Discussion Papers in Economics

Race to collusion: Monitoring and incentive contracts for loan officers under multiple-bank lending

Kanishka Dam and Prabal Roy Chowdhury

July, 2020

Discussion Paper 20-05

Indian Statistical Institute, Delhi
Economics and Planning Unit
7, S. J. S. Sansanwal Marg, New Delhi, 110016, India

Race to collusion: Monitoring and incentive contracts for loan officers under multiple-bank lending*

Kaniṣka Dam† Prabal Roy Chowdhury‡

July 8, 2020

Abstract

The possibility of vertical collusion between an informationally opaque borrower and corruptible loan officers, to whom the task of monitoring is delegated, in bank-loan officer-borrower hierarchies shapes the incentive contracts for the loan officers. Collusive threats exacerbate incentives for over-monitoring, and make monitoring efforts of the banks strategic complements because of a novel 'race-to-collusion' effect—a hitherto unexplored effect of multiple-bank lending. Thus, delegation contract solves the free-riding problem in the presence of monitoring duplication, and may lead to higher levels of per-bank monitoring in multiple-bank lending. Moreover, over-monitoring, albeit inefficient relative to the optimal contract in the absence of race-to-collusion, may enhance social welfare. We further show that the collusion-proofness principle may fail to hold under multiple-bank lending.

*We are grateful to Karna Basu, Fahad Khalil, Dilip Mookherjee, David Pérez-Castrillo, Ashok Rai, Uday Rajan, Sudipta Sarangi, Jaume Sempere, Rajdeep Sengupta, Kostas Serfes, Ricardo Serrano-Padial, Thierry Verdier, and the participants at the Mexico City Economic Theory Conference, 29th Annual Conference on Contemporary Issues in Development Economics, and 15th Annual Conference on Economic Growth and Development for insightful comments.

†Center for Research and Teaching in Economics (CIDE), 3655 Carretera Mexico-Toluca, Lomas de Santa Fe, 01210 Mexico City, Mexico. E-mail: kaniska.dam@cide.edu

‡Indian Statistical Institute, Delhi Center. 7 Sahid Jit Singh Sansanwal Marg, New Delhi, India. E-mail: prabalrc@isid.ac.in

1 Introduction

In recent years, borrowing from multiple lenders, such as lending consortia and syndicates, has been pervasive. Ongena and Smith (2000) document that more than 85% of the firms across twenty European countries tend to maintain multiple lending relationships. Detragiache, Garella, and Guiso (2000) and Farinha and Santos (2002) find similar results in relation to small business lending. However, a strand of literature suggests, in a scenario where informationally opaque borrowers require monitoring and due diligence,[1] that lending relationships with multiple investors may lead to lower level of monitoring because of the free-riding problem and costly monitoring duplication (e.g. Sufi, 2007).

The main objective of our paper is to develop a framework for financial intermediation that can reconcile these apparently conflicting pieces of evidence. To this end, we analyze a one-period model with one firm and multiple lenders in which mis-behaving by the firm can only be prevented by costly monitoring. The novel feature of our model is that (unlike e.g. Winton, 1995), monitoring activity is delegated to loan officers. This in turn opens up the possibility of collusion between the borrower and loan officers.[2] We establish that the likelihood of such *vertical collusion* generates some hitherto unexplored effects, in particular what we term the *race-to-collusion* effect. Further, we shall argue that this effect ensures that in certain cases per-bank monitoring under two-bank lending can exceed that under single-bank lending.

The firm, that owns a constant-returns-to-scale technology project, borrows startup capital from bank(s) at a given loan rate, and we compare and contrast two cases, single-bank and two-bank lending. This firm can either be diligent (high effort), or mis-behave (low effort), with the effort level being non-observable by the banks in the absence of any monitoring. In case the borrower is diligent, the project generates a verifiable but random cash flow, but there is no private benefit. On the other hand, if the borrower mis-behaves, then the borrower has a non-verifiable private benefit, but there are no verifiable returns from the project. Further, the private benefit to the firm in case of mis-behaving exceeds its return from being diligent, so that in the event of non-monitored lending the firm necessarily mis-behaves, and given that in this case there is no verifiable return, the banks obtain no repayment and fail to break-even. Thus non-monitored lending is not feasible. Banks have no access to any monitoring technology however and thus the task of monitoring is delegated to loan officers. Berger and Udell (2002) recognize the importance of loan officers in producing [soft] information in the context of small business (SME) lending. In fact they argue that "the relevant relationship in SME lending is the loan officer-borrower relationship, not the bank-entrepreneur relationship".

