager or the risk has to be passed on to the borrower.

In the VR-Control framework, capital charges are not estimated as a function of an exposure's marginal contribution to actual portfolio credit risk ('unexpected loss' / Credit Value at Risk or CVaR), but via the so-called 'multiplier approach' where the capital charge is determined by the expected loss[10]. In other words, diversification, concentration, and size effects are not priced directly into the loan and have to be transmitted to the front-line bankers by way of guidelines on which industries are desirable additions to the portfolio and which are not. As long as all front-office bankers are kept informed about the bank's portfolio strategy in a timely manner, this need not impair the operation of incentive schemes. From the point of view of incentive scheme design, however, it would be preferable if the 'real' economic capital cost of a loan were considered in the estimation of expected profit. This would strengthen the allocative role of prices and obviate the need for additional guidelines. One argument against pricing the actual marginal contribution to portfolio credit risk into a loan is that this not necessary when credit risk can be easily traded. The loans under consideration, however, are non-tradable. Another argument that came up in discussions but the relevance of which I have not been able to establish is that the portfolios of many of the cooperative banks are so small that even small changes to portfolio composition can affect the marginal contribution of a given exposure to portfolio credit risk considerably. This would be a concern if the fluctuations were due to the portfolio credit risk model not working as accurately as one would want it to be for small portfolio sizes.

No matter whether the multiplier or a full-fledged portfolio approach is used, given the characteristics of the borrower and the existing portfolio of loans, the banker does not have any influence on the capital charge which in either case is determined by a computer. It was, however, pointed out to me by one interviewee that deciding on the industry that a borrower is in is not always straightforward and that there may be cases in which a banker has discretion to change an industry classification to one that attracts a lower capital charge.

This completes the discussion of the pricing and risk rating methodology. I have shown in which ways there is scope for human influence and discussed some of the implications that this might have for incentive schemes. More fundamental technical issues are discussed in a separate section below.

[10]Controversially, there is no provision in the New Capital Accord for rewarding banks with well-diversified portfolios with lower capital charges.

7.4 Findings in detail

In this sub-section I discuss possible objectives for incentive schemes. Following the introduction, I describe some desirable features of incentive schemes and give some institutional and organizational background. The discussion of possible objectives for front-office employees (relationship managers) begins with comments on the use of profit targets, followed by the analysis of several other candidate objectives. The risk objective must be at the heart of any lending-related incentive scheme and is discussed in detail. The sub-section ends with a discussion of objectives for credit analysts and the turn-around/liquidation unit.

7.4.1 Introduction

My interviewees were by and large in favor of explicit incentive schemes. Several banks were already using incentive arrangements of some kind or other and the experiences with such schemes differed greatly. Sometimes employees were reported to be entirely unresponsive to financial incentives; in another case, a bank had been able to move to a compensation structure centered on commissions rather than salary and was very successful. Time and time again, the role of 'personality' emerged as a theme. It was asserted that there is a 'salesperson personality' that one either has or does not have, but that cannot be 'created' through any amount of training or a very cleverly designed incentive scheme. As a general rule, older employees were described by human resource managers as finding the transition to contingent pay more difficult – maybe the result of a certain complacency brought about by many years of relatively easy profits and steady, seniority-driven increases in pay. The emphasis on personality might be interpreted as support for the sorting function of incentive schemes [stressed, for example, in the famous Safelite paper by Lazear (2000b)]: Maybe the real impact of contingent pay arrangements is not (or not in the main) in giving employees incentives to exercise more effort, but in leading to desirable turnover in the workforce in that low-productivity employees (assumed to dislike contingent pay) leave and are replaced by high-productivity employees (supposedly in favor of contingent pay)[11]. I find the idea that whether an incentive scheme 'works' or does not 'work' might depend more on some innate traits of the employees affected than its design or the way it is implemented very

[11] In a similar way, one might think that a company not offering contingent pay might suffer the exodus of its most productive employees. My interviewees did not suggest that this was a problem, but did sometimes indicate that the absence of incentive pay was perceived as unfair by some high-productivity employees.

interesting with potentially far-reaching implications.

7.4.2 Desirable features of incentive schemes

Above I outlined a number of (from a regulator's point of view) desirable features of lending-related incentive schemes: They should permit an evaluation prior to the maturity of a loan while still giving incentives for post-disbursement monitoring; they should be able to deal with staff turnover; and performance evaluation should be objective. In the following I will outline several further desirable features; the first two are drawn from theory (and common sense) while the importance of the remainder was stressed by the practitioners I interviewed.

Two additional key criteria can be deduced straight from standard principal-agent-theory:

1. Objectives should align the incentives of agent and principal. This is relatively straightforward for front-office employees, but more difficult for credit analysts.

2. To the extent possible, the objectives used for evaluation should be under the control of the employees concerned and should not be subject to exogenous influences.

Baron and Kreps (1999) make the case for external consistency of human resource management policies (with the 'Five Factors': the external environment, the workforce, the organization's culture and strategy, and the technology of production; see Ch. 2), as well as internal consistency (Ch. 3). My contribution focuses on incentive arrangements, and I will not discuss human resource management practices in as encompassing a way as Baron and Kreps do. The interested reader is referred to the discussion in their book. I would like to highlight that:

- I focus on one possible element of human resource management policy, namely incentive arrangements, and on its fit with the technology of production (rather than all five factors considered by Baron and Kreps);
- I see no reason a priori why contingent pay should not be expected to be effective in the banks that I visited, although occasionally interviewees noted that there might be tension between the consensual tradition in the cooperative sector and pay arrangements that focus on the performance of the individual;
- Consistency may be an issue in developing incentive arrangements for front-office and back-office. The front-office's primary role is to bring in business; the back-office's role is to protect the bank against unwanted risk. The conflict inherent in these objectives is in principle desired; however, it is intuitively plausible that if both sides had very steep incentive

schemes based on their respective objectives, gridlock might result.

I do take into consideration the cautionary comments in Baron and Kreps (1999) on pay-for-performance (Ch. 11; especially, pp. 256–258 and pp. 268–271). Of Baron and Kreps' 'non-economic caveats' (pp. 256–258), only one (fairness) was brought up in the interviews, which may have something to do with the fact that the use of contingent pay was not widespread in the sample banks and that – with few exceptions – any resulting differences in income are small. Of the five arguments that Baron and Kreps state against pay-for-performance, the first two may have 'technical' remedies. These candidate solutions constitute an important part of my findings.

1. Misalignment of incentives / multi-tasking considerations / 'gaming': It may be possible that through judicious combination of the objectives, the right balance can be attained. I will make it clear in the discussion where candidate objectives might lead to dysfunctional reactions on the part of employees.

2. Employees are made to bear excessive risk: This point will be addressed in detail below. Given the cyclical nature of the banking business, it is a key concern.

With respect to the following three arguments, I can only stress the importance of communication. It may be that a scheme that has been developed in good faith, with employee input and participation at all stages of development and implementation, and that is being proposed by a management that commands the trust of its workforce is able to overcome these issues[12].

3. Incentive schemes might not be considered legitimate in the eyes of employees.

4. Incentive schemes may be adjusted inefficiently rarely for essentially 'political' reasons.

5. The 'extrinsic motivation' aspect of pay-for-performance arrangements may undermine intrinsic motivation.

The last point is one so controversial that I refrain from further discussion here and take it up again in the section on fundamental doubts with respect to contingent pay arrangements in Eggenberger (2006c).

Lastly, I outline some of the features that the practitioners found desirable. They were strongly in favor of objective evaluations, although it was frequently suggested that complete objectivity was not feasible and that some subjective component should be retained so that special

[12]Empirical work suggesting that there is scope for overcoming self-serving bias is described in Babcock and Loewenstein (1997), and I would expect their findings to apply to negotiations on incentive arrangements.

circumstances could be taken into account. Objectives should be such that an employee at any time can accurately determine how he or she is doing at negligible cost and should be transparent throughout the bank; all levels of management should be included (i.e., evaluated according to clear criteria). Fairness (as mentioned above) is an essential feature and economists would do well to remember that fairness is what is perceived as fair. There should only be a small number of objectives, about three to five.

7.4.3 Some organizational and institutional background

In the market segment that I am considering, one relationship manager is assigned to one customer – that banker is the point of contact for all requests. One front-office employee typically works together with a number of different credit analysts on different credit files (and conversely), but because of the high fixed (and sunk) cost of understanding a customer relationship, the customer's business and accounts, the initial assignment of relationship manager and credit analyst will typically not be changed[13, 14]. The number of customers assigned to a relationship manager in this way typically varies from approximately 30 to 250 and depends on such factors as size, complexity, and risk. All business transacted with a customer will typically be credited to that customer's relationship manager, regardless of his or her actual involvement. This is a concern with respect to incentive schemes: Consider, for example, the case where the relationship manager needs to bring in a colleague (a specialist in some area in which the relationship manager's expertise is limited). If the revenue is credited only to the relationship manager, the specialist has less-than-optimal incentives; if the revenue is credited to the specialist, the relationship manager might attempt to avoid bringing in the colleague[15]. One possible technical solution to this problem would be the internal billing of services between units (profit-centers).

[13]Having said that, one could imagine that employees would try to take advantage of staff turnover to get rid of difficult or time-consuming clients by moving these from their own portfolio to the portfolio assigned to the then-vacant position. Should this happen to any significant extent, this would have potentially serious implications for the fairness of contingent pay arrangements.

[14]Board of Governors of the Federal Reserve System (2003b) discusses possible alternative organizational arrangements for the cooperation between front- and back-office with respect to risk rating (pp. 65–66).

[15]Interestingly, in one of the banks outside the sample that I talked to, allowing both specialist and relationship manager to credit the full amount of revenue to their respective accounts did not solve the problem: Relationship managers were still very reluctant to bring in someone else, maybe for fear that the client might contact the specialist directly on the next occasion.

This, however, is currently not yet technically feasible in my sample banks; neither are relationship managers nor the back-office organized as profit-centers. Revenues are tracked on a per-employee basis in the front-office but not costs, so transparency with respect to costs is incomplete. Economists do not usually pay much attention to issues of accounting and information technology; here, however, the possibility of billing services between profit-centers would make a material difference. Note, for example, that the cost of evaluating loan applications is considerable but is not usually charged to the applicant (by the relationship manager) nor the relationship manager (by the credit analyst). It would be highly desirable to directly charge applicants a fee for the creditworthiness assessment and the regular meetings that take place in an ongoing lending relationship and it was encouraging to see that some banks in the sample have moved quite far in that direction. Apart from the issue of incentive arrangements, there is an additional, related reason for implementing profit-center accounting. It is already possible today to outsource the back-office function for certain standardized products (primarily residential real estate loans), and it is hard to see how what are essentially make-or-buy decisions on keeping or outsourcing certain processes can be made without transparency about costs. The heterogeneity of business loans in terms of complexity plays a role in at least two important ways:

1. The output of the back-office is hard to measure, at least as long as the work is not differentiated into individual, standardized components. This is different, for example, for residential real estate loans which are highly standardized and therefore allow direct comparisons of employees' output.
2. Assuming that profit targets are to be a component of an incentive scheme, it is very hard to tailor portfolios such that no relationship manager (or credit analyst) has an obvious advantage over any other. This could be a real concern if it leads to an incentive arrangement being perceived as unfair. For the mass and mass affluent retail markets standardized tools are available that use a range of variables (from socio-economic data to commuting patterns to the bank's market share etc) to estimate the profit potential of a group of customers. Similar tools are not available for the business market. In addition, creating portfolios with approximately equal potential may not be management's first priority. Other criteria for assigning customers to relationship managers may be

- location (especially in rural areas some degree of geographical con-

centration of customers in a portfolio may be required)

- industry (not a major concern for the relatively small banks that I visited, as there is usually not that much scope for specialization by industry – however, larger banks will have industry specialists)

Sometimes the back-office will be organized around a finer division of labor: Certain tasks (such as financial statement analysis, sending out requests, compliance/reporting, collateral evaluation) are then centralized. In such cases the principles that are discussed here can be applied with some modifications.

7.4.4 Possible objectives for front-office employees

It is a key insight from the multi-tasking literature[16] that if one cannot give balanced incentives for all the separate tasks that make up a job, it may be best not to give explicit incentives at all. In addition, complexity is a real-world concern of great importance. An incentive scheme may be cleverly designed; if it is so complicated that it is not understood, it will not serve its purpose[17]. Any effort to develop explicit incentive arrangements to a large extent hinges on developing a package of objectives that covers all aspects of the relationship manager's job and gives the right weight to each. It is important to note that the credit risk aspect cannot be separated from the other tasks of the relationship manager; in particular, a relationship manager needs to bring in business. (In fact, sales and the profits they generate are a source of protection to the bank when risk is taken in a disciplined way.) Although I am looking at incentive arrangements from a regulatory point of view, it would, therefore, not be enough to ask: "How can I punish the relationship manager if performance with respect to risk management leaves something to be desired?" Any such lopsided scheme would lead to resentment on the part of the employees concerned and likely not find acceptance among the regulated institutions. There is no contradiction between looking at incentive arrangements from a regulatory point of view (where the primary concern must be downside risk, in this case, credit risk) and taking into account the profit objective of the practitioner. It is not inconceivable that such regulation leaves both regulator and bank better off; there is ample evidence suggesting that the argument "if something were profit-maximizing, companies would already be doing it" does not always apply.

[16]The classic reference is Holmstrom and Milgrom (1991). Baron and Kreps (1999) provide an excellent discussion (pp. 251-253, 268-270).

[17]For a real-world example of how this happened at Analog Devices, see Baron and Kreps (1999) (pp. 221–222)

The main objectives that I will consider in the following are:

- Profit objective
- Customer satisfaction
- Risk objective

Secondary (or possibly temporary) objectives are:

- Borrowers' financial statements submitted in a timely manner
- Reduction in overdrafts
- Number of submitted applications complete
- Announcing intention to leave position as early as possible

These seven candidate objectives emerged as potentially useful objectives in the interviews.

Possible objectives for front-office employees: Profit objective

To an economist, using the profit a front-office employee generates (by summing up the 'profit contributions IV' for that employee's sales) as one key objective seems an obvious choice: If the accounting infrastructure is in place, why look at anything other than profits (which should, if measured correctly, align the interests of the bank or regulator and the employee)? When a pure profit objective is used, the employee does not require any additional guidance (except with respect to portfolio considerations as discussed above as long as these are not yet decentralized and automatically taken into account in the bank's risk management and pricing software) and has the right incentives to negotiate well on behalf of the bank. All information is assumed to be conveyed via the 'profit contributions IV' as estimated, and the front-line banker is given the freedom to decide which products to sell. He or she would then rationally choose to concentrate his or her sales efforts on those products that bring the highest profit (given the time required to sell them)[18]. A profit target also would give the relationship manager an incentive to price a lot of services that are provided for free today: Loans, for example, are typically bundled with the creditworthiness assessment, ongoing monitoring and financial advice and so on. The more innovative banks in the sample have started billing for the evaluation of loan applications and the regular consultations separately (and have noted that relationship managers are much better prepared for their meetings when they know that they have to send the customer a bill afterwards). Whereas currently banks typically make binding loan offers to the applicants that expose the bank to adverse term structure movements, a profit objective

[18]The experiences of banks that pay commissions for the sale of certain products but not others bear out that this is what would happen in practice.

would give relationship managers an incentive to pass that risk on to the prospective borrower wherever possible. The question arises why the use of profit objectives could ever not be optimal from the point of view of the bank.

- The profit estimates might be biased. Below I will outline concerns with respect to credit risk measurement in a separate section. Here I wish to discuss the implications of using a profit objective when profit estimates are obtained in present value terms. While this ensures apples are compared with apples, to arrive at present value estimates for certain products, a number of behavioral assumptions need to be made. If one cannot not have much faith in these, differences between actual and estimated profits might be large and systematic. Examples of product sales requiring such assumptions to be made are loans with option elements (Will options be exercised and in which circumstances?) and demand deposits (How long will the money be held at the bank?). In the case of long-term loans, one also needs to estimate the administrative costs associated with future risk ratings, financial statement analysis, discussions with the borrower (if these are not billed separately) many years out into the future, when there is actually little transparency with respect to these costs.
- Certain externalities might arise that are not priced[19]. Balance sheet structure, for example, might be a concern if estimated profitability gives misleading signals. For example, the bank may be reluctant to lose savings and time deposits to aggressive direct-banking competitors (paying above market rates trying to build market share) even though it would be cheaper to switch to wholesale funding rather than to match the competitor's rates. In this example, deposits are assumed to remain a cheap source of funding in the long-run. Cross-selling considerations ("If the time deposit disappears, then so will all the other business that the customer is doing with us: insurance, credit card etc.") may be playing a role, too. By using appropriately calculated subsidies, the front-office banker could, in this example, be insulated against rate fluctuations. However, 'fiddling' with prices is problematic. At any rate, one would want to see any decisions taken on what would be de facto subsidies to be based on facts and data rather than rules of thumb and gut feeling. Another illustration of an externality are customers whose direct profitability does not reflect their indirect profitability to the bank (e.g., a close relative of a very important business customer). Under a pure profit

[19]Most of the 'externalities' that I came across, however, are likely side-effects of any explicit incentive scheme based on individual performance and so it is not the use of a profit objective as such that is the problem (see below).

objective, an employee has an incentive to not spend time on or even get rid of all unprofitable customers unless he or she is compensated in some way.

- Profit targets possibly lead to inefficiently high degrees of specialization. If product sales require a certain expertise, an employee might find it in his or her best interest to neglect certain products and concentrate on those he or she already knows well. This is a real concern in banks which would prefer to have well-rounded relationship managers able to give advice on a wide range of products. Over time, the loss of a reputation for giving good advice might constitute an externality with tangible costs to the bank.
- Commissions paid by insurance companies are typically paid according to a non-linear schedule (the more policies are sold, the higher the commission per policy). This is difficult to capture via the accounting software.
- Even leaving credit risk concerns aside for now, the use of market-driven prices exposes the employee to a lot of risk because of interest rate fluctuations. Incentive schemes might need to be adjusted frequently and any such changes are problematic. This is one short-coming of a profit objective that would not disappear using quantity targets instead.
- The consistent use of profit targets has important implications for other areas of bank management. For example, it might render a preferred marketing tool, campaigns that heavily promote and 'push' certain products, ineffective. Traditional 'loss-leaders' (such as, in the case of my sample banks, certain insurance products) would likely also disappear unless a relationship manager found it in his or her interest to sell the product and incur the associated loss. Pure profit targets also are not easy to reconcile with current planning policies that involve quantities. Asset growth is still considered an objective by many bank managers (as is growth in the number of customers), but is a fairly meaningless concept in a profit-maximization framework. These illustrations highlight the need for incentive arrangements to be consistent with other policies of the bank.

A relationship manager operating almost independently as a quasi-entrepreneur marks one extreme; at the other extreme he or she might be given detailed quantitative targets for the sale of certain products[20].

[20]If the employee were paid exclusively on a commission basis, one might indeed ask why he or she still is an employee of the bank; to all intents and purposes the employee is an independent entrepreneur. Implicit in my discussion is the use of contingent compensation as a more or less important supplement to a salary. The 'pure commission' model is not as far fetched as it might sound; many mortgage originators in the United States are paid this way (see section 4.4).

Use of quantitative targets would solve some of the problems outlined above. However, such an objective is far more complicated than a simple profit objective; one would need to weigh different components and determine the consequences of not meeting the target for individual products. (Product quotas can have their own pernicious consequences: For example, they give employees incentives to sell certain products no matter what in order to fill the quota.) For sales above target, one would still need to tell employees how much the extra sales are worth in terms of additional pay. In the highly heterogeneous business market, it would be especially unhelpful to give bankers binding targets; also, the bankers in this area are the most experienced in the bank.

While in theory it might be possible to combine prices and quantities as tools for managing the front-office, it is my impression that any resulting packages of objectives for employees are likely to be too complex. On the basis of the interviewees' support for a profit objective, my impression is that a profit objective, if necessary supplemented with certain additional elements, is the most reasonable choice. However, this is not a question that can be decided in the abstract. The point of the discussion here is to outline various concerns that a regulator thinking about making incentive arrangements would need to consider.

In whichever way the bank's need to generate profits is transformed into an objective for the front office, the bank needs to stick to it consistently for a set, pre-announced period of time (typically a year or maybe three months during the trial phase). The consistency requirement underlines the extent to which well-designed objectives discipline not only the employee, but also top management. A number of issues (for example, the value of cross-selling opportunities when 'loss-leaders' are used) would need to be thoroughly analyzed. Ad hoc changes to objectives will undermine acceptance of such a scheme; so will ratcheting up the objectives after the trial phase without a convincing explanation and renegotiation. Fiddling with prices or quantitative objectives will likely be highly demoralizing when this is to the disadvantage of the employee. In fact, constant changes to prices will be confusing for employees regardless of whether these are in their favor or not[21].

A few words on risk are called for within the profit objective context. Credit risk will be discussed below in connection with the risk objective, but as noted above fluctuations in market rates will have an impact on front-office employees even when the interest-rate risk for every transaction is hedged. Agency theory suggests that the risk of term structure

[21] An increase in price will make it more difficult to sell the product, but also increase the profit.

movements and the associated fluctuations in profitability should be assumed by the bank, but how to separate the effects of changes in interest rates from the effects of employee effort is still not straightforward. The standard solution for dealing with risk, relative performance evaluation, has its own problems [see p. 253 in Baron and Kreps (1999)] and, in the case of the banks that I visited, can be ruled out as a sensible approach: While relationship managers work largely on their own, they rely on the cooperation of their colleagues to a sufficient degree for relative evaluation to be very problematic. This may be different for large banks.

This issue leads to the question of how high a target should be set. Using a profit objective it would be reasonable to demand from a front-office employee that he or she earn all fixed costs (basic salary, benefits, office space, other overheads etc) before a bonus is awarded[22]. While that may be a good rule of thumb, certain adjustments need to be made under an accounting approach where all profits from a transaction are credited at once. The main reason for this is that under that system, a borrower in the portfolio (who might, say, have taken out a ten-year loan two years ago) makes a lot of work (that the relationship manager does not, in most banks, bill to the customer) and does not necessarily generate additional profits. In practice a relationship manager will normally be able to generate substantial additional sales from an existing customer, but not adjusting for the workload that comes with existing customers might still be problematic. It also might be unfair to someone taking over a portfolio to not have as much time for generating new business as his or her colleagues. A more sensible solution than adjusting targets would be to de-bundle the different services and sign separate contracts with the customer (application fees, the loan itself, monitoring fees), but even this would not help if all revenue components were credited at the moment a deal closes. As was repeatedly pointed out in the interviews, gauging the sales and profit potential of a portfolio of business customers is difficult.

Certain problems, driven primarily by such a profit objective, are inherent in the implementation of explicit incentive schemes focusing on individual performance:

1. One important issue revolves around the tradeoff between current and future profits, and its impact on customer satisfaction and loyalty. (Are product recommendations driven by the customer's needs or by the

[22]The bonus could be a fixed percentage of all profits above the threshold. Such a linear incentive scheme – similar to piece rates – has the advantage of great simplicity. The classic reference in the theoretical literature dealing with linear schemes is Holmstrom and Milgrom (1987), although it is not exactly a practitioner paper.

profit they command? Do front-office employees become too aggressive and alienate customers?) Interviewees repeatedly related first-hand experiences (not necessarily from the sample banks) with salespeople who sold very aggressively and moved on to another bank before the damage they had done to customer relationships became apparent. This issue was considered so important that the interviewees thought it appropriate to counterbalance the profit objective with a customer satisfaction objective (see below).

2. Another side-effect of any performance evaluation that focuses on individual performance is a reduced willingness to help colleagues and be a team-player. Interviewees consistently indicated that correcting for this with an explicit teamwork objective (everyone evaluates everyone else) might be possible, although any such bureaucratic device should clearly be a last resort. The issue of teamwork has particular relevance in the present context, as transactions beyond a certain size and risk require the approval of a banker's superior. Since it takes much less time to reject a loan application than to carefully review it, supervisors may have an incentive to reject inefficiently frequently. A solution to that problem might be to tie the reviewer's remuneration to that of the more junior employee.

3. The more senior the position, the greater the extent of managerial responsibilities. The performance of an employee as a manager of people is inherently difficult to evaluate objectively. One possibility, again, is to tie a manager's remuneration to that of his or her staff. However, depending on the nature of the contributions expected from a manager, employees' remuneration may not be a very good proxy for the quality of managing and subjective elements may need to be added.

4. Organizational arrangements may come under pressure. In several interviews it was noted that in a setting where employees are paid a fixed salary, the tolerance for unclear responsibilities and areas of authority, an illogical division of labor and other organizational inadequacies will be greater than in an environment where dysfunctional organizational practices have direct adverse consequences for remuneration.

Team profit objectives are likely essential for branch staff, but the consensus among the interviewees was that the work of the relationship managers under consideration here is not so interlinked with that of their colleagues that a team objective would make sense.

The bottom line is that an objective linking a relationship manager's remuneration with the profit he or she generates for the bank is conceivably a useful objective from the point of view of both bank and regulator.

However, many practical issues would need to be resolved. Some of these (such as the heterogeneity of business customers that makes it difficult to design comparable portfolios for different relationship managers and may, therefore, undermine the fairness requirement) are mainly a headache for the bank. Other problems (such as possibly inaccurate profitability estimates) are more directly a concern from the regulator's point of view. On the basis of my exploratory interviews it is not possible to gauge the extent to which such difficulties make it inadvisable to use a profit objective (or, indeed, any of the objectives discussed below).

Possible objectives for front-office employees: Customer satisfaction

The use of explicit performance evaluation that focuses on the individual may make it necessary to forestall overly aggressive selling. One way of doing this would be to regularly contact the relationship managers' key contacts on the customer side and ask them for a detailed assessment of the relationship manager according to a standardized list of questions. Many banks have call-centers that could perform this task during quiet periods; in view of the relatively small number of customers in the business segment the cost of such assessments was considered by all interviewees with whom I discussed this issue worth incurring. This kind of customer survey is not carried out systematically at present; one could use the results of the assessment not just for remuneration and coaching purposes, but also for broader business-development efforts. The relevance of such an objective from the point of view of a bank regulator is not obvious, but it is possible that customer complaints about aggressive sales tactics are of value as an early warning sign.

Possible objectives for front-office employees: Risk objective

The purpose of this objective is to give relationship managers an incentive to devote an efficient amount of effort to the credit risk aspects of the lending relationship, including thorough risk assessment; careful monitoring of borrowers and good preparation for follow-up meetings; and expeditious signaling of developing problems.

One objective that I considered but rejected were assessments by the auditors. I learned in the interviews that auditing is far from an exact science. In a number of banks, interviewees described that they had been audited twice by different sets of auditors during a year (statutory annual audit plus an audit ordered by the financial regulator at random on a sampling basis) and that different auditors came to different con-

clusions when looking at the same credit files. Building audit results into relationship managers' compensation would likely only lead to endless discussions first with the auditors and then with bank management. Auditors also concentrate on only a subset of (particularly large and/or risky) loans and would likely not review a good-sized sample of loans for each and every relationship manager. This problem would be less severe in case internal auditors' opinions were used, since management has more leeway in getting internal auditors to widen the scope of their investigations. However, using internal auditors' assessments as the basis for performance evaluation still runs the risk of turning the exercise into a blame game. Having said this, the out-of-sample bank previously mentioned which uses the assessment of relationship managers' risk-related work by credit analysts for compensation purposes reported that this approach was working fine. Because of the subjective character of the performance signals generated in this way, I will not discuss this evaluation technique further here. If it could be shown to work more generally, it would be an interesting and pragmatic idea (although, again, difficult to implement as a regulatory instrument).

One related objective that one might consider are what could be referred to as 'autopsy results' where defaulted loans are analyzed carefully by an independent unit in the bank to see if the default can be attributed to any material errors on the part of employees[23]. The consensus among my interviewees was that clear-cut responsibilities can be established in only relatively few cases. A loan loss is not usually attributable to a single flawed decision and any flawed decision usually cannot be attributed to any one clearly discernible mistake on the part of a specific employee (although there are exceptions: missing signatures on key documents, guarantees not renewed after a certain number of years). While I believe the systematic analysis of defaulted loans would be helpful regardless, using such assessments for performance evaluation purposes has the critical disadvantage that mistakes are only detected if and when a borrower defaults.

In the following, I discuss a more promising indicator of the quality of the relationship manager's risk-related work that could be said, with some simplification, to compare actual loan losses on a portfolio with expected losses. I will consider a measure that I term 'adjusted gross loan losses' (the gross loan losses are adjusted for migrations, changes in collateral value, and other variables) which has the advantage that

[23]To some extent, this kind of analysis of why and how things went wrong occurs today when a credit file is transferred from the front-office to the work-out unit or liquidation.

changes in credit quality short of default are considered[24] and that a fairly continuous, at the very least yearly, assessment is possible since all borrowers have to be re-rated at least yearly. The relationship manager's work on credit risk aspects is reflected in the risk rating of a borrower (albeit often with a considerable delay of maybe a year or two because fresh financial statements come in yearly intervals); using the measure defined below, an improved risk rating would pay off for the relationship manager in terms of remuneration, as would an increase in collateral value or additional collateral and guarantees. As I will outline below, the scheme may also give relationship managers incentives to signal problems early on. 'Adjusted gross loan losses' would be computed as follows:

1.
   - Reduction in provisions no longer required (1)
   - plus reduction in expected losses due to increases in measured creditworthiness (2)
   - unless interest rate were to be adjusted immediately to reflect the lower risk (3)
   - plus reduction in expected losses due to additional collateral or increases in collateral value or additional guarantees (4)
   - unless interest rate were to be adjusted immediately to reflect the lower risk (3)
   - plus any other increase in risk premia due to a negotiated increase in interest rate (3)
   - minus increase in provisions additionally required and write-offs (1)
   - minus increase in expected losses due to decreases in measured creditworthiness (2)
   - unless interest rate were to be adjusted immediately to reflect the higher risk (3)
   - minus increase in expected losses due to decreases in collateral value etc (4)
   - unless interest rate were to be adjusted immediately to reflect the higher risk (3)
   - minus any other decrease in risk premia due to a negotiated decrease in interest rate (3)
   - ='Adjusted gross loan losses' (on relationship manager's portfolio)

[24]This is referred to as the mark-to-market paradigm for traded credit risk and mark-to-model paradigm for non-tradable risk.

For now I will assume that all the items listed above refer to flow figures for the evaluation period which I assume to be one year. I also assume that there are no loans in the portfolio at the beginning of the year and that all loans are made at the beginning of the year for a term of exactly one year. This avoids certain complications which I deal with below.

1. Provisions and write-offs are the main credit-risk-related costs. As the creditworthiness of a borrower declines and the loan becomes doubtful, first a provision is created for the amount of the likely loss. If the actual loss is less than expected, provisions can always be cancelled. If the work-out unit does good work and manages to nurse financially fragile borrowers back to health, the provisions for the likely losses on these borrowers can be credited back (minus the costs incurred by the work-out unit, see below).

2. Changes in measured creditworthiness are primarily changes in the probability of default. It is quite clear that the expected loss on a loan goes up when the probability of default goes up and conversely. The basic assumption behind including these migrations in the above sum is that a monetary equivalent to the change in creditworthiness can be computed. As I will argue below, estimating these monetary equivalents (which are essentially expected losses or, more accurately, differences in expected losses) is fraught with difficulty, but cannot be avoided if one is to take into account credit migrations short of default.

3. Current practice among my sample banks is to fix the interest rate for the term of the loan in the case of a fixed-rate loan. Where a loan is at variable rates, the benchmark rate is typically a market rate (say, LIBOR plus X basis points). This means that changes in credit risk are not taken into account even in the case of loan agreements with very long terms. Loan rates are in principle not renegotiated even when the creditworthiness of a customer has decreased substantially. In practice, it may be possible that individual relationship managers succeed in persuading a borrower that the contractual interest rate is no longer appropriate, but it is not standard procedure to regularly revise interest rates. Interestingly, I have found almost no references to regular risk-related adjustments of interest rates in the literature[25]. This may be because loan covenants in many or even most cases allow the bank to call in the loan if the financial condition of the borrower deteriorates. At any rate, I believe that in view of the possible shortcomings of current risk rating technology which I discuss below, regular rate adjustments

[25]Federal Reserve System Task Force on Internal Credit Risk Models (1998) mentions the practice of 'grid pricing'(p. 23, FN 22).

should at least be considered (even though their aggregate pro-cyclical effects could be considerable).

4. It is quite clear that changes in the value of collateral and guarantees affect the expected loss to the bank. A relationship manager can directly 'work' on these by convincing a borrower to put up more collateral or to find additional guarantors.

My sample banks' accounting and IT systems are able to calculate the above measure with some minor manipulation; this is essential for practical implementation. I would think that computation of (some variant of) the above measure should not be a problem for other banks either. The measure as such is not terribly helpful - incurring loan losses is just a normal part of the banking business, after all. Some benchmark is required that allows an assessment of whether the 'adjusted gross loan losses' on a relationship manager's portfolio are 'high', 'too high', or 'low'. The obvious candidate benchmark is the expected loss on the portfolio (calculated as the sum of the risk premia priced into the loans); the idea being that a relationship manager with lower loan losses than expected is doing a good job. Using the difference between the above measure of actual loan losses and expected losses to assess the quality of a relationship manager's risk-related work raises a number of questions which I address next.

While the relationship manager does have some influence on this measure, surely exogenous risk factors beyond the control of the banker play an important, maybe even dominant role. This, of course, is the classic risk-incentive trade-off that is at the heart of the principal-agent literature. As stated previously, relative performance evaluation cannot be used as a remedy in the (small) banks that I visited, although larger banks might consider comparing bankers that do not need to cooperate with each other (as long as the risk factors affecting them are the same or at least highly similar; this may not be the case across regions and industries). The crucial question with respect to the above candidate objective is whether it is possible to sufficiently isolate employee performance from exogenous fluctuations. I cannot give a definitive answer on the basis of my interviews, but I do outline below some relevant findings.

The expected loss priced into a loan is estimated on a stand-alone basis. On the portfolio level, the expected loss is equal to the sum of losses expected on the individual loans, but the portfolio loss distribution cannot be obtained by modeling the performance of the component loans as independent random variables. Rather, defaults are correlated (over time and cross-sectionally) in complicated ways and estimating these

correlations is one of the key challenges for credit risk modeling[26]. The sources cited above attest to the difficulty of the task. In economic terms, one way of looking at the correlations is to think of these as being driven by exogenous risk factors (such as exchange rates or prices of key commodities) that affect all borrowers simultaneously.