Entrusting loan officers with the responsibility of monitoring can however be problematic in that these agents may collude with the borrowing firm, and mis-report (e.g. Stein, 2002; Liberti and Mian, 2009). While the monitoring level is contractible, its outcome, i.e., whether monitoring has been successful, or not, is not verifiable. In the event that a loan officer succeeds in monitoring, he can either make the firm behave, or, in exchange for a bribe, allow the firm to mis-behave (in certain cases), reporting to his employer that he has failed in his monitoring. Thus, the optimal contract involves each bank designing

[1] Hölmstrom and Tirole (1997) show that informational opacity is severe with low-net worth entrepreneurs, and hence, they are not able to secure non-monitored finance.

[2] There is a plethora of evidence on collusion among loan officers and borrowers. Winton (1995) cites several press reports in the 90's on graft on behalf of monitoring agents that had led to large investor losses. Hertzberg, Liberti, and Paravisini (2010) find evidence of loan officer incentives to misreport borrower behavior in a multinational U.S. bank lending to small and medium sized enterprises (SMEs) in Argentina, and show that a rotation policy that reassigns loan officers to borrowers is able to mitigate such incentive problems in the sense that loan officer reporting tends to be more accurate when rotation is anticipated. Uchida, Udell, and Yamori (2012) find similar evidence for Japanese banks in relation to SME lending.

an incentive scheme (a share of repayment) for its loan officer such that he has no incentive to engage in bribery. One interesting feature of this framework is that in case of multiple-bank lending the optimal contract between one bank and its loan officer depends on that of the other, i.e., contracts are subject to externality. Using a parsimonious model, we answer the following relevant questions pertaining to the internal organization of a bank under multiple-bank lending:

(A) How does the possibility of vertical collusion shape incentive contracts for loan officers in a bank?

(B) Building on (A), does multiple-bank lending elicit higher monitoring efforts and induce banks to provide stronger incentives to their loan officers relative to single-bank lending?

The optimal contract under both monopoly, as well as two-bank lending consists of a stipulated monitoring effort and a share of repayment. Under either scenario, the contractual structure is critically affected by the presence of collusion possibilities, with each bank essentially maximizing the surplus arising out of the bank-loan officer relationship subject to a *no-collusion constraint* that ensures that the loan officer has no incentive to collude, and a *participation constraint* for the loan officer.[3] The incentive pay of the loan officer plays a dual role here. An increase in incentive pay not only relaxes the participation constraint, thus allowing the bank to mandate higher levels of monitoring, it makes collusion less attractive. Given these similarities in the problem structure, naturally the optimal contract under both forms of lending exhibit certain similarities, though our focus is on the differences between the two. We start by describing the similarities.

Let firm quality denote the firm's expected per unit return (gross of interest payments) from behaving diligently net of her per unit payoff from mis-behaving. Clearly, when firm quality is high (less informationally opaque borrower), the firm has little to gain from mis-behaving, and hence, little incentive to collude with the loan officer(s) as well. Moreover, the participation constraint ensures that optimally the incentive pay of the loan officer cannot be too small, so that the return from truthful reporting is high. Taken together, these two facts ensure that the no-collusion constraint does not bind at the optimum, and hence, loan officer incentives become perfectly aligned with those of the banks. As a result, the outcome coincides with the non-delegation outcome, i.e., the case when each bank would have directly monitored the borrower. Collusion however becomes a more serious issue when firm quality is low, so that the potential gains from mis-behaving, and therefore from collusion, is high. Thus the incentive pay must be increased beyond non-delegation levels so that the no-collusion constraint binds. This in turn ensures that the participation constraint is slack at the non-delegation level of monitoring, so that the bank can enforce higher levels of monitoring. Consequently, under both lending modes, we observe inefficient *over-monitoring* relative to the non-delegation case when firm quality is low.