These exogenous risk factors have important implications not only for the unexpected (portfolio) loss distribution, but also for the estimation of expected loss. In my sample banks expected losses (risk premia) are invariant to exogenous risk factors. In other words, the expected loss priced into each loan as a risk premium is calculated under the implicit assumption of exogenous factors (as well as the relationship manager's effort level) taking some 'average' value. The assumption is a merely implicit one as the expected loss is obtained as a historical 'average' based on data from a sample period spanning several years (ideally at least one full credit cycle) during which one would expect the exogenous variables to move around quite a bit. Exogenous risk factors not entering the estimation explicitly means that expected loss is not conditioned on, for example, the business cycle. It is theoretically possible to estimate conditional transition matrices (which describe the migrations of loans between risk classes) and use these for pricing. To the best of my knowledge, such work is being done by academics [see, for example, Bangia et al. (2002)] but is not (yet) routinely applied by practitioners.

Given the definition of expected losses as an unconditional historical average, there is no reason for expecting actual losses to be close to expected losses during any given year. In a real-world setting, one would typically see lower-than-expected portfolio losses in six or seven out of ten years and higher losses in other years. These deviations of actual losses from expected losses are the unexpected losses that the bank holds capital against. One should expect this phenomenon to be even more pronounced at the level of a single relationship manager's portfolio at which the above objective proposes to measure performance. The reason for this is that a single relationship manager's portfolio is of necessity less diversified than the entire bank's loan book.

Clearly, the relationship manager's monitoring effort shifts the loss distribution [see Carey (1998)], but that observation is, of course, far too general a basis on which to construct an incentive scheme. In theoretical work one can, loosely speaking, map observed outcomes back into effort

[26]Treacy and Carey (1998), for example, note that 'For many firms, industry supply and demand cycles are as important or more important than the overall business cycle in determining cash flow.' (p. 899) Also see Wilson (1998) who discusses the role of macroeconomic influences.

by asking: Given the realization of, say, output, how likely is it that the agent has chosen the desired effort level? However, our understanding of real-world credit portfolio loss distributions is far too rudimentary (and data too scarce) to be able to translate this idea into practical incentive schemes. To recapitulate, one would ideally want to know the portfolio distribution of losses conditional on the values taken by exogenous risk factors, and one also would like to know how effort shifts the distribution. Such estimates are to all intents and purposes unobtainable. Comparing the above measure of actual losses against expected loss without controlling for exogenous risk factors thus exposes the relationship manager to a lot of risk. Put simply, when the economy booms, actual losses will be much lower than expected losses; the converse is true when the economy slows down. Agency theory and common sense suggest that incorporating such an objective into an incentive scheme is problematic.

However, my interviews have made me less pessimistic and suggest that these concerns can be partly defused with low-tech workarounds and rules-of-thumb which may not convince a theorist, but may very well be workable in practice.

A possible solution to the problem is to adjust the target by negotiation between management and employees ex ante, for example at the beginning of the year, or even ex post. It is important to remember the research question: "What scope is there for devising effective lending-related incentive arrangements in practice? Is it possible to construct incentive schemes that align the interests of a regulator with those of the bankers concerned?" The purpose of the interviews was not to find a solution that would satisfy a theorist, but rather to find one that stands a chance of working in practice and could be imposed by a regulator. The bankers I interviewed were not so concerned about the exogenous (business cycle, industry cycle) risk as to think the above risk objective unworkable. They pointed out that banking is a cyclical business and that this is known to employees. Sometimes exogenous factors are favorable and sometimes they are not. As long as the risk to an employee's remuneration is limited, shifting some amount of risk from the bank to the employee seems acceptable. No matter what the overall business environment, relationship managers and credit analysts were said to have considerable scope for working on risk-related issues. With a few exceptions it was felt that the gain from having a risk objective in place would still outweigh the complications due to any 'unjust' performance evaluations that might result from the use of a relatively crude scheme. While attempts to isolate the impact of exogenous risk factors using econome-

ric tools are not likely to find much acceptance (because one would likely need to be a fairly good econometrician to understand the procedure), simple rules of thumb agreed on ex ante or even ex post could go a long way towards sheltering employees from exogenous risk factors and make a risk objective acceptable to employees. These risk factors will vary for different portfolios, but plausible proxies for exogenous risk include the GDP growth rate or the output gap, the number of insolvencies, and other such easy-to-understand variables[27]. For example, the appropriate value of the target might be negotiated in light of the GDP forecast (with a provision that renegotiation is possible in the case of very strong deviations of GDP from the forecast). While GDP growth forecasts are notoriously unreliable for a period of more than a few months ahead (see the discussion of risk measurement tools below), acceptance may well be more important than accuracy. Assuming that the historical loss experience is a good guide to future losses, one would then expect cumulative expected losses and actual losses to be roughly equal over a full credit cycle; however, a full credit cycle is too long a time frame for compensation purposes. (Note that one of the key requirements was that at least annual performance evaluations should be possible.)

It is possible that there will be a move toward the use of conditional transition matrices for risk rating and pricing as credit risk tools improve and more high-quality data become available. A scheme that uses the (unconditional) expected loss on a portfolio as a benchmark as discussed here could at the least be used as a transitional arrangement that serves more to create awareness on the part of employees than to determine a substantial part of compensation.

In sum, a risk objective worth considering that found support among the interviewees (its shortcomings notwithstanding) is to consider the difference between the expected losses (risk premia) that a relationship manager has priced into the loans in his or her portfolio and 'adjusted gross loan losses' as a measure of the relationship manager's diligence with respect to credit risk-related work. While I wish to focus on the

[27] As far as German data are concerned, when I analyzed the numbers I found that it is not straightforward to detect a negative relationship between banks' aggregate loan losses (the accounting term for this is Risikovorsorge in official statistics) and contemporaneous GDP growth or GDP growth lagged by one year. This seems to reflect accounting issues rather than the lack of a relationship between the business cycle and loan losses. Risikovorsorge is a net term that adds up various figures that are heavily influenced by management discretion. To the extent that banks operate in other jurisdictions the business cycle of which is not correlated with the domestic business cycle, one would expect this to smooth the aggregate figure somewhat. Also, the Risikovorsorge figures are available only yearly, so the timing and duration of business cycle movements might distort the economic relationship between defaults and macroeconomic variables further.

conceptual issues and do not wish to make any suggestions as to how a good performance with respect to this objective should be translated into a bonus, I do wish to stress that it is not realistic to expect the measure to take a value around zero and that one should adjust the target in the light of exogenous risk factors in a transparent, easy-to-understand way. Very small portfolio sizes compound these problems.

Moving beyond a one-year horizon For the sake of completeness, it should be pointed out that things are complicated by the fact that in reality the loan portfolio does not start at zero at the beginning of the year. Neither does it end at zero at the end of the year. Loans are added and repaid continuously. In particular, any relationship manager will begin an evaluation period with a stock of loans and a starting balance of risk premia available to cover losses. Risk premia (expected losses on an individual loan) are computed in present value terms and cover the entire loan term. This means that in case of maturities longer than one year (the usual risk horizon and by assumption the period covered by the performance evaluation), the risk premium needs to be apportioned appropriately. In the above I had assumed that all loans have a maturity of exactly one year so that this effect could be neglected[28].

Incentives for giving 'early warning' signals As noted previously, it is very important that relationship managers signal difficulties on the part of their customers as early as possible. The evidence cited above suggests that this is a problem in many banks. Relationship managers apparently tend to ignore problems or postpone dealing with them in the hope that things will improve on their own. This may be human, but it means that borrowers' financial difficulties, manageable at an early stage, often become very serious by the time the borrower's credit file is transferred to the work-out unit and given 'intensive care'. The objective

[28] For an illustration of how it might be problematic when the term of the evaluation period (assumed to be one year) differs from that of the loan term, assume that a two-year loan has a probability of default of zero during the first year and 50 % during the second year. The relationship manager would be required to price a very substantial risk premium into the loan. The present value of that risk premium would be added to the insurance pool in year one when – by assumption – the borrower will not default with certainty. This would make the risk performance of the relationship manager appear better than it is during the first year, possibly giving him or her incentives to leave the bank at the end of year one when the bonus for the good performance is paid out. The drastic example aside, adding new loans to the portfolio will in all likelihood lead to a temporary increase in risk premia available to cover losses without a commensurate increase in defaults – the longer the term of the loan, the more pronounced this effect (which is related to the 'aging effect' discussed above). In what one might loosely speaking call the steady state, defaults on loans made in earlier years would, of course, offset the temporary upward blip in risk premia available to cover losses.

proposed here may help to give relationship managers an incentive to signal problems early on in that the resulting loss (which influences their remuneration) should be lower the earlier problems are signaled.

Transfer of a credit file from the front-office to a specialized work-out unit is not currently priced in the banks I visited; i.e., the work-out unit is not an independent profit-center and does not charge the front-office for its work. I will discuss possible objectives for the work-out unit in more detail below. Here I only wish to note that the transfer ('sale') of a loan out of the front-office should result in a stringent write-down of the loan to liquidation value (or possibly even zero) in order to have a benchmark for how well the work-out and/or liquidation units are doing and to give them an incentive to exert effort. The reduction in loan value will initially take the form of a provision. That provision can be (at least partly) cancelled if the borrower returns to financial health. Earlier signaling of problems makes it more likely that such a provision can be cancelled and should therefore be in the relationship manager's interest.

It should be noted that in such a profit-center organization where a relationship manager has to 'pay' his or her colleagues in the work-out unit to take care of problem loans, there might be an incentive for the relationship manager to try to resolve difficulties on his or her own. However, the front-office bankers typically do not have the time nor the specific skill set to deal with complex problem cases. The net effect of giving the relationship manager a specific risk objective (on the one hand an incentive to not involve others, on the other hand an incentive to involve others early on as that would reduce the expected cost of a problem loan) is hard to gauge, but my interviewees were confident that on balance the risk objective should help to get relationship managers to signal difficulties earlier.

Variable interest rate contracts As noted above, the banks in my sample currently smooth lending interest rates to a significant degree. Smoothing interest rates for a given credit risk is straightforward, as the bank can always hedge this risk in the market. However, banks also tend to 'freeze' the credit risk assessment at the moment a loan is negotiated and do not pass downgrades on to the borrower in the form of higher interest rates. Nor, conversely, does a borrower benefit from an upgrade via lower interest rates. As noted above, credit risk is a risk that the bank cannot hedge directly; it is also difficult to conceive of good 'indirect' hedges that are cheap and have little basis risk (even at the portfolio level). For example, if a bank could purchase protection

against a recession that would to some extent obviate the need for direct protection against defaults. However, the requisite hedging instrument to all intents and purposes does not exist. If it did exist and were reasonably priced, it is not clear how well such an insurance contract would allow the bank to hedge the risk that it really cares about, namely credit losses (and not the recession as such which is merely an important risk factor driving credit losses). Having said that, hedging non-tradable credit risk indirectly is a fascinating idea that cannot be discussed in the detail that it deserves here.

In practice banks self-insure against credit risk by charging a risk premium, the expected loss discussed above, to borrowers. Therefore, if the expected loss is estimated correctly, possible upgrades or downgrades should have been considered in the estimation of the required risk premium. Expected losses are estimated assuming that the historical loss experience is a good guide to the future. Given the relatively short data histories that many banks use to estimate transition matrices, that may be a rather strong assumption[29]. I believe that it would be preferable to adjust interest rates to assessed credit risk. This would effectively pass some exogenous risk on to the borrower and thereby help the relationship manager reach a risk objective as defined above[30].

'Gaming' issues As described above, (partly) subjective assessments by the relationship managers affect the risk rating of a borrower. The relationship managers (jointly with the credit analyst in charge of the customer) may modify the risk grade assignment by way of an override. Credit analysts, who also should have a risk objective, have some degree of leeway in estimating collateral values. This could be a problem for the incentive scheme, since better risk assessments and higher estimated collateral values result in a better (measured) performance and thus higher remuneration. Treacy and Carey (1998) describe some of the effects observed when profitability is risk-adjusted, i.e., measured more accurately by pricing in expected loss and cost of capital:

'Interviews indicated clearly that the introduction of risk-sensitive profitability analysis puts significant new pressures on the risk grading system. Pressure to rate loans favorably arises because expected losses

[29]For example, note that it is an empirical regularity that a transition matrix calculated for a two-year period is different from the one-year matrix multiplied with itself. However, because of a lack of data many banks have to rely on a one-year matrix to estimate transition matrices for time horizons beyond a year.

[30]However, the aggregate macroeconomic implications of such an approach are potentially quite far-reaching and are being much discussed in connection with the New Capital Accord.

and capital allocations are lower for lower risk loans. Some institutions found that many loans were upgraded shortly after the introduction of profitability analysis, although the overall degree of the shift was small. One institution specifically mentioned an upward bias of about one-half grade relative to previous rating practice. Many noted that the number of disagreements in which relationship managers pressed for more favorable ratings increased once such systems were put into place.' (p. 919)

Treacy and Carey do not state whether in their sample the more exact estimate of loan value affected compensation directly through incentive arrangements; if anything that should exacerbate such conflicts of interest.

In defense of the measure proposed, the scope for relationship managers to affect the risk assessment is limited in the banks I visited; arbitrary assessments would likely not pass basic plausibility checks. The relationship managers' work in this respect is checked by the credit analysts, their supervisors, and external as well as internal auditors. While in principle the remuneration of the supervisors and credit analysts also might benefit from optimistic risk assessments, at least on the part of the credit analysts a strong credit culture should be expected to have a countervailing effect.

At any rate, in the banks that I visited it also would be possible to use only the quantitative risk rating (which cannot be influenced by employees) for remuneration purposes.

Being able to adjust interest rates as a borrower's credit risk changes also would reduce the incentive to further influence remuneration through biased assessments. Finally, the transition procedure for replacing one relationship manager with another suggested below should help. Variable remuneration should be paid out only a year after the end of the evaluation period; the bonus payment would thus act as a bond further reducing incentives to be over-optimistic (as this would likely not, in the end, lead to higher remuneration).

Transition procedure for transfer of portfolio Occasionally, a relationship manager (or the credit analyst) leaves and is replaced with a new employee who becomes responsible for a portfolio of loans that he or she did not negotiate and the performance of which will, under the proposals outlined here, affect compensation. Such an arrangement is likely to lead to acceptance problems unless the transition procedure is fair to all parties concerned. One possible procedure that found the support of the interviewees is to effectively consider part of the outgoing relationship manager's remuneration as a performance bond and pay it, for

example, in quarterly installments deferred by one year. On taking over the portfolio, the new relationship manager (or credit analyst) has one year (the usual risk-rating rhythm) to rate all borrowers himself. The differences in risk assessment could then be debited or credited to the former relationship manager's remuneration. Once the new relationship manager has rated all borrowers, responsibility for the portfolio is his' or her's alone. It is in the interest of the new relationship manager to be as pessimistic as possible, whereas the old relationship manager will be as optimistic as possible. This conflict of interests is desired, as it will get each side to reveal all the information that supports its case. Ideally, the old relationship manager is still an employee of the bank and can negotiate directly. Having to come to an agreement on maybe two-hundred or more borrowers, there will be some compromising and give-and-take and the interviewees were confident that the new and old employees would normally find a consensus. In really critical cases, there is always the option of getting the supervisor to decide. As noted above, any relationship manager's scope for influencing the risk rating (and thus the scope for disagreement) is fairly limited anyway. Should the incumbent leave the bank, he or she may not be able to directly negotiate with the successor. In that case, the new relationship manager could negotiate with the credit analyst (assuming that the credit analyst is not leaving at the same time) whose interests are largely the same as those of the outgoing relationship manager since the risk objective should be an objective for the credit analyst, too (see below).

Risk objective for supervisors Much as was proposed with respect to the profit target, relationship managers' performance with respect to the risk objective could be used as a factor in determining their supervisors' remuneration. This would give supervisors incentives to encourage their juniors to not neglect risk-related work in order to achieve higher sales. One potentially critical aspect of such an arrangement is that this might lead supervisors (who need to act as referees whenever two employees cannot agree on a risk rating, for example) to be less objective than they would otherwise be.

One last, rather subtle, observation with respect to the objective proposed here is that over the medium-to-long term one should expect a ratchet effect: Current transition matrices have been obtained for an 'average' monitoring effort. Assuming that the risk objective has the desired effect and makes relationship managers exert higher effort, future actual loan performance will be better than it has historically been. However, as more data become available and are used to update the

transition matrices, the average performance will gradually improve and converge toward a new transition matrix for high monitoring effort. This means that over time it will become more and more difficult to do as well or better than expected.

In summary, by comparing actual and expected loan losses on a portfolio, it may be possible to measure the quality of a relationship manager's risk-related work. While such an objective may expose employees to considerable risk, simple work-arounds may be able to defuse this problem. It is not possible to say with confidence that such an arrangement would be effective in practice, but my interviews suggest that it may well work.

Possible objectives for front-office employees: Secondary (or possibly temporary) objectives

In the interviews, a number of possible objectives emerged that should be considered secondary (or possibly temporary) objectives in the sense that they capture aspects of the relationship manager's work that are not directly covered by the preceding objectives, but that also are not key in aligning the incentives of employee, bank, and regulator. I will discuss these in the following.

Possible objectives for front-office employees: Borrowers' financial statements submitted in a timely manner

As discussed above, the earlier a borrower submits financial statements the better from the point of view of credit risk management. A relationship manager has considerable influence on how quickly borrowers draw up their annual accounts and could use this influence. An objective measure could easily be defined as the percentage of time that statements are submitted before the deadline, possibly weighted by expected loss (which captures both loan volume and riskiness). Assuming the time allowed is the calendar year, submission on 01 July represents '50% of time not used', 01 April represents '75% of time not used' etc.

Possible objectives for front-office employees: Reduction in overdrafts

Within certain limits, overdrafts for business customers are not subject to the 'four-eyes-principle,' according to which all loan decisions have to be approved by both credit analyst and relationship manager. Instead it is the relationship manager alone who can approve a request for an overdraft (such a request will typically arrive electronically or by phone when paying a bill on behalf of the customer would mean that the regular line of credit is exceeded). Such overdrafts are seemingly lucrative business for the bank because of high penalty interest rates and overdraft

fees. However, they are very worrying from a credit risk management perspective (and very time-consuming: In one of the sample banks, relationship managers spend up to an hour per day authorizing overdrafts). The very fact that a customer needs an overdraft signals a difficult liquidity situation which is not a good sign. Therefore, should a bank have handled overdrafts in a lax way in the past, one objective for relationship managers could be to reduce overdrafts over time to an acceptable level. This should be done gradually because it is not reasonable to expect borrowers who have become used to generous approvals of overdrafts to adjust to a stricter policy from one day to the next. It is straightforward to formulate a performance measure that captures a relationship manager's skill in this respect.

Possible objectives for front-office employees: Number of applications submitted complete

In the cooperation between front-office and back-office, it is a frequent complaint from front-office employees that credit analysts take too long to come to a decision. Back-office employees, on the other hand, often criticize their colleagues for handing in incomplete applications. Quite apart from a decision on an incentive scheme, any bank should define very clearly when an application is complete. There may be a gray area in the case of certain complex borrowing arrangements where it is not straightforward which documents are required to evaluate the request. However, my interviewees argued that by and large it is possible to define when an application is complete. In cases where this is a sufficiently serious problem, one might consider using an objective for relationship managers whereby they are penalized for a more than negligible number of incomplete applications.

Possible objectives for front-office employees: Announcing intention to leave position as early as possible

The earlier a relationship manager announces the intention to leave, the better from a credit risk management perspective. An orderly transfer to the successor not only allows for the successor to be personally introduced, but also enables the leaving relationship manager to impart to the successor a lot of 'soft' information about the borrower that is not written down anywhere. This is undoubtedly valuable (although it is hard to quantify the value of that information). A well-planned transfer also would lead to a smooth transition with respect to the risk objective outlined above. A bank could not only guarantee that the early announcement of the intention to leave a position has no negative consequences for an employee; it could also pay a bonus to an employee who

gives notice much earlier than required by law.

7.4.5 Possible objectives for back-office employees

The primary function of credit analysts is to protect the bank and ensure that risks are accepted in a disciplined manner. Giving them balanced objectives via explicit incentive arrangements is even less straightforward than it is for the front-office. In particular their remuneration must not be linked to the profit generated by the relationship managers they are working with, for this would skew their incentives in an undesirable way. The same problem would arise if credit analysts were to receive a bonus for making cross-selling recommendations: They can often, through their analysis of the financial statements, develop ideas for further product sales to customers. However, to reward them for such additional leads gives them an incentive to approve loans that they might not otherwise approve since cross-selling is obviously only possible when a loan application is approved and not when it is denied. On the other hand, my interviews suggest that excessive risk aversion on the part of credit analysts is occasionally a real concern and that turning the back-office into a profit-center would be very helpful with respect to incentive arrangements. Such a move would allow a credit analyst to directly bill his or her front-office colleague for loan evaluations and would provide the back-office with a direct source of revenue. In particular, a back-office profit center would probably not hold the considerable buffer in manpower that many banks currently deem necessary, but rather would charge differentiated prices for evaluating loan applications at different guaranteed turnaround times. While such efficiency gains would not directly be of interest to a regulator, the resulting increased profitability of banks would function as a welcome cushion against adverse shocks.

Possible objectives for back-office employees: Risk objective

The risk objective discussed above ought to be an important element of a credit analyst's incentive scheme. Since it is only the relationship manager who has direct contact with the customers, that might be considered problematic. However, it was, in fact, the back-office managers among my interviewees who were particularly interested in the use of the risk objective, as they felt that this would make the contribution of the credit analysts measurable in a reasonably objective way. Credit analysts may not be in contact with the customer, but no lending decision and no rating decision can be made without their approval. They are well-placed to provide very important inputs to the credit risk management process:

They analyze the financial statements in depth and develop suggestions on the basis of their analysis; they should see warning signs such as a decline in sales or an increase in the use of a credit line; they can ask the relationship manager to discuss points of concern in his or her next meeting with the customer; they are typically in charge of keeping collateral values up-to-date; and so on. The dominant view among the interviewees was that there is sufficient scope for credit analysts to use their influence and that the use of a risk objective for credit analysts is warranted.

Possible objectives for back-office employees: Turn-around times

This objective would most likely be of less interest to a regulator, but would probably need to be part of a balanced package of explicit incentives for credit analysts. As mentioned above, use of such an objective requires a clear definition of when an application is complete and when, therefore, the clock starts ticking for the credit analyst. Concurrent use of the risk objective should ensure that the quality of loan decisions does not suffer.

Possible objectives for back-office employees: Complexity-adjusted number of applications

Like the preceding objective, this one would not be of much relevance to a regulator, but would likely be monitored by the bank (though not necessarily become part of an incentive arrangement in practice, since the credit analysts cannot themselves acquire further customers and so cannot be held responsible for the number of applications they process directly). The purpose is to ensure that a credit analyst does a 'sufficient' amount of work. Once again, in the highly heterogeneous business market it would be important to adjust the number of loans for complexity; for example by disaggregating every loan application into its component parts (which, as noted, my interviewees thought possible).

Possible objectives for back-office employees: Announcing intention to leave position as early as possible

The above comments on the same objective for front-office employees apply.

7.4.6 Possible objectives for the turn-around/liquidation unit

The rationale for having such a turn-around/liquidation unit[31] in the first place is that dealing with loans near or in default requires a different skill set and is highly time-consuming. It therefore makes sense for such loans to be taken out of normal front-office operations and given to specialists. As noted previously, speed in dealing with developing problems is of the essence. The risk objective for relationship managers discussed above ought to give relationship managers incentives to signal problems as early as possible. The banks that I visited were sufficiently small to envisage an incentive scheme for the turn-around/liquidation unit as a team incentive scheme for the few employees that would typically staff such a unit. However, the objectives outlined here could be broken down to individuals if the need arose. The turn-around/liquidation unit is in many ways the easiest actor in the entire credit risk management process to provide incentives for. In fact, some banks most likely have incentive schemes for their turn-around units in place. Although such schemes are conceptually fairly straightforward, there is evidence suggesting that many banks are not yet using them systematically [see Cermele, Donato and Mignanelli (2002)].

The ideas discussed here apply to business loans from a certain size onwards only. All retail loans with the exception of mortgages are too small for them to be dealt with by a highly qualified and highly paid specialist banker in a cost-effective way. Such (mostly unsecured) loans are typically bundled and auctioned off or sold directly to specialist firms that benefit from economies of scale and less expensive personnel (and, occasionally, collection techniques that, while entirely legal, a bank would not necessarily want to be associated with).

Salvaging as much as possible from business loans in or near default takes considerable skill and creativity, but given that the loan has lost much or most of its value already, can be a highly lucrative business. A turn-around/liquidation unit should be organized as a profit center; incentive pay can then simply be a share of the unit's profits. From a regulator's point of view, the key decision to be taken is that on liquidating or attempting a turn-around. That decision should be taken by a unit that has the incentives to make efficient liquidation decisions. That this is far from a merely academic issue is borne out, for example, by the experience of Japan during its long banking crisis. Peek and Rosengren

[31] For a general description of the role of a turn-around/liquidation unit by a practitioner, see Wiggers (2002)

(2005) note that

'while the evergreening of loans in Japan insulated many severely troubled Japanese firms from market forces and may have prevented a bank capital crunch, that behavior nonetheless exacerbated economic problems for the economy by promoting the allocation of an increasing share of bank credit to many of the firms least likely to use it productively.' (p. 1165)

Economywide financial problems were thus allowed to fester for years and years when it would have been much better to cut the losses at the beginning. Having a statutory requirement for an independent unit in a bank to decide on liquidation when a certain criterion is met (e.g., contractual payment more than ninety days late) would help to reduce delays in efficient liquidations.

When is liquidation efficient? What is efficient from the point of view of the bank need not be efficient overall. Things get more complicated still when there is not only a single debt-holder but several. These issues have been treated exhaustively in the theoretical literature. My purpose here is to propose an outline of a procedure that should be acceptable to both a banking regulator and a bank and which my interviewees indicated should work in practice. It is an outline only as the topic is too encompassing for it to be dealt with at length here. In particular, I avoid all legal considerations which – in the case of companies near or in bankruptcy – are likely to be very important. For example, liquidating collateral via foreclosure in case of non-payment is legally different from forcing the borrower into bankruptcy (although the two will often go together and they are treated as one here).

From the bank's point of view, liquidation will be preferable to a turn-around attempt if liquidation results in lower losses. However, gauging the likely profitability of either liquidation or a turn-around attempt is fraught with uncertainty (How much will the collateral be worth? How much fresh money will a turn-around attempt require and what are its chances of succeeding?). A turn-around/liquidation unit that is paid as a function of the profits it generates has an incentive to acquire the information and skills necessary to perform this task well; its incentives are aligned with those of the bank, albeit imperfectly[32].

The interviewees agreed that one effective way to define the turn-around/liquidation unit's success is to:

[32]It is clear that in theory an agent who bears the full cost of an action but only gets to keep a fraction of the resulting rewards does not have optimal incentives to exert effort. What is not clear is whether that is always a problem in real-life effort decisions.

- have it (fictitiously) purchase loans in difficulty or default from the front-office at a realistic (i.e., knock-down) price;
- let it retain a share of the profits obtained by either implementing a successful turnaround plan and selling the loan back to the front-office or liquidating the collateral.

The scope for a turn-around manager or liquidator to generate profits for the bank is not in doubt. The requisite skills include, for example, negotiation technique and a good business sense. An entrepreneurial turn-around manager might, for example, agree with a borrower to have relatives provide extra funds in return for a payment holiday. He or she might steer the borrower toward a turn-around consultant who can help with a necessary restructuring[33] or decide not to sell collateral immediately when, say, the local real estate market is in a slump. Apparently, re-possessed property sold at a public auction fetches a very much lower price than property skillfully sold directly through negotiations. My interviewees gave many additional examples of how skill and effort can make a difference.

The key issues to be solved in implementing the above arrangement are how to determine the price that the front-office receives for a loan and which criteria should determine when a borrower is sufficiently healthy to be returned to the front-office. In neither case did the interviews yield clear-cut rules. One approach that is always available in theory is to sell ailing loans to an outside party (say, a specialized consulting or investment firm) in a true market transaction with no promise to buy back the loan. (That approach, of course, would obviate the need to discuss possible incentive arrangements for a turn-around unit.) In the case of the small banks I visited, such an approach is not likely to be feasible as the same information problems that make small- and medium-sized businesses so opaque are compounded in the case of distressed borrowers. However, Cermele, Donato and Mignanelli (2002) suggest that this might be an option for larger borrowers. It is quite clear that it is in the interest of the relationship manager to obtain as high a price as possible whereas the turn-around/liquidation unit would prefer to pay as little as possible. A reasonable benchmark would be the amount obtainable from immediate sale of the collateral. But how would this be determined? Clearly this issue holds the potential for considerable conflict. The simplest thing may be not to determine a 'price' at all, but to fully write down a loan and to let the turn-around unit operate purely on a

[33]It was a frequent observation on the part of the interviewees that while there are many consultants, not many of them are able to bring results. A bank should be expected to have much greater transparency about quality than the borrower.

contingency basis. Alternatively one could rely on senior management to mediate. Similar issues arise with respect to the criterion that determines whether a borrower has been successfully turned around. There are other practical issues that I was unable to explore in the interviews. I do wish to note, however, that the fact that the employees concerned have to work together for many years and on many transactions makes it likely that some sort of compromise will be found. It emerged in the interviews that – apart from dealing with emerging problems rapidly – the continued involvement of the previous relationship manager is essential as he or she still has a lot of valuable information that cannot otherwise be obtained. Under the scheme proposed here, it also would be in the previous relationship manager's interest to actively participate in the turn-around or liquidation. While he or she has to take a loss initially, all revenues or proceeds from the loan in excess of the written-down loan value could be shared between the turn-around/liquidation unit and the former relationship manager. These are good reasons, then, to assume that the problems discussed here would not be critical in practice.

This ends the discussion of my findings. I first reviewed desirable features of incentive arrangements and provided some institutional and organizational background on the sample banks. I then described in detail possible objectives for line employees in lending to smaller and medium-sized businesses. The risk objective is at the heart of the discussion and was reviewed at length.

7.5 Potential Shortcomings of Current Risk Measurement Tools

As the preceding sections have made clear, reliable risk measurement/management tools are an essential prerequisite for lending-related incentive arrangements. It is worrying, then, that the risk management tools currently in use are possibly less accurate than one might want them to be. I discuss this concern in the present section.

None of my comments in this section should be construed as a statement on the risk rating tools used by the banks in my sample. The following discussion is based entirely on material in the public domain.

As shown in section 6 of Eggenberger (2006c), the literature is full of examples of incentive arrangements that had disastrous consequences because they were based on inappropriate or misleading performance measures. Unfortunately, it is unlikely that current credit risk measurement tools are as accurate as one would like them to be. To make the

term 'accurate' operational, I will phrase the discussion here in terms of the assessment of the probability of default of individual borrowers. I should note that real-world credit risk assessment is considerably more complex and involves the estimation of additional variables such as loss-given-default and, in particular, correlations and portfolio risk measures such as Credit Value at Risk (CVaR). [The estimation of credit portfolio risk is critically examined, for example, by Löffler (2003).] A borrower will either default or not, so it is impossible to say whether the risk assessment [estimate of the probability of default] of an individual borrower was ex ante correct. However, by bringing statistical techniques to bear on ex ante assessments and outcomes for a large number of loans, we can get an idea of the accuracy of the rating tool used. These validation or back-testing methods are still in their infancy as far as credit risk is concerned and a detailed discussion is far beyond the scope of this overview. However, I will present some of the concerns that have been voiced as far as they are relevant for my question of interest, namely 'Is it possible to develop lending-related incentive arrangements?'

This section has three parts: I first discuss cross-sectional credit risk ratings. The main part of this section, however, will deal with the time-dimension of credit risk and a number of related issues (such as the accuracy of economic forecasts). This is where the main conceptual problems and unresolved issues are to be found. A brief conclusion follows.

It is simplest to think about cross-sectional or relative risk mentally holding the state of the business cycle constant. If we consider a number of loans, some classified as low-risk, some classified as high-risk, and we let some time elapse, we would expect to see relatively more high-risk borrowers default (the usual horizon for this kind of exercise is one year). Making this simple idea operational with the objective of comparing the accuracy of different risk rating methodologies or even quantifying the degree of accuracy of a single risk rating tool is anything but straightforward, however: In practice, for example, one would want to consider multiple risk classes and migrations not just into the default state but also from one (non-default) risk class to another. Data issues are a considerable problem, for by definition of the long risk horizon there are relatively few observations. However, even if one had ample data, the statistical techniques for quantitatively validating rating tools are in their infancy. The Deutsche Bundesbank (2003b), for example, for the time being anticipates making extensive use of qualitative validation criteria in addition to quantitative techniques. [For a discussion of candidate quantitative techniques, see Sobehart, Keenan and Stein (2000).] The absence of a

consensus on the validation and back-testing methodology means that I cannot present here an agreed-upon list of shortcomings of current risk-rating tools. However, that we seem to be far away from anything like a standardized quality control mechanism surely is significant in itself. Not surprisingly, in a few cases where researchers had access to proprietary credit file data, systematic mistakes have been discovered. For example, Oehler, Volmar and Schark (2003) document that the pooling of data by several banks for the purpose of jointly developing a risk rating tool (which is common practice for smaller banks, for example, in Germany) may introduce significant bias. Avery, Calem and Canner (2004) show that the credit scores for their data set of US consumers systematically do not incorporate situational information (that means, for example, that a loan applicant who routinely overspends for no good reason may be assessed in the same way as an applicant who has defaulted on a loan because of a temporary medical problem; however, going forward the two individuals may well be very different credit risks). Roszbach (2004), in a more involved econometric exercise, shows the credit scoring tool in use at the Swedish bank from which he obtained his data set to be inefficient. He finds

'no evidence of banks' behaving in a way that is consistent with profit maximization'.

Although he cannot completely rule out alternative explanations for the inconsistencies he documents,

'none of these suggestions agree (...) with the practices reported by the lending institution that provided the data. Rather, the results bear strong evidence of a lending institution that has attempted to minimize risk or maximize a simple return function, without success.' (p. 956)

Of course, maximum accuracy may not be a financial institution's priority regardless of cost. It may conceivably be profit-maximizing to use a less accurate credit risk rating tool that is also less expensive to develop and maintain. However, it does not appear to be the case that banks are at present able to deliberately make such a decision with full knowledge of the respective costs and benefits. More likely, credit risk measurement tools still are not fully developed so that some institutions are working with systematically flawed methods. That, of course, would complicate the use of lending-related incentives.