We next turn to comparing the outcomes under one- and two-bank lending. Our central result is that, relative to single-bank lending, two-bank lending yields lower monitoring efforts and weaker incentives when firm quality is high. By contrast, when firm quality is low, the results are reversed—two-bank lending induces higher monitoring efforts, as well as stronger incentives. We further argue that this result can be traced to two countervailing effects that emerge under two bank lending—namely, a *free-riding* effect and a *race-to-collusion* effect.

First consider the case when firm quality is high, so that the firm and loan officers have little incentive to collude. Thus the no-collusion constraint does not bind. Moreover, reports by loan officers are public

[3]In Section 6.1 later, we examine if imposing the no-collusion constraint(s) is without loss of generality.

information, allowing each bank to free-ride on the monitoring done by the other bank. Consequently, individual monitoring efforts become strategic substitutes which leads to the under-provision of monitoring in the two-bank lending equilibrium (similar to Carletti, 2004; Khalil, Martimort, and Parigi, 2007). Given that collusive considerations play no role in this case, two-bank lending induces lower monitoring effort relative to single-bank lending because of the negative externality arising from the free-riding effect.

We then consider the case when firm quality is low. Interestingly, in this case the individual monitoring efforts become strategic complements, which we call the race-to-collusion effect. This arises due to the following reason. Each bank faces a trade-off between extracting additional rent from its loan officer and providing incentives to deter collusion. With an increase in monitoring by any loan officer, the participation constraint of the other bank's loan officer gets relaxed. But the concerned bank cannot extract the additional rent by lowering incentive pay because, with firm quality being low the no-collusion constraint of each loan officer binds, and consequently, lowering incentive pay would violate the no-collusion constraint. The only way to exploit the relaxation of the participation constraint is to increase the loan officer's monitoring effort. An increase in the monitoring effort of any loan officer therefore induces an increase in the effort of the other. As firm quality decreases, the strategic complementarity effect becomes stronger (steeper best reply functions), so that when firm quality is low, the race-to-collusion effect dominates the free-riding effect. As a result, two-bank lending leads to over-provision of monitoring efforts relative to single-bank lending.

We then perform two conceptual exercises in a bid to disentangle the two aforementioned effects. First, consider a scenario where the two banks merge, maximizing joint profits so that the free-riding channel is shut down, but employing two agents so that the race-to-collusion effect is still present. As expected, the absence of free-riding ensures that both per-bank monitoring as well as incentives are higher in this case relative to the baseline two-bank lending framework in which banks act independently. Second, we consider a scenario where the two loan officers coordinate their activity, in that they jointly bargain with the firm over the bribe amount in case collusion becomes feasible. Thus in this case, the race-to-collusion effect is absent, and not surprisingly we find that both incentives and per-bank monitoring are lower relative to the baseline two-bank framework.

We finally extend the analysis in several directions. First, recall that we solve for the equilibrium under the assumption that the *collusion-proofness principle* (Laffont and Rochet, 1997) holds, i.e., imposing the no-collusion constraints is without loss of generality. We demonstrate that, as long as firm quality is sufficiently high, this is indeed true i.e., the collusion-proofness principle does hold. Second, we endogenize the loan rates, using the zero-profit condition for each bank (competitive banking market). We find that two-bank lending implies a higher loan rate if and only if firm quality is low. Moreover, we find that our central results—namely, over-monitoring resulting from binding no-collusion constraints if borrower quality is low, and second, that monitoring is higher (lower) under two-bank lending when firm quality is low (high), remain valid.

2 Related literature and our contribution

The extant literature has provided diverse explanations for multiple lending. von Thadden (1992) and Padilla and Pagano (1997) argue that the presence of multiple investors can reduce the ability of any single lender to holdup. Dewatripont and Maskin (1995) argue that multiple-bank lending may be a way to make re-financing of unprofitable projects more complicated, thereby ameliorating the soft budget-constraint problem. Bolton and Scharfstein (1996) demonstrate that multiple borrowing makes debt renegotiation more complicated, thereby reducing strategic default by entrepreneurs. It has also been argued that multiple-bank lending serves as an instrument to diversify loan risks (e.g. Carletti, Cerasi, and Daltung, 2007). Finally, Detragiache et al. (2000) argue that financial arrangement with many banks provides protection against information loss following one bank having distress.[4] While, like this literature, the present paper also finds that multiple-bank lending can improve firm performance in the presence of information opacity, it differs from this literature in several respects. First, the channel via which this operates, namely the race-to-collusion effect is novel to the literature on multiple-bank lending. Second, multiple-bank lending enhances monitoring; however, it comes at the expense of efficiency as the presence of many banks exacerbates incentives for over-monitoring. This is in contrast to the aforementioned literature, where multiple-bank lending decreases the fundamental inefficiency, e.g. re-financing of unprofitable projects (as in Dewatripont and Maskin, 1995).