At this point, one legitimate question is why, apparently, the experience and expertise developed by validating and back-testing internal market risk models (which banks have been allowed to use since the so-called Market Risk Amendment to the original Basel Accord) do not

carry over to credit risk models. First, note that what I am primarily concerned with here are rating tools to assess individual exposures. Credit portfolio models are beyond the scope of my discussion. Market risk models, however, are typically targeted at assessing Value-at-Risk (the maximum loss that is not exceeded with a given probability) for a portfolio of securities. (I conjecture, but have not verified, that that makes it difficult to use VaR in the performance evaluation of individual traders, unless these had a genuine portfolio rather than concentrated holdings of a few securities.) Second, new data to fine-tune market-risk models are generated every day and are abundantly and cheaply available at high quality. The return distributions on securities, even if not quite Normal, should also be easier to work with than those on a portfolio of loans. However, even though the quality of market risk measurement and management tools does not appear to be a top concern of either practitioners or regulators, there are good reasons for thinking that there is less to the current state of the art in that field than meets the eye. As far as I can tell, the most fundamental criticism is that such models are inherently flawed because they fail to account for the feedback mechanism that exists between market movements, models, and trading behavior. Danielsson (2002) develops this point and explains that a risk modeler is not like a meteorologist who cannot affect the weather; rather, a risk modeler indirectly also influences the data she models. The consequence of that is that

'If risk measurements influence people's behavior, it is inappropriate to assume market prices follow an independent stochastic process. This becomes especially relevant in times of crisis when market participants hedge related risks leading to the execution of similar trading strategies. The basic statistical properties of market data are not the same in crisis as they are during stable periods; therefore, most risk models provide very little guidance during crisis periods.' (p. 1274)

In other words, the models in use are likely to fail precisely when we would most like to be able to rely on them. Danielsson discusses a range of problems with current market risk models. Most importantly with respect to the discussion here, he explicitly warns against exporting the flaws of the current market risk modeling approach to the credit risk area. I wish to take two points away from this sub-section that quite naturally lead to the discussion of the time dimension of credit risk: First, current market risk measurement technology may be far less sophisticated than one might think. In particular, extreme market movements seem to occur far more frequently than models allow for. A similar problem in the case

of credit risk measurement is that the possibility of recessions occurring might not be accurately assessed. Second, neither in market nor in credit risk measurement do banks appropriately take into account that market movements and changes in the state of the economy, respectively, are not independent of their decisions. Both of these problems affect the desirability and feasibility of incentive schemes.

To restrict the discussion of the time dimension of credit risk rating to a discussion of recessions as the only driver of systematic risk is, of course, a simplification. However, concentrating on this particular aspect is not misleading, as macroeconomic fluctuations are indeed a main driver of credit risk [see, for example, Wilson (1998)]. It creates conceptual clarity to distinguish only two states of the world, expansion and recession. The key issue with respect to the time dimension of credit risk, then, is whether the probability of a recession occurring during the term of a loan is adequately taken into account (I abstract here from loans other than term loans and a host of complications such as the likely presence of lags, the duration of recessions etc). Underestimating this probability will lead to underpricing and overestimating this probability will lead to overpricing. Assume that the state of the economy is the only risk factor driving loan performance. If it were possible to hedge against a recession, then the price of the insurance could simply be factored into the loan rate. However, direct insurance against most macroeconomic risks is not available as yet (the recent attempts of two investment banks to sell 'economy options' notwithstanding). In practice, it may be possible to replicate such insurance contracts, but the costs will likely be prohibitive and there will still be basis risk. (An essential assumption of the entire discussion is that market prices for the credit risk of the particular debtor are not available.)

One could, of course, obtain estimates of the likelihood of a recession occurring over some pre-defined time horizon. However, to the best of my knowledge, all attempts at systematically obtaining reliable recession forecasts thus far have been found wanting, be they based on surveys, the extraction of information from market prices, or forecasting. Forecasting in particular has led to much disappointment [see Fildes and Stekler (2002a) as well as their discussants' comments in the same issue for a thorough discussion; Diebold (1998) provides a survey and background and Stamp (2002) an introduction; Loungani (2001) presents evidence on the failure to predict recessions in particular]. Borio (2003), p. 11 et seq., argues that there are two schools of thought on the topic of risk measurement through time. One view holds that changes in risk through

time are and will continue to be unforecastable; this is the stance taken by, for example, Goodhart (2002):

'Neither markets, nor regulators, can reasonably predict that it is now, say, one year to go till the next downturn, so we had better raise our risk premia, or capital requirements, in advance of that eventuality.' (p. 2)

The other view, to which Borio himself subscribes, is that as methods improve, there will be substantial progress. [Borio illustrates his optimism by showing that banking crises can be predicted with a reasonable degree of confidence. However, while that is certainly good news, we are still a long way away from predicting events that give us less advance warning and require less of a build-up phase such as recessions.] There is, however, consensus that our current ability to forecast material business cycle developments, in particular turning-points, is poor.

Unfortunately, the conclusion that the current credit risk rating techniques seem to draw from this observation is that the best that can be done is to assume that tomorrow will be like today. [I am painting with a very broad brush here. For a more thorough treatment of how cyclical aspects are captured in current rating tools, see Allen and Saunders (2003) and Allen and Saunders (2004), who look far beyond the estimation of default probabilities, as well as – for a slightly less technical treatment - Lowe (2002).] In their discussion of quantitative credit risk models, for example, Borio, Furfine and Lowe (2001) note that

'While the various models have different structures, most tend to extrapolate recent history in one way or another, so that good current economic conditions signal good future prospects.' (p. 20)

The standard philosophy in banks seems to be the 'point-in-time' approach whereby risk is assessed based on current conditions and with a horizon of one year. Presumably as a consequence of this short horizon, banks seem to be systematically surprised by macroeconomic developments and the consequences that these have for their loan portfolios. For example, Laeven and Majnoni (2003) find that they systematically start provisioning too late. All this is a considerable concern for regulators in connection with the internal-ratings based capital requirements under the Basel II framework which will amplify the procyclical effect of banks' risk measurement techniques [see, for example, Caterineu-Rabell, Jackson and Tsomocos (2003)]. An alternative approach labeled 'through-the-cycle rating' said to be employed by the rating agencies (the agencies themselves are careful not to use that term) does not actually strip out cyclical fluctuations either: While the assigned risk grade of a borrower

may well stay unchanged through a recession, the probability of default that is implicitly associated with the risk grade may vary by up to a factor of four between recession and expansion [e.g., compare the conditional transition matrices for an initial rating of B on p. 462 of Bangia et al. (2002); the implied default probability goes up from 1.96 % to 9.2 %. Also see Nickell, Perraudin and Varotto (2000).] Banks' credit ratings are more likely to be directly calibrated to default probabilities and, therefore, more likely to be revised over the cycle. At any rate, if capital requirements have to be adjusted one-for-one with changes in risk, the resulting fluctuations are enormous. In a knock-on effect, high capital requirements curtail banks' ability to extend credit which, in the aggregate, will further slow the economy, thus worsening credit quality in the real sector, forcing banks to increase their capital and restrict their lending still further. It is quite clear that an effect is in operation that is rather similar to what has been described for the case of market risk measurement above. A number of remedies to deal with this excessive procyclicality have been proposed; among these are, for example, 'dynamic provisioning' [pioneered by Spain; see Fernández de Lis, Martínez Pagés and Saurina (2000) as well as Mann and Michael (2002)] or

'to try to relate prudential requirements much more to the rates of growth of variables, relative to their longer term trends, and not to their levels.' [Goodhart (2002), p. 2]

Even if these approaches were effective from the point of view of a regulator, i.e., on an aggregate level, it is still not clear how that would lead to correct risk assessments being transmitted to loan officers.

This brings the discussion in this section to an end. The objective was to describe problems with current risk measurement techniques that might affect the implementation of lending-related incentive schemes. On the balance of the material that I have reviewed, I cannot rule out that risk measurement problems are sufficiently severe that they constitute a critical obstacle to the implementation of incentive schemes. However, assessing the extent of possible problems would only be possible on the basis of more in-depth empirical work.

7.6 Conclusion

Based on interviews with approximately 120 senior practitioners drawn from a sample of German banks, this section investigated the scope for devising effective lending-related incentive arrangements that could double as a tool of banking regulation. The results suggest that such schemes

may – some caveats notwithstanding – be feasible and that regulators should explore this idea further. The possible objectives presented here have not yet been tested in practice; this research was exploratory in nature and focused on obtaining practitioners' views of the feasibility of incentive schemes and on identifying candidate objectives that could become part of such schemes. The bankers I interviewed were by and large in favor of the ideas outlined. The results reported here are not representative, of course, but the issues and questions discussed are likely to be relevant beyond the sample banks. Many questions remain, some of which can only be answered after an incentive scheme has been implemented. For example, the objectives discussed may need to be combined to create a balanced scheme, and there is no guarantee that the right balance can be achieved and that, in particular, employees would be induced to show the right degree of risk aversion.

In a next step, regulators (who have an easier time obtaining access to confidential information than outside researchers; this is especially valuable with respect to something as sensitive as compensation) should seek more comprehensive information on incentive schemes in operation and develop a better understanding of how lending-related incentives are or could be structured. If researchers associated with the regulators were able to gain access to a wider range of banks, this might afford an excellent opportunity to collect data for 'insider econometrics' and study the actual impact of different incentive structures (or lack thereof) on various measures of performance directly. That effort would be well invested. Enormous sums are currently being spent on upgrading banks' technology, possibly creating a false sense of security while weaknesses due to counterproductive incentives might be neglected.

One would be foolish to assume that people do not respond to incentives (whatever these incentives may be; they do not necessarily need to be financial). The priority for regulators should be to review existing incentive arrangements with a view to prohibiting dysfunctional elements as part of the Supervisory Review Process. Rewarding employees for non-risk-adjusted loan volume is an example of an incentive arrangement that a regulator, in the light of more conclusive evidence, might want banks to get rid of. One can also think about less defensive, more sophisticated approaches through which a regulator deliberately tries to steer employee behavior in desirable ways. To attempt to target just the right behavior through financial incentives may be ambitious, but my findings suggest that this route may well be worth exploring. At any rate, a regulator has ways to create awareness for potentially problematic incentive arrange-

ments short of imposing mandatory rules. My research is a first step in understanding these issues.

The focus throughout was on explicitly linking remuneration to objective performance measures, as that is what a regulator could most easily impose. In practice, a bank would want to integrate an explicit incentive scheme into its human resource management strategy and complement it with additional elements. Intrinsic motivation of employees may go a long way towards ensuring that risk is taken in a disciplined way and while this clearly cannot be imposed by a regulator, it is something that a bank would do well to foster. In this respect, one may have reservations regarding explicit contingent-pay schemes no matter how well they are designed as these might crowd out intrinsic motivation. I discuss this point as well as other doubts with respect to explicit incentive schemes in section 6 in Eggenberger (2006c). These issues are not central to my analysis, but ought to be considered in thinking about regulating incentive arrangements.

One important unknown is the quality of risk management tools. As argued above, the performance of a relationship manager or credit analyst can hardly be assessed independently of that of the risk rating tools used. There is some evidence to suggest that – the enormous technical progress of recent years notwithstanding – current rating tools may have systematic flaws.

In summary, my findings do not provide conclusive proof that the regulation of incentive compensation in banks should be a tool of financial regulation. However, the results of my fieldwork constitute a first stab at the question of whether there is scope for devising effective lending-related incentive arrangements that align the interests of a regulator with those of the banks. My hope is that regulators will build on these findings and use their influence to further explore this sensitive but critically important area.

Chapter 8

Conclusion

This thesis has contributed to the literature on the link between the intensity of bank competition and the allocation of capital in three main ways.

Following an extensive literature summary (chapter 2), I extended the seminal contribution by Broecker (1990) and showed that contrary to Broecker's conjecture, his bank competition game does not have an equilibrium in pure strategies even when banks are asymmetric (chapter 3). In chapter 4 I modified the auction framework used by Broecker and allowed applicants to renegotiate their offers. I showed that the modified game has an equilibrium in pure strategies where for a given number of banks an applicant's interest rate decreases with the number of offers obtained. Surplus unambiguously increases with entry as long as additional banks have an incentive to enter, so the allocation of capital – in contrast to much of the rest of the theoretical literature – is shown to improve with the intensity of competition. Clearly, the question is whether an auction framework is appropriate for modeling bank competition. I argue that it is not and that modeling creditworthiness tests should be robust to the empirically highly relevant phenomenon of test results becoming (partially) public and thus available to competitors.

My second contribution was to reconcile a discrepancy between the empirical and theoretical literatures on bank competition and capital allocation. In the literature review, I showed that the theoretical literature is somewhat pessimistic with respect to the effect of more intense bank competition on the allocation of capital whereas the empirical literature was shown to have found large gains from more intense bank competition. I developed the argument that the main role of more intense competition may be that of a discipline device. When other mechanisms of corporate governance fail for whatever reason (which they do, as the evidence on X-inefficiency in banks suggests), a major benefit of more

intense competition may be that it forces managers to work harder and thus leads to improvements in the allocation of capital. The existing theoretical work cannot capture this effect, because it assumes that banks are profit-maximizing entities. I motivated and formalized my conjecture and demonstrated that in the presence of an effort-minimizing bank, the allocation of capital unambiguously improves as competition becomes more intense.

There may be a more general point here, too. Information problems and related obstacles to an efficient allocation of capital (such as the moral hazard problem that I model on the part of the borrowers) often motivate suspicion with respect to the unfettered operation of markets. Indeed, starting with the famous 'lemons' paper [Akerlof (1970)], much research has been devoted to demonstrating how such problems may lead to a market shutdown. A well-known example from the banking literature rationalizing a partial market failure is Stiglitz and Weiss (1981). The contribution that chapter 5 makes is to appeal to the empirical evidence and suggest that in some settings the key issue in improving the allocation of capital may not be information problems (which to some extent can be overcome, albeit at a cost), but agency problems. As I show, the policy implications of the two kinds of problems may be diametrically opposed.

Finally, Chapters 6 and 7 deal with the transmission mechanism that links the intensity of bank competition and the allocation of capital by investigating the subject of lending-related incentive arrangements; a subject about which little is known. The existing theoretical literature treats banks as black boxes, so the work reported in chapters 6 and 7 can be read as a complement to the theoretical work in the preceding chapters: One reason to study real-world incentive structures is as an extension to theoretical work akin to a robustness check.

As far as the real-world applicability of theoretical work is concerned, one crucial question is how the pressures that the bank is facing at the institution level are transmitted throughout the organization. As an economist, one is trained to look for and analyze incentives as a key element of this transmission mechanism. The interview-based fieldwork I presented in chapters 6 and 7 therefore addressed the question 'What scope is there for devising effective lending-related incentive arrangements in practice?' While it would be intellectually dishonest to claim to be able to give a definitive answer to the question on the basis of the fieldwork reported here, my findings indicate that it may be possible to construct sensible schemes. I discussed a number of candidate objectives in detail and presented obstacles and potential problems discovered in the

interviews. These findings are a substantial step forward and shed light on the problems inherent in designing and implementing lending-related incentives.

My hope is that other researchers will build on my findings and further explore this sensitive but critically important area. In a next step, regulators (who have an easier time obtaining access to confidential information than outside researchers; this is especially valuable with respect to something as sensitive as compensation) should seek more comprehensive information on incentive schemes in operation and develop a better understanding of how lending-related incentives are or could be structured. If researchers associated with the regulators were able to gain access to a wider range of banks, this might afford an excellent opportunity to collect data for 'insider econometrics' and study the actual impact of different incentive structures (or lack thereof) on various measures of performance and thence the allocation of capital directly.

Chapter 9

Bibliography

Aghion, Philippe, Mathias Dewatripont and Patrick Rey (1997), Corporate Governance, Competition Policy, and Industrial Policy, European Economic Review, vol. 41, 797-805

Aghion, Philippe, Mathias Dewatripont and Patrick Rey (1999), Competition, Financial Discipline, and Growth, Review of Economic Studies, vol. 66, no. 4, 825-852

Aghion, Philippe and Rachel Griffith (2005), Competition and Growth, Cambridge, MA: MIT Press, 2005

Aghion, Philippe, Christopher Harris, Peter Howitt and John Vickers (2001), Competition, Imitation, and Growth with Step-by-Step Innovation, Review of Economic Studies, vol. 68, 467-492

Aghion, Philippe and Peter Howitt (1996), Research and Development in the Growth Process, Journal of Economic Growth, vol. 1, March 1996, 49-73

Agrawal, Raj, Stephen Findley, Sean Greene, Kathryn Huang, Aly Jeddy, William W. Lewis and Markus Petry (1996), Capital Productivity: Why the US Leads and Why It Matters, McKinsey Quarterly, no. 3, 1996,

Akerlof, George (1970), The Market for "Lemons": Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, vol. 84, no. 3, August 1970, 488-500

Allen, Franklin and Douglas Gale (2000b), Corporate Governance and Competition, Ch. 2, in: Vives, Xavier (ed.), Corporate Governance: Theoretical and Empirical Perspectives, Cambridge: Cambridge University Press, 2000b, 23-94

Allen, Franklin and Douglas Gale (2004), Competition and Financial Stability, Journal of Money, Credit, and Banking, vol. 36, no. 3, 453-480

Allen, Franklin, Hans Gersbach, Jan-Pieter Krahnen and Anthony Santomero (2001), Competition among Banks: Introduction and Con-

ference Overview, European Finance Review (now: Review of Finance), vol. 5, 1-11

Allen, Linda and Anthony Saunders (2003), A Survey of Cyclical Effects in Credit Risk Measurement Models, BIS Working Paper, January 2003

Allen, Linda and Anthony Saunders (2004), Incorporating Systemic Influences into Risk Measurements: A Survey of the Literature, Journal of Financial Services Research, vol. 26, no. 2, 161-192

Almazan, Andres (2002), A Model of Competition in Banking: Bank Capital vs Expertise, Journal of Financial Intermediation, vol. 11, 87-121

Almazan, Andres and Javier Suarez (2003), Managerial Compensation and the Market Reaction to Bank Loans, Review of Financial Studies, vol. 16, no. 1, 237-261

Amel, Dean and Martha Starr-McCluer (2002), Market Definition in Banking: Recent Evidence, Antitrust Bulletin, vol. 47, no. 1, Spring 2002, 63-89

Avery, Robert, Paul Calem and Glenn Canner (2004), Consumer Credit Scoring: Do Situational Circumstances Matter? Journal of Banking and Finance, vol. 28, 835-856

Babcock, Linda and George Loewenstein (1997), Explaining Bargaining Impasse: The Role of Self-Serving Biases, Journal of Economic Perspectives, vol. 11, no. 1, Winter 1997, 109-126

BaFin (Bundesanstalt für Finanzdienstleistungsaufsicht) (2002), Mindestanforderungen an das Kreditgeschäft der Kreditinstitute, Bonn: 20 December 2002

Baker, George (2000), The Use of Performance Measures in Incentive Contracting, American Economic Review, vol. 90, no. 2, May 2000, 415-420

Bandiera, Oriana, Gerard Caprio, Patrick Honohan and Fabio Schiantarelli (2000), Does Financial Reform Raise or Reduce Saving? Review of Economics and Statistics, vol. 82, no. 2, May 2000, 239-263

Banerjee, Priyodorshi (2005), Information Acquisition under Uncertainty in Credit Markets, Review of Financial Studies, vol. 18, no. 3, 2005, 1075-1104

Bangia, Anil, Francis Diebold, André Kronimus, Christian Schagen and Til Schuermann (2002), Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing, Journal of Banking and Finance, vol. 26, 445-474

Bank for International Settlements (BIS) (2001), Annual Report 2001 (01.04.2000 to 31.03.2001), Basel: 2001

Baron, James M. and David M. Kreps (1999), Strategic Human Resources, New York etc: John Wiley & Sons, Inc., 1999

Bartel, Ann (2004), Human Resource Management and Organizational Performance: Evidence from Retail Banking, Industrial and Labor Relations Review, vol. 57, no. 2, January 2004, 181-203

Basel Committee on Banking Supervision (1999c), Credit Risk Modelling: Current Practices and Applications, Basel: April 1999

Basel Committee on Banking Supervision (2000a), Principles for the Management of Credit Risk, Basel: September 2000

Basel Committee on Banking Supervision (2000c), Summary of Responses Received on the Report "Credit Risk Modelling: Current Practice and Applications", Basel: May 2000

Basel Committee on Banking Supervision (2000d), Range of Practice in Banks' Internal Rating Systems, Basel: January 2000

Basel Committee on Banking Supervision (2004), International Convergence of Capital Measurement and Capital Standards - A Revised Framework, Basel: June 2004

Bauer, Paul, Allen Berger, Gary Ferrier and David Humphrey (1998), Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods, Journal of Economics and Business, vol. 50, 85-114

Berger, Allen, Aslı Demirgüç-Kunt, Ross Levine and Joseph G. Haubrich (2004), Bank Concentration and Competition: An Evolution in the Making, Journal of Money, Credit, and Banking, vol. 36, no. 3, June 2004 (Part 2), 433-451

Berger, Allen, Rebecca Demsetz and Philip Strahan (1999), The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future, Journal of Banking and Finance, vol. 23, 135-194

Berger, Allen, Robert DeYoung, Hesna Genay and Gregory Udell (2000), Globalization of Financial Institutions: Evidence from Cross-Border Banking Performance, Brookings-Wharton Papers on Financial Services, 23-158

Berger, Allen and Timothy Hannan (1998), The Efficiency Cost of Market Power in the Banking Industry: A Test of the "Quiet Life" and Related Hypotheses, Review of Economics and Statistics, vol. 80, no. 3, August 1998, 454-465

Berger, Allen and Loretta Mester (1997), Inside the Black Box: What explains Differences in the Efficiencies of Financial Institutions? Journal of Banking and Finance, vol. 21, 895-947

Berger, Allen and Gregory Udell (1995), Relationship Lending and Lines of Credit in Small Firm Finance, Journal of Business, vol. 68, no. 3, July 1995, 351-381

Bernanke, Ben and Mark Gertler (1989), Agency Costs, Net Worth, and Business Fluctuations, American Economic Review, vol. 79, no. 1, March 1989, 14-31

Bertoletti, Paolo and Clara Poletti (1996), A Note on Endogenous Firm Efficiency in Cournot Models of Incomplete Information, Journal of Economic Theory, vol. 71, 303-310

Besanko, David and Anjan Thakor (1987), Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets, International Economic Review, vol. 28, no. 3, October 1987, 671-689

Besanko, David and Anjan V. Thakor (1992), Banking Deregulation: Allocational Consequences of Relaxing Entry Barriers, Journal of Banking and Finance, vol. 16, 909-932

Bester, Helmut (1985), Screening vs. Rationing in Credit Markets with Imperfect Information, American Economic Review, vol. 75, no. 4, September 1985, 850-855

Bewley, Truman (1999), Work Motivation, Federal Reserve Bank of St. Louis Review, vol. 81, no. 3, May-June 1999, 35-49

Bewley, Truman (2002), Interviews as a Valid Empirical Tool in Economics, Journal of Socio-Economics, vol. 31, 343-353

Bikker, Jacob (2003), Testing for Imperfect Competition on EU Deposit and Loan Markets with Bresnahan's Market Power Model, Kredit und Kapital, vol. 36, no. 2, 167-212

Bikker, Jacob and Katharina Haaf (2002a), Measures of Competition and Concentration in the Banking Industry: A Review of the Literature, Economic & Financial Modelling, vol. 9, no. 2, Summer 2002, 53-97

Bikker, Jacob and Katharina Haaf (2002b), Competition, Concentration, and their Relationship: An Empirical Analysis of the Banking Industry, Journal of Banking and Finance, vol. 26, 2191-2214

Bird, Anat (2002), An incentive plan will work only if you target the right behavior, American Banker, vol. 167, no. 95, 17 May 2002, 9

Black, Sandra and Philip Strahan (2002), Entrepreneurship and Bank Credit Availability, Journal of Finance, vol. 56, no. 6, December 2002, 2807-2833

Blinder, Alan (1991), Why are Prices Sticky? Preliminary Results from an Interview Study, American Economic Review, vol. 81, no. 2, May 1991, 89-96

Board of Governors of the Federal Reserve System (2002), Report to

the Congress on the Availability of Credit to Small Businesses, Washington, DC: September 2002

Board of Governors of the Federal Reserve System (2003b), Draft Supervisory Guidance with request for comment: Internal Ratings-Based Systems for Corporate Credit and Operational Risk Advanced Measurement Approaches for Regulatory Capital, Federal Reserve System, Docket No. OP-1153 (also available from www.federalreserve.gov), Washington, DC: 04 August 2003

Board of Governors of the Federal Reserve System and U.S. Securities and Exchange Commission (2000), Report to the Congress on Markets for Small-Business and Commercial-Mortgage-Related Securities, Washington, DC: September 2000

Bolton, Patrick and Mathias Dewatripont (2005), Contract Theory, Cambridge, MA: The MIT Press, 2005

Bölz, Karlheinz and Stefan Thorsch (2002), Kapital ersetzt kein internes Kontrollsystem, Handelsblatt, 15 May 2002

Bonaccorsi di Patti, Emilia and Giovanni Dell'Ariccia (2004), Bank Competition and Firm Creation, Journal of Money, Credit, and Banking, vol. 36, no. 2, April 2004, 225-251

Boot, Arnoud W.A. and Anjan V. Thakor (2000), Can Relationship Banking Survive Competition? Journal of Finance, vol. 55, no. 2, April 2000, 679-713

Borio, Claudio (2003), Towards a Macroprudential Framework for Financial Supervision and Regulation? BIS Working Paper, February 2003

Borio, Claudio, Craig Furfine and Philip Lowe (2001), Procyclicality of the Financial System and Financial Stability: Issues and Policy Options, BIS Papers, 2001

Bouckaert, Jan and Hans Degryse (2004), Softening Competition by Inducing Competition in Credit Markets, Journal of Industrial Economics, vol. 52, no. 1, March 2004, 27-52

Broecker, Thorsten (1990), Credit-Worthiness Tests and Interbank Competition, Econometrica, vol. 58, no. 2, March 1990, 429-452

Caminal, Ramon and Carmen Matutes (2002), Can Competition in the Credit Market Be Excessive? manuscript available at: http:// www.iae.csic.es/ caminal/pub/ excescomp.pdf (last accessed on 24 March 2005), July 2002

Canoy, Marcel, Machiel van Dijk, Jan Lemmen, Ruud de Mooij and Jürgen Weigand (2001), Competition and Stability in Banking, The Hague: December 2001

Cao, Melanie and Shouyong Shi (2001), Screening, Bidding, and the

Loan Market Tightness, European Finance Review (now: Review of Finance), vol. 5, 21-61

Caouette, John, Edward Altman and Paul Narayanan (1998), Managing credit risk: The next great financial challenge, New York etc: Wiley & Sons, Inc., 1998

Carey, Mark (1998), Credit Risk in Private Debt Portfolios, Journal of Finance, vol. 53, no. 4, August 1998, 1363-1387

Carletti, Elena and Philipp Hartmann (2002), Competition and Stability: What's Special about Banking? European Central Bank Working Paper Series, May 2002

Caterineu-Rabell, Eva, Patricia Jackson and Dimitrios Tsomocos (2003), Procyclicality and the New Basel Accord - Banks' Choice of Loan Rating System, Bank of England Working Paper, March 2003

Caves, Richard E. and Associates (1992), Industrial Efficiency in Six Nations, Cambridge, MA: MIT Press, 1992

Caves, Richard E. and David R. Barton (1990), Efficiency in US Manufacturing Industries, Cambridge, MA: MIT Press, 1990

Cermele, Michele, Maurizio Donato and Andrea Mignanelli (2002), Good Money from Bad Debt, McKinsey Quarterly, no. 1, 103-111

Cetorelli, Nicola (1997), The Role of Credit Market Competition on Lending Strategies and on Capital Accumulation, Federal Reserve Bank of Chicago Working Paper, 06 December 1997

Cetorelli, Nicola (1999), Competitive Analysis in Banking: An Appraisal of the Methodologies, Federal Reserve Bank of Chicago Economic Perspectives, Q1 1999, 2-15

Cetorelli, Nicola (2001), Competition among Banks: Good or Bad? Federal Reserve Bank of Chicago Economic Perspectives, Q2, 38-48

Cetorelli, Nicola (2003), Life-Cycle Dynamics in Industrial Sectors: The Role of Banking Market Structure, Federal Reserve Bank of St. Louis Review, July/August 2003, 135-147

Cetorelli, Nicola and Michele Gambera (2001), Banking Market Structure, Dependence, and Growth: International Evidence from Industry Data, Journal of Finance, vol. 56, no. 2, April 2001, 617-648

Cetorelli, Nicola and Pietro Peretto (2000), Oligopoly Banking and Capital Accumulation, Federal Reserve Bank of Chicago Working Paper, December 2000

Cetorelli, Nicola and Philip Strahan (2006), Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets, Journal of Finance, vol. 61, no. 1, February 2006, 437-461

Chan, Yuk-Shee, Stuart I. Greenbaum and Anjan V. Thakor (1986),

Information Reusability, Competition, and Bank Asset Quality, Journal of Banking and Finance, vol. 10, 243-253

Chan, Yuk-Shee and Anjan V. Thakor (1987), Collateral and Competitive Equilibria with Moral Hazard and Private Information, Journal of Finance, vol. 42, no. 2, June 1987, 345-363

Chiappori, Pierre-André, David Perez-Castrillo and Thierry Verdier (1995), Spatial Competition in the Banking System: Localization, Cross-Subsidies, and the Regulation of Deposit Rates, European Economic Review, vol. 39, 889-918

Chiesa, Gabriella (1998), Information Production, Banking Industry Structure, and Credit Allocation, Research in Economics, vol. 52, 409-430

Citron, Laura and Richard Walton (2002), International Comparisons of Company Profitability, Economic Trends, vol. 587, October 2002, 21-34

Claessens, Stijn and Daniela Klingebiel (2001), Competition and Scope of Activities in Financial Services, World Bank Research Observer, vol. 16, no. 1, Spring 2001, 19-40

Claessens, Stijn and Luc Laeven (2005), Financial Dependence, Banking Sector Competition, and Economic Growth, Journal of the European Economic Association, vol. 3, no. 1, March 2005, 179-207

Coco, Giuseppe (2000), On the Use of Collateral, Journal of Economic Surveys, vol. 14, no. 2, 191-214

Comanor, William S. (1994), Review of Caves, Richard E. and Ass., Industrial Efficiency in Six Nations, Cambridge, MA: MIT Press, 1992, Journal of Economic Literature, vol. 32, no. 3, September 1994, 1232-1234

Committee on the Global Financial System (2003), Credit Risk Transfer, Basel: January 2003

Crespi, Rafel, Miguel A. Garcia-Cestona and Vicente Salas (2004), Governance Mechanisms in Spanish Banks: Does Ownership Matter? Journal of Banking and Finance, vol. 28, 2311-2330

Crouhy, Michel, Dan Galai and Robert Mark (2001), Risk Management, New York: McGraw-Hill, 2001

Daníelsson, Jón (2002), The Emperor has No Clothes: Limits to Risk Modelling, Journal of Banking and Finance, vol. 26, 1273-1296

Davis, Kevin (1995), Bank Deregulation, Supervision, and Agency Problems, Australian Economic Review, no. 111, 3rd Quarter 1995, 43-54

De Haas, Ralph and Ilko Naaborg (2005a), Internal Capital Markets

in International Banks: Implications for European Transition Countries, De Nederlandsche Bank working paper, August 2005

De Haas, Ralph and Ilko Naaborg (2005b), Does Foreign Bank Entry Reduce Small Firms' Access to Credit? Evidence from European Transition Economies, De Nederlandsche Bank working paper, August 2005

Degryse, Hans and Steven Ongena (2004), The Impact of Technology and Regulation on the Geographical Scope of Banking, Oxford Review of Economic Policy, vol. 20, no. 4, 571-590

Degryse, Hans and Steven Ongena (2005), Distance, Lending Relationships, and Competition, Journal of Finance, vol. 60, no. 1, February 2005, 231-266

Dell'Ariccia, Giovanni (2000), Learning by Lending, Competition, and Screening Incentives in the Banking Industry, IMF Working Paper, 28 February 2000

Dell'Ariccia, Giovanni (2001), Asymmetric Information and the Structure of the Banking Industry, European Economic Review, vol. 45, 1957-1980

Dell'Ariccia, Giovanni (2003), Banking Systems in a Financially Integrated World, IMF Research Bulletin, vol. 4, no. 1, March 2003, 1, continued on 4-6

Dell'Ariccia, Giovanni, Ezra Friedman and Robert Marquez (1999), Adverse Selection as a Barrier to Entry in the Banking Industry, Rand Journal of Economics, vol. 30, no. 3, Autumn 1999, 515-534

Dell'Ariccia, Giovanni and Robert Marquez (2004), Information and Bank Credit Allocation, Journal of Financial Economics, vol. 72, 185-214

Deutsche Bundesbank (2003b), Validierungsansätze für interne Ratingsysteme, Deutsche Bundesbank Monatsbericht, September 2003, 61-74

Diamond, Douglas (1984), Financial Intermediation and Delegated Monitoring, Review of Economic Studies, vol. 51, no. 3, July 1984, 393-414

Diamond, Douglas and Philip Dybvig (1983), Bank Runs, Deposit Insurance, and Liquidity, Journal of Political Economy, vol. 91, no. 3, June 1983, 401-419

Diamond, Peter (1965), National Debt in a Neoclassical Growth Model, American Economic Review, vol. 55, no. 5, December 1965, 1126-1150

Diebold, Francis (1998), The Past, Present, and Future of Macroeconomic Forecasting, Journal of Economic Perspectives, vol. 12, no. 2, Spring 1998, 175-192

Dinç, I. Serdar (2000), Bank Reputation, Bank Commitment, and the Effects of Competition in Credit Markets, Review of Financial Studies, vol. 13, no. 3, Autumn 2000, 781-812

Döhring, Jens and Jürgen Hromadka (2002), Die Ausfallrisikobepreisung, Bankinformation, June 2002, 36-41

Dyck, Alexander and Luigi Zingales (2004), Private Benefits of Control: An International Comparison, Journal of Finance, vol. 59, no. 2, April 2004, 537-600

Edwards, Franklin (1977), Managerial Objectives in Regulated Industries: Expense-Preference Behavior in Banking, Journal of Political Economy, vol. 85, no. 1, February 1977, 147-162