The paper closest in spirit to ours is by Carletti et al. (2007) who analyze a model with multiple lenders. There is double-sided moral hazard, both in the effort exerted in by the borrowers, and in the level of bank monitoring. Multiple-bank lending induces a trade-off between the benefits of greater diversification and costs arising from free-riding and monitoring duplication. Whenever the first effect is stronger, which happens for example if the monitoring costs are sufficiently large, and/or inside equity is small, multiple-bank lending induces over-provision of per-project monitoring compared with single-bank lending. Our paper adds to the literature by showing that multiple-bank lending can induce higher monitoring even when project returns are correlated, so that there is little diversification. We argue that this happens when collusion between the borrower and loan officers is a serious concern, via the race-to-collusion effect. Thus, our paper is complementary to Carletti et al. (2007), showing that over-monitoring can happen under diverse scenarios.

Prior to Carletti et al. (2007), both Carletti (2004) and Khalil et al. (2007) analyze the effect of multiple-bank lending on monitoring incentives.[5] In Carletti (2004) the central issue is of *ex ante* moral hazard, with monitoring aimed at preventing borrower misbehavior, whereas in Khalil et al. (2007) the central problem is of costly state verification. Further, while Carletti (2004) aims at explaining the endogenous choice of alternative lending modes by a borrower, Khalil et al. (2007) focus on whether, under multiple-bank lending, equilibrium contracts exhibit debt-like properties (as in Winton, 1995). Both these papers though make the same fundamental point that, in the presence of multiple investors, free-riding due to strategic interaction among banks leads to under-provision of monitoring.[6] The present paper however differs from both the aforementioned works in that we explicitly take into account agency problems, in particular collusion possibilities, in bank-monitor-borrower hierarchies. This induces an interplay between the free-riding effect unearthed in Carletti (2004) and Khalil et al. (2007), and collusion incentives for the loan officers and the borrower. Interestingly, Khalil et al. (2007) also demonstrate that in case the principals coordinate on the amount of monitoring, but not on the transfers, then there may be over-monitoring relative to the case when the principals merge. By contrast, we examine a scenario

[4]While this literature mostly focuses on informational issues, by way of contrast Parlour and Rajan (2001) sustain multiple lending in a framework with complete information.

[5]While there are other papers that talk of lender monitoring, some assume that information acquisition about borrowers is a by product of lending (e.g. Padilla and Pagano, 1997), or that the magnitude and cost of monitoring is exogenously fixed (e.g. von Thadden, 1992).

[6]While Diamond (1984) makes a fundamental point that non-cooperative contracting can lead to over-monitoring, neither Diamond (1984), nor much of the literature following from it explicitly allow for strategic interactions among lender, and the consequent contractual externalities. Of course, notable exceptions include Carletti (2004) and Khalil et al. (2007).

where there is no coordination among the principals, either in monitoring or in incentive pay, with the result being driven by collusive possibilities.

The analysis of collusion in hierarchies goes back to Tirole (1986), who argued that while collusion possibilities among supervisors and agents creates possibilities for additional inefficiencies, employing a supervisor can still be beneficial. Kofman and Lawarée (1993) extend this analysis to allow for external auditors who may have less information relative to internal auditors, but are more honest and can keep a check on collusive behavior. The present paper, while related, differs in that the monitors are symmetric, and both can prevent collusion by the other monitor if he is honest herself. Another related paper is Mookherjee and Png (1995) who study the issue of how to compensate corruptible law enforcers, finding that the collusive possibilities may generate non-obvious results, in particular that it can lead to over-monitoring. Mookherjee and Png (1995) differ from the present paper in that they do not allow for multiple law enforcers as that is not their focus. Moreover, in the theoretical literature on collusion, there is an increasing focus on the "collusion-proofness principle", i.e., whether one can, without loss of generality, restrict attention to equilibrium contracts that involve no collusion. While we discuss this literature in greater details later in Section 6.1, here we mention some important papers in this literature, Laffont and Rochet (1997), Laffont and Martimort (2000) and Tressel and Verdier (2011), among others.