Edwards, Franklin (1999), Hedge Funds and the Collapse of Long-Term Capital Management, Journal of Economic Perspectives, vol. 13, no. 2, Spring 1999, 189-210

Eggenberger, Knut (2006a), A Note on Product Market Competition: Theory and Evidence, LSE manuscript (available from the author upon request: K.Eggenberger@ lse.ac.uk), 20 December 2005

Eggenberger, Knut (2006b), Introduction and Context (Bank Competition and Capital Allocation: A Survey of the Literature - Introduction), LSE manuscript (available from the author upon request: K.Eggenberger@lse.ac.uk), 20 December 2005

Eggenberger, Knut (2006c), Lending-related Incentive Schemes: Supplementary Material, LSE manuscript (available from the author upon request: K.Eggenberger@ lse.ac.uk), 23 January 2006

Engelbrecht-Wiggans, Richard, Paul R. Milgrom and Robert Weber (1983), Competitive Bidding and Proprietary Information, Journal of Mathematical Economics, vol. 11, no. 2, April 1983, 161-169

English, William and William Nelson (1998), Bank Risk Rating of Business Loans, Federal Reserve Finance and Economics Discussion Series Working Paper No. 51, November 1998

Ervin, Wilson and Joseph Seidel (2002), Basel Capital Accord Must Leave Some Room for Human Judgement, American Banker, vol. 167, no. 167, 30 August 2002, 6

Erxleben, Sascha and Bernhard Krob (2002), Turbo für Kreditgenossenschaften, Bankinformation, March 2002, 28-32

Farrell, M.J. (1957), The Measurement of Productive Efficiency, Journal of the Royal Statistical Society - Series A (General), vol. 120, no. 3, 253-290

Federal Reserve System Task Force on Internal Credit Risk Models (1998), Credit Risk Models at Major U.S. Banking Institutions: Current

State of the Art and Implications for Assessments of Capital Adequacy, Washington, DC: May 1998

Fernández de Lis, Santiago, Jorge Martínez Pagés and Jesús Saurina (2000), Credit Growth, Problem Loans, and Credit Risk Provisioning in Spain, Bank of Spain Working Paper, October 2000

Fildes, Robert and Herman Stekler (2002a), The State of Macroeconomic Forecasting, Journal of Macroeconomics, vol. 24, 435-468

Frame, Scott, Aruna Srinivasan and Lynn Woosley (2001), The Effect of Credit Scoring on Small-Business Lending, Journal of Money, Credit, and Banking, vol. 33, no. 3, August 2001, 813-825

Freixas, Xavier and Jean-Charles Rochet (1997), Microeconomics of Banking, Cambridge, MA etc: MIT Press, 1997

Gehrig, Thomas (1998a), Screening, Cross-Border Banking, and the Allocation of Credit, Research in Economics, vol. 52, 387-407

Gehrig, Thomas (1998b), Screening, Market Structure, and the Benefits from Integrating Loan Markets, University of Freiburg manuscript, July 1998

Gehrig, Thomas and Rune Stenbacka (2001), Screening Cycles, CEPR Discussion Paper, August 2001

Gensler, Gary (1999), Testimony before the House [of Representatives] Committee on Banking and Financial Services (also available at http://www.ustreas.gov/press/releases/rr3137.htm), Washington, DC: 06 May 1999

Gilbert, R. Alton and Adam Zaretsky (2003), Banking Antitrust: Are the Assumptions Still Valid? Federal Reserve Bank of St. Louis Review, November/December 2003, 29-52

Goodhart, Charles (1996), An Incentive Structure for Financial Regulation, LSE Financial Markets Group Special Paper, July 1996

Goodhart, Charles (2002), Changes in Risk Through Time: Measurement and Policy Responses, Paper given at the 38th Annual Conference on Bank Structure and Competition at the Federal Reserve Bank of Chicago, 09 May 2002

Graziano, Clara and Bruno M. Parigi (1998), Do Managers Work Harder in Competitive Industries? Journal of Economic Behavior and Organisation, vol. 34, 489-498

Green, Alison and David Mayes (1991), Technical Inefficiency in Manufacturing Industries, Economic Journal, vol. 101, no. 406, May 1991, 523-538

Greenbaum, Stuart and Thakor Thakor (1995), Contemporary Financial Intermediation, Orlando, FL: The Dryden Press, 1995

Greenspan, Alan (1998), Private-sector refinancing of the large hedge fund, Long-Term Capital Management, [Testimony before the Committee on Banking and Financial Services, U.S. House of Representatives (also available at http:// www.federalreserve.gov/ boarddocs/testimony/ 1998/19981001.htm)], Washington, DC: 01 October 1998

Greenwich Associates (2002), Middle Market Banking: Credit Becomes More Critical, Greenwich, CT: Greenwich Associates, 2002

Griffith, Rachel (2001), Product Market Competition, Efficiency and Agency Costs: An Empirical Analysis, Institute for Fiscal Studies Working Paper No. 2001/12, June 2001

Griffith, Rachel and Rupert Harrison (2004), The link between product market reform and macro-economic performance, Economic Paper No. 209, Brussels: European Commission - Directorate General for Economic and Financial Affairs, August 2004

Group of Ten (2001), Report on Consolidation in the Financial Sector, January 2001

Gual, Jordi (1999a), Deregulation, Integration, and Market Structure in European Banking, EIB Papers, vol. 4, no. 2, 34-48

Gual, Jordi (1999b), Deregulation, Integration, and Market Structure in European Banking, Journal of the Japanese and International Economies, vol. 13, 372-396

Guiso, Luigi, Paolo Sapienza and Luigi Zingales (2003), The Cost of Banking Regulation, University of Chicago Working Paper, June 2003

Guzman, Mark (2000), The Economic Impact of Bank Structure: A Review of Recent Literature, Federal Reserve Bank of Dallas Economic and Financial Review, Q2 2000, 11-25

Guzman, Mark (2000b), Bank Structure, Capital Accumulation, and Growth: A Simple Macroeconomic Model, Economic Theory, vol. 16, 421-455

Hainz, Christa (2003), Bank Competition and Credit Markets in Transition Economies, Journal of Comparative Economics, vol. 31, 223-245

Hanazaki, Masaharu and Akiyoshi Horiuchi (2003), A Review of Japan's Bank Crisis from the Governance Perspective, Pacific-Basin Finance Journal, vol. 11, 305-325

Hannan, Timothy (1979), Expense-Preference Behavior in Banking: A Re-Examination, Journal of Political Economy, vol. 87, no. 4, August 1979, 891-895

Hannan, Timothy and Ferdinand Mavinga (1980), Expense Preference and Managerial Control: The Case of the Banking Firm, Bell Jour-

nal of Economics, vol. 11, no. 2, Autumn 1980, 671-682

Hart, Oliver (1983), The Market Mechanism as an Incentive Scheme, Bell Journal of Economics, vol. 14, no. 2, Autumn 1983, 366-382

Hart, Oliver and John Moore (1995), Debt and Seniority: An Analysis of the Role of Hard Claims in Constraining Management, American Economic Review, vol. 85, no. 3, 567-585

Hauswald, Robert (2005), personal communication, email exchange on his forthcoming RFS paper (with Robert Marquez), 04 July 2005

Hauswald, Robert and Robert Marquez (2005), Competition and Strategic Information Acquisition in Credit Markets, Review of Financial Studies, forthcoming, volume, issue, and page numbers not available at the time of writing

Hay, Donald A. and Guy S. Liu (1997), The Efficiency of Firms: What Difference does Competition make? Economic Journal, vol. 107, May 1997, 597-617

Hayek, Friedrich August (1945), The Use of Knowledge in Society, American Economic Review, vol. 35, no. 4, September 1945, 519-530

Hellwig, Martin (1991), Banking, Financial Intermediation and Corporate Finance, in: Giovannini, Alberto and Colin Mayer (eds), European Financial Integration, Cambridge etc: CUP, 1991, 35-63

Helper, Susan (2000), Economics and Field Research: "You can observe a lot just by watching", American Economic Review, vol. 90, no. 2, May 2000, 228-232

Hendricks, Kenneth, Robert Porter and Charles Wilson (1994), Auctions for Oil and Gas Leases with an Informed Bidder and a Random Reservation Price, Econometrica, vol. 62, no. 6, November 1994, 1415-1444

Hermalin, Benjamin E. (1992), The Effects of Competition on Executive Behavior, Rand Journal of Economics, vol. 23, no. 3, Autumn 1992, 350-365

Herring, Richard (1999), Credit Risk and Financial Instability, Oxford Review of Economic Policy, vol. 15, no. 3, 63-79

Hicks, J.R. (1935), Annual Survey of Economic Theory: The Theory of Monopoly, Econometrica, vol. 3, no. 1, January 1935, 1-20

Holmstrom, Bengt (1979), Moral Hazard and Observability, Bell Journal of Economics, vol. 10, no. 1, Spring 1979, 74-91

Holmstrom, Bengt (1982), Moral Hazard in Teams, Bell Journal of Economics, vol. 13, no. 2, Autumn 1982, 324-340

Holmstrom, Bengt (1999), Managerial Incentive Problems: A Dynamic Perspective, Review of Economic Studies, vol. 66, no. 1, January

1999, 169-182

Holmstrom, Bengt and Paul Milgrom (1987), Aggregation and Linearity in the Provision of Intertemporal Incentives, Econometrica, vol. 55, March 1987, 303-328

Holmstrom, Bengt and Paul Milgrom (1991), Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, Journal of Law, Economics, and Organization, vol. 7, 24-52

Holmstrom, Bengt and Jean Tirole (1989), The Theory of the Firm (Chapter 2), in: Schmalensee, Richard and Robert Willig (eds), Handbook of Industrial Organization, Volume 1, Amsterdam: North-Holland, 1989, 61-133

Holmstrom, Bengt and Jean Tirole (1997), Financial Intermediation, Loanable Funds, and the Real Sector, Quarterly Journal of Economics, vol. 112, no. 3, 663-691

Hommel, Ulrich and Hilmar Schneider (2003), Financing the German Mittelstand, EIB Papers, vol. 8, no. 2, 52-90

Hoppe, Heidrun and Ulrich Lehmann-Grube (2002), Specialization in Credit Markets, University of Hamburg manuscript, 28 January 2002

Hoppe, Heidrun and Ulrich Lehmann-Grube (2003), The Hotelling Model with Customer Screening, University of Bonn manuscript, December 2003

Horn, Henrik, Harald Lang and Stefan Lundgren (1994), Competition, Long-Run Contracts, and Internal Inefficiencies in Firms, European Economic Review, vol. 38, 213-233

Horn, Henrik, Harald Lang and Stefan Lundgren (1995), Managerial Effort Incentives, X-Inefficiency, and International Trade, European Economic Review, vol. 39, 117-138

Hughes, Joseph and Loretta Mester (1998), Bank Capitalization and Cost: Evidence of Scale Economies in Risk Management and Signaling, Review of Economics and Statistics, vol. 80, no. 2, May 1998, 314-325

Hughes, Joseph, Loretta Mester and Choon-Geol Moon (2001), Are Scale Economies in Banking Elusive or Illusive? Evidence Obtained by Incorporating Capital Structure and Risk-Taking into Models of Production, Journal of Banking and Finance, vol. 25, 2169-2208

Hyytinen, Ari (2003), Information Production and Lending Market Competition, Journal of Economics and Business, vol. 55, 233-253

International Monetary Fund (2000), Chapter VI: The Role of Foreign Banks in Emerging Markets, in: International Monetary Fund, International Capital Markets Report 2000, Washington, DC: International Monetary Fund, 2000, 152-182

International Monetary Fund (2001), Financial Sector Consolidation in Emerging Markets, Chapter Five in: International Capital Markets Report 2001, Washington, DC: 2001

Jagannathan, Ravi and Shaker B. Srinivasan (1999), Does Product Market Competition Reduce Agency Costs? North-American Journal of Economics and Finance, vol. 10, 387-399

Jagannathan, Ravi and Shaker B. Srinivasan (2000), Does Product Market Competition Reduce Agency Costs? NBER Working Paper Series, January 2000

James, Christopher (1996), RAROC Based Capital Budgeting and Performance Evaluation: A Case Study of Bank Capital Allocation, Wharton Financial Institutions Center Working Paper, 22 September 1996

Januszewski, Silke, Jens Köke and Joachim K. Winter (2002), Product Market Competition, Corporate Governance and Firm Performance: An Empirical Analysis for Germany, Research in Economics, vol. 56, 299-332

Jayaratne, Jith and Philip Strahan (1996), The Finance-Growth Nexus: Evidence from Bank Branch Deregulation, Quarterly Journal of Economics, vol. 111, no. 3, August 1996, 639-670

Jayaratne, Jith and Philip Strahan (1997), The Benefits of Branching Deregulation, Federal Reserve Bank of New York Economic Policy Review, vol. 3, no. 4, December 1997, 13-29

Jayaratne, Jith and Philip Strahan (1998), Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking, Journal of Law and Economics, vol. 41, no. 1, April 1998, 239-273

Jayaratne, Jith and Philip Strahan (1999), The Benefits of Branching Deregulation, Regulation, vol. 22, no. 1, 8-16

Jensen, Michael C. (1986), Agency Costs of Free Cash Flow, Corporate Finance and Takeovers, American Economic Review, vol. 76, no. 2, 323-329

Jensen, Michael C. (1993), The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems, Journal of Finance, vol. 48, no. 3, 831-880

Jensen, Michael C. and William H. Meckling (1976), Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, Journal of Financial Economics, vol. 3, 305-360

John, Kose, Anthony Saunders and Lemma Senbet (2000), A Theory of Bank Regulation and Management Compensation, Review of Financial

Studies, vol. 13, no. 1, Spring 2000, 95-125

Jorion, Philippe (2000), Risk-Management Lessons from Long-Term Capital Management, European Financial Management, vol. 6, no. 3, 277-300

Kim, Moshe, Eirik Gaard Kristiansen and Bent Vale (2005), Endogenous Product Differentiation in Banking Markets: What do Borrowers pay for? Journal of Banking and Finance, vol. 29, 681-699

Kirmße, Stefan, André Ehlerding and Annette Putzer (2001), Die verbundeinheitlichen Richtlinien für das Kreditgeschäft, Bankinformation, August 2001, 67-76

Kirmße, Stefan and Sven Jansen (2001), BVR-II-Rating: Das verbundeinheitliche Ratingsystem für das mittelständische Firmenkundengeschäft, Bankinformation, February 2001, 67-71

Kirmße, Stefan and Stefan Schweizer (2001), VR-CreditPortfolioManager - Portfolioorientierte Kreditrisikosteuerung für den genossenschaftlichen Verbund, Bankinformation, March 2001, 20-25

Klemperer, Paul (1999), Auction Theory: A Guide to the Literature, Journal of Economics Surveys, vol. 13, no. 3, 227-286

Krahnen, Jan Pieter and Martin Weber (2001), Generally Accepted Rating Principles: A Primer, Journal of Banking and Finance, vol. 25, 3-23

Kramer, Jost (2001), Ratingsysteme in Banken, Bankinformation, September 2001, 4-11

Krishna, Pravin (2001), On Competition and Endogenous Firm Efficiency, Economic Theory, vol. 18, 753-760

Kroon, Gerhard and Frank Pool (2002), Professionelle Kundengeschäftssteuerung mit Hilfe der DB-III-Rechnung, Bankinformation, April 2002, 26-34

Laeven, Luc and Giovanni Majnoni (2003), Loan Loss Provisioning and Economic Slowdowns: Too Much, Too Late? Journal of Financial Intermediation, vol. 12, 178-197

Lardy, Nicholas (1998), China's Unfinished Economic Revolution, Washington, DC: Brookings, 1998

Lazear, Edward P. (2000b), Performance Pay and Productivity, American Economic Review, vol. 90, no. 5, December 2000, 1346-1361

Lazear, Edward P. and Sherwin Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, Journal of Political Economy, vol. 89, no. 5, October 1981, 841-864

Leibenstein, Harvey (1966), Allocative Efficiency vs. 'X-Inefficiency', American Economic Review, vol. 56, no. 3, 392-415

Litan, Robert E., Paul Masson and Michael Pomerleano (eds) (2001), Open doors: foreign participation in financial systems in developing countries, Washington, DC: Brookings Institution Press, 2001

Llewellyn, David (2000), Some Lessons for Regulation from Recent Bank Crises, Open Economies Review, vol. 11 (S1), 69-109

Llewellyn, David (2002), An Analysis of the Causes of Recent Banking Crises, European Journal of Finance, vol. 8, no. 2, June 2002, 152-175

Loayza, Norman, Klaus Schmidt-Hebbel and Luis Servén (2000), What Drives Private Saving Across the World? Review of Economics and Statistics, vol. 82, no. 2, May 2000, 165-181

Löffler, Gunter (2003), The Effects of Estimation Error on Measures of Portfolio Credit Risk, Journal of Banking and Finance, vol. 27, 1427-1453

Loungani, Prakash (2001), How Accurate are Private-Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth, International Journal of Forecasting, vol. 17, 419-243

Lowe, Philip (2002), Credit Risk Measurement and Procyclicality, BIS Working Paper, September 2002

Lown, Cara, Donald Morgan and Sonali Rohatgi (2000), Listening to Loan Officers: The Impact of Commercial Credit Standards on Lending and Output, Federal Reserve Bank of New York Economic Policy Review, July 2000, 1-16

Mann, Fiona and Ian Michael (2002), Dynamic Provisioning: Issues and Application, Bank of England Financial Stability Review, December 2002, 128-136

Manove, Michael and A. Jorge Padilla (1999), Banking (Conservatively) with Optimists, Rand Journal of Economics, vol. 30, no. 2, Summer 1999, 324-350

Manove, Michael, A. Jorge Padilla and Marco Pagano (2001), Collateral vs. Project Screening: A Model of Lazy Banks, Rand Journal of Economics, vol. 32, no. 4, Winter 2001, 726-744

Marquez, Robert (2002), Competition, Adverse Selection, and Information Dispersion in the Banking Industry, Review of Financial Studies, vol. 15, no. 3, Summer 2002, 901-926

Martin, Stephen (1993), Endogenous Firm Efficiency in a Cournot Principal-Agent Model, Journal of Economic Theory, vol. 59, 445-450

Matthews, Steven (1984), Information Acquisition in Discriminatory Auctions, in: Boyer, M. and R.E. Kihlstrom (eds), Bayesian Models in Economic Theory, Amsterdam: North-Holland, 1984, 181-207

McAfee, R. Preston and John McMillan (1987), Auctions and Bid-

ding, Journal of Economic Literature, vol. 25, no. 2, 699-738

McCloskey, Deirdre and Stephen Ziliak (1996), The Standard Error of Regressions, Journal of Economic Literature, vol. 34, no. 1, March 1996, 97-114

McDonough, William (1998), Statement before the Committee on Banking and Financial Services, U.S. House of Representatives (also available at http:// www.newyorkfed.org/ newsevents/speeches/ 1998/mcd981001.html), New York: 01 October 1998

Mester, Loretta (1997), What's the Point of Credit Scoring? Federal Reserve Bank of Philadelphia Business Review, September 1997, 3-16

Milgrom, Paul and Robert Weber (1982), A Theory of Auctions and Competitive Bidding, Econometrica, vol. 50, no. 5, 1089-1122

Moody's Investors Service (2004), Special Comment: Default & Recovery Rates of Corporate Bond Issuers (January 2004), New York etc: Moody's, 2004

Mookherjee, Dilip (1984), Optimal Incentive Schemes with Many Agents, Review of Economic Studies, vol. 51, no. 3, July 1984, 433-446

Murillo-Zamorano, Luis (2004), Economic Efficiency and Frontier Techniques, Journal of Economic Surveys, vol. 18, no. 1, 33-77

Nadler, Paul (2003), Incentive Programs can make people do the wrong thing, American Banker, vol. 168, no. 165, 27 August 2003, 5

Nadler, Paul (2000a), Don't let incentive compensation beanball your bank, American Banker, vol. 165, no. 94, 16 May 2000, 5

Nadler, Paul (2000b), Incentive programs can have more risk than reward, American Banker, vol. 165, no. 123, 27 June 2000, 6

Nakamura, Leonard I. (1996), The Informational Impact of Competition in Loan Screening, Manuscript, Federal Reserve Bank of Philadelphia, August 1996

Nalebuff, Barry J. and Joseph E. Stiglitz (1983a), Information, Competition, and Markets, American Economic Review, vol. 73, no. 2, May 1983, 278-283

Nalebuff, Barry J. and Joseph E. Stiglitz (1983b), Prizes and Incentives: Towards a General Theory of Compensation and Competition, Bell Journal of Economics, vol. 14, no. 1, Spring 1983, 21-43

Neuberger, Doris (1998), Industrial Organization of Banking: A Review, International Journal of the Economics of Business, vol. 5, no. 1, 97-118

Nickell, Pamela, William Perraudin and Simone Varotto (2000), Sta-

bility of Rating Transitions, Journal of Banking and Finance, vol. 24, 203-227

Nickell, Stephen (1996), Competition and Corporate Performance, Journal of Political Economy, vol. 104, no. 4, August 1996, 724-746

Nickell, Stephen, Daphne Nicolitsas and Neil Dryden (1997), What makes firms perform well? European Economic Review, vol. 41, 783-796

Nowak, Helge (2002), Ratings für alle Segmente, Bankinformation, May 2002, 22-28

Oehler, Christian, Ulrike Volmar and Egbert Schark (2003), Datenpools für interne Ratings, Zeitschrift für das gesamte Kreditwesen, no. 9, 01 May 2003, 24-29

Oliver Wyman & Company (1996), Credit Pricing: Winning in Inefficient Markets, Oliver Wyman Report, vol. 6, no. 8, 1-4

Oliver Wyman & Company (2000), Themes: Pricing, www.oliverwyman.com/servlet/ShowPublish? page=per_themes_ detail&id=7, access date: last access date unknown; I have a pdf copy of the cited material

Ong, Michael (1999), Internal Credit Risk Models, London: Risk Books, 1999

Paal, Beatrix, Bruce Smith and Ke Wang (2005), Monopoly vs Competition in Banking: Some Implications for Growth and Welfare, University of Tokyo manuscript, 22 January 2005

Padilla, A. Jorge and Marco Pagano (1997), Endogenous Communication among Lenders and Entrepreneurial Incentives, Review of Financial Studies, vol. 10, no. 1, Spring 1997, 205-236

Padoa-Schioppa, Tommaso (2001), Bank Competition: A Changing Paradigm, European Finance Review (now: Review of Finance), vol. 5, 13-20

Parigi, Bruno (1998), Competition in banking: a survey of the literature (translated into Italian), in: Polo, M., Industria Bancaria e Concorrenza, Bologna: Il Mulino, 1998, 19-66

Park, Albert, Loren Brandt and John Giles (2003), Competition Under Credit Rationing: Theory and Evidence from Rural China, Journal of Development Economics, vol. 71, 463-495

Parkinson, Patrick (Associate Director, Division of Research and Statistics) (1999), Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management, Testimony before the Committee on Banking and Financial Services, U.S. House of Representatives (also available at http://www.federalreserve.gov/ boarddocs/testimony/ 1999/19990506.htm), Washington, DC: 06 May 1999

Pavcnik, Nina (2002), Trade Liberalization, Exit, and Productivity

Improvements: Evidence from Chilean Plants, Review of Economic Studies, vol. 69, 245-276

Peek, Joe and Eric Rosengren (2005), Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan, American Economic Review, vol. 95, no. 4, September 2005, 1144-1166

Persico, Nicola (2000), Information Acquisition in Auctions, Econometrica, vol. 68, no. 1, 135-148

Petersen, Mitchell A. and Raghuram G. Rajan (1995), The Effect of Credit Market Competition on Lending Relationships, Quarterly Journal of Economics, vol. 110, no. 2, May 1995, 407-443

President's Working Group on Financial Markets (1999), Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management (also available at http://www.ustreas.gov/press/releases/reports/hedgfund.pdf), Washington, DC: 28 April 1999

Purroy, Pedro and Vicente Salas (2000), Strategic Competition in Retail Banking under Expense Preference Behavior, Journal of Banking and Finance, vol. 24, 809-824

Radecki, Lawrence (1999), Banks' Payments-Driven Revenues, Federal Reserve Bank of New York Economic Policy Review, July 1999, 53-70

Raith, Michael (2003), Competition, Risk, and Managerial Incentives, American Economic Review, vol. 93, no. 4, September 2003, 1425-1436

Rajan, Raghuram and Luigi Zingales (1998), Financial Dependence and Growth, American Economic Review, vol. 88, no. 3, June 1998, 559-586

Rathmann, Christa (2002), Die BaFin schmiedet mit den 'MaK' ein heißes Eisen, Börsenzeitung, 17, 25 July 2002

Ridder, Thomas (2002), Portfoliosteuerung der Adressrisiken, Bankinformation, July 2002, 17-24

Riordan, Michael (1993), Competition and Bank Performance: A Theoretical Perspective, in: Mayer, Colin and Xavier Vives (eds), Capital Markets and Financial Intermediation, Cambridge: CUP, 1993, 328-343

Roe, Mark (2001), Rents and their Corporate Consequences, Stanford Law Review, vol. 53, 1463 et seq.

Roszbach, Kasper (2004), Bank Lending Policy, Credit Scoring, and the Survival of Loans, Review of Economics and Statistics, vol. 86, no. 4, November 2004, 946-958

Ruckes, Martin (2004), Bank Competition and Credit Standards, Review of Financial Studies, vol. 17, no. 4, 1073-1102

Rule, David (2001), The Credit Derivatives Market: Its Development

and Possible Implications for Financial Stability, Bank of England Financial Stability Review, June 2001, 117-140

Salop, Steven C. (1979), Monopolistic Competition with Outside Goods, Bell Journal of Economics, vol. 10, no. 1, Spring 1979, 141-156

Santomero, Anthony (1997), Commercial Bank Risk Management: An Analysis of the Process, Journal of Financial Services Research, vol. 12, no. 2/3, 83-115

Santomero, Anthony (2003), Process and Progress in Risk Management, Federal Reserve Bank of Philadelphia Business Review, First Quarter 2003, 1-5

Saunders, Anthony (1997), Financial Institutions Management, 2nd edition, Chicago etc: Irwin, 1997

Saunders, Anthony (1999), Credit Risk Measurement, New York etc: John Wiley & Sons, Inc., 1999

Scharfstein, David (1988), Product-Market Competition and Managerial Slack, Rand Journal of Economics, vol. 19, no. 1, Spring 1988, 147-155

Schargrodsky, Ernesto and Federico Sturzenegger (2000), Banking Regulation and Competition with Product Differentiation, Journal of Development Economics, vol. 63, 85-111

Schierenbeck, Henner (2001), Ertragsorientiertes Bankmanagement, Vol. 1 and 2 (a third volume consists of cases and is less interesting), Wiesbaden: Gabler, 7th edition, 2001

Schmidt, Klaus M (1997), Managerial Incentives and Product Market Competition, Review of Economic Studies, vol. 64, no. 2, April 1997, 191-213

Schnitzer, Monika (1999a), On the Role of Bank Competition for Corporate Finance and Corporate Control in Transition Economies, Journal of Institutional and Theoretical Economics, vol. 155, no. 1, March 1999, 22-46

Schnitzer, Monika (1999b), Enterprise Restructuring and Bank Competition in Transition Economies, Economics of Transition, vol. 7, no. 1, 133-155

Schranz, Mary (1993), Takeovers Improve Firm Performance: Evidence from the Banking Industry, Journal of Political Economy, vol. 101, no. 2, 299-326

Schure, Paul, Rien Wagenvoort and Dermot O'Brien (2004), The Efficiency and the Conduct of European Banks: Developments after 1992, Review of Financial Economics, vol. 13, 371-396

Shaffer, Sherrill (1994), Bank Competition in Concentrated Markets,

Federal Reserve Bank of Philadelphia Business Review, March/April 1994, 3-16

Shaffer, Sherrill (1998), The Winner's Curse in Banking, Journal of Financial Intermediation, vol. 7, 359-392

Shiller, Robert, Robert Gordon and Herschel Grossman (1991), Why are Prices Sticky? Discussion, American Economic Review, vol. 81, no. 2, May 1991, 97-100

Shleifer, Andrei (1985), A Theory of Yardstick Competition, Rand Journal of Economics, vol. 16, no. 3, Autumn 1985, 319-327

Shleifer, Andrei and Robert Vishny (1997), A Survey of Corporate Governance, Journal of Finance, vol. 52, no. 2, 737-783

Sjöström, Tomas and Martin L. Weitzman (1996), Competition and the Evolution of Efficiency, Journal of Economic Behavior and Organisation, vol. 30, 25-43

Small Business Administration (2004), FAQ: What is a small business? (also available on the internet at http:// app1.sba.gov/ faqs/faqindex.cfm? areaID=24), Washington, DC: 2004

Smith, R. Todd (1998), Banking Competition and Macroeconomic Performance, Journal of Money, Credit, and Banking, vol. 30, no. 4, November 1998, 793-815

Sobehart, Jorge, Sean Keenan and Roger Stein (2000), Benchmarking Quantitative Default Risk Models: A Validation Methodology, New York: Moody's Investors Service, 2000

Stahl, Dale O. (1988), Bertrand Competition for Inputs and Walrasian Outcomes, American Economic Review, vol. 78, no. 1, March 1988, 189-201

Stamp, David (2002), Market Prophets - Can Forecasters Predict the Financial Future? London etc: Pearson Education, 2002

Stennek, Johan (2000), Competition increases X-efficiency: A Limited-Liability Mechanism, European Economic Review, vol. 44, 1727-1744

Stiglitz, Joseph E. and Andrew Weiss (1981), Credit Rationing in Markets with Imperfect Information, American Economic Review, vol. 71, no. 3, June 1981, 393-410

Strahan, Philip (2003), The Real Effects of U.S. Banking Deregulation, Federal Reserve Bank of St. Louis Review, July/August 2003, 111-128

Sutton, John (1991), Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration, Cambridge, MA: MIT Press, 1991

Taylor, Lowell (1999), Commentary: Work Motivation, Federal Re-

serve Bank of St. Louis Review, vol. 81, no. 3, May / June 1999, 51-53

Thakor, Anjan V. (1996), Capital Requirements, Monetary Policy, and Aggregate Bank Lending: Theory and Empirical Evidence, Journal of Finance, vol. 51, no. 1, March 1996, 279-324

The Economist (2000a), Suckers for Nickels, The Economist, 08 January 2000

The Economist (2004f), Frustration, The Economist, 17 July 2004

The Economist (2005a), The Insidious Charms of Foreign Investment, The Economist, in: A Survey of India and China, 7-9, 05 March 2005

Treacy, William F. and Mark Carey (1998), Credit Risk Rating at Large U.S. Banks, Federal Reserve Bulletin, November 1998, 897-921

Treacy, William F. and Mark Carey (2000), Credit Risk Rating Systems at Large US Banks, Journal of Banking and Finance, vol. 24, 167-201

Tröge, Michael (2000a), The Insider's Curse, Northwestern University manuscript, 24 July 2000

Tröge, Michael (2000b), Information Acquisition, Specialization, and Credit Market Competition, Northwestern University manuscript, 25 July 2000

Tröge, Michael (2002), The Structure of the Banking Sector, Credit Screening, and Firm Risk, ESCP-EAP manuscript, 22 March 2002

Tybout, James R. (1992), Review of Caves, Richard E. and David R. Barton, Efficiency in US Manufacturing Industries, Cambridge, MA: MIT Press, 1990, Journal of Economic Literature, vol. 30, no. 1, March 1992, 185-187

Van Damme, Eric (1994), Banking: A Survey of Recent Microeconomic Theory, Oxford Review of Economic Policy, vol. 10, no. 4, 14-33

Vickers, John (1995), Concepts of Competition, Oxford Economic Papers, vol. 47, no. 1, January 1995, 1-23

Villas-Boas, J. Miguel and Udo Schmidt-Mohr (1999), Oligopoly with Asymmetric Information: Differentiation in Credit Markets, Rand Journal of Economics, vol. 30, no. 3, Autumn 1999, 375-396

Vives, Xavier (2001b), Competition in the Changing World of Banking, Oxford Review of Economic Policy, vol. 17, no. 4, 535-547

Voigt, Susan and Tammo Andersch (2003), Unternehmenskrisen und ihre Bewältigung-die Sicht der Banker, Die Bank, March 2003, 192-195

Walton, Richard (2000), International Comparisons of Profitability, Economic Trends, vol. 554, January 2000, 33-46

Wang, Cheng and Stephen Williamson (1998), Debt Contracts and Financial Intermediation with Costly Screening, Canadian Journal of

Economics, vol. 31, no. 3, August 1998, 573-595

Wiggers, Ferdinand (2002), Modell für Krisenfrüherkennung und Intensivbetreuung, Betriebswirtschaftliche Blätter, no. 08, August 2002, 363 et seq

Williamson, Oliver (1963), Managerial Discretion and Business Behavior, American Economic Review, vol. 53, no. 5, December 1963, 1032-1057

Willig, Robert (1987), Corporate Governance and Market Structure, in: Razin, Assaf and Efraim Sadka (eds), Economic Policy in Theory and Practice, Basingstoke: Macmillan, 1987, 481-494

Wilson, Thomas (1998), Portfolio Credit Risk, Federal Reserve Bank of New York Economic Policy Review, October 1998, 71-82

Winton, Andrew (1995), Delegated Monitoring and Bank Structure in a Finite Economy, Journal of Financial Intermediation, vol. 4, 158-187

Winton, Andrew (1997), Competition Among Financial Intermediaries when Diversification Matters, Journal of Financial Intermediation, vol. 6, 307-346

Wuffli, Peter and David Hunt (1993), Fixing the Credit Problem, McKinsey Quarterly, No. 2, 93-106

Yafeh, Yishay and Oved Yosha (2001), Industrial Organization of Financial Systems and Strategic Use of Relationship Banking, European Finance Review (now: Review of Finance), vol. 5, 63-78

Yanelle, Marie-Odile (1989), The Strategic Analysis of Intermediation, European Economic Review, vol. 33, 294-301

Yanelle, Marie-Odile (1997), Banking Competition and Market Efficiency, Review of Economic Studies, vol. 64, no. 2, April 1997, 215-239

Zarutskie, Rebecca (forthcoming), Evidence on the Effects of Bank Competition on Firm Borrowing and Investment, Journal of Financial Economics,

Ziliak, Stephen and Deirdre McCloskey (2004), Size Matters: The Standard Error of Regressions in the American Economic Review, Journal of Socio-Economics, vol. 33, 527-546