Finally, the issue of multiple-bank lending is related to a broader literature on non-exclusive contracts (e.g. Kahn and Mookherjee, 1998; Parlour and Rajan, 2001), with the central theme that such contracts impose an externality on the other agents. Externality across contractual relationships also arises in the present paper since any changes in the incentives being offered to one agent affects repayment performance, thus impacting the payoffs of the other bank, and also the other loan officer. The present paper however extends the literature on non-exclusive contracts in several dimensions. First, it allows for a three-layered hierarchy, rather than a two-layered one. Second, it analyzes lending without collateral.

3 The Model

The economy, which spans five dates $t = 1, \ldots, 5$, consists of three classes of risk neutral agents—a firm (borrower), two banks (lenders), and two loan officers (monitors). The firm owns a risky project: $I > 0$ dollars invested in the project yield a verifiable cash flow of yI (with $y > 1$) if it succeeds, and nothing if the project fails. The probability of success depends on borrower behavior. If she (or the manager overseeing the firm) behaves diligently (works), then the project succeeds with probability $p \in (0, 1)$. On the other hand, if the firm misbehaves (shirks), the project fails (probability of success is zero), but it generates a non-verifiable private benefit $v > 0$ per dollar invested. These actions are not publicly verifiable, and hence, there is moral hazard at the firm level.[7] We assume that the project is economically viable only when the firm is diligent, i.e., $py > 1 > v$. Let $z \equiv py - v$ represent 'firm quality', which can be low either due to low expected per unit cash flow py [even if the firm is diligent], or because the private moral hazard v is high.

The borrower does not have any fund of her own, and hence, the investment must be externally funded. Each bank invests \$1 in the firm at an exogenously given loan rate $r \geq 1$.[8] The banks have a

[7]Following Hölmstrom and Tirole (1997), an equivalent interpretation is that the borrower can choose between two versions of the project—a "good project" that yields yI with probability p_H and no private benefits, and a "bad project" that yields yI with probability p_L and a positive private benefit v, where $p_H = p > 0$ and $p_L = 0$.

[8]In Section 6.2, we endogenize loan rates allowing for the possibility that the charge different loan rates.

per unit opportunity cost of capital equal to 1. We assume $v \geq p(y-r)$, i.e., $pr \geq z$, so that the firm has incentives to shirk by diverting funds to private uses. Given that borrower actions cannot be contracted upon, non-monitored lending is not feasible because the firm will necessarily misbehave and the banks will not break even. In other words, bank monitoring brings down the extent of borrower moral hazard. Formally, if the firm is monitored at an intensity $\pi \in [0, 1]$, then with probability π the borrower can be obliged to behave diligently (e.g. Hölmstrom and Tirole, 1997; Repullo and Suárez, 2000). Monitoring in the present context is what Aoki (1994) classifies as *interim monitoring*, which is an instrument to ameliorate [borrower] moral hazard,[9] as opposed to *ex-ante monitoring* or *screening* which is meant to improve upon adverse selection problems, or *ex-post monitoring* that is used to verify the [hidden] financial states of the borrowing firm.[10]

Banks however do not possess any monitoring technology, and hence, must delegate monitoring to loan officers.[11] tThe (aggregate) monitoring intensity π thus depends on the individual monitoring efforts $m_i, m_j \in [0, 1]$ exerted by loan officers i and j, respectively. In particular, the aggregate monitoring function is given by:

$$\pi(m_i, m_j) = 1-(1-m_i)(1-m_j) = m_i + m_j - m_i m_j, \tag{1}$$

which is increasing in individual monitoring effort because $\pi_i(m_i, m_j) = 1 - m_j \geq 0$. However, increased effort by one loan officer decreases the marginal benefit of monitoring by the other because $\pi_{ij}(m_i, m_j) = -1 < 0$, suggesting that free-riding may arise in equilibrium. For loan officer i, m_i comes at a cost of

$$C(m_i) = \frac{1}{2} c m_i^2,$$

where $c > 0$. We assume that

Assumption 1 $\quad 1 + \frac{c}{2} \leq pr \leq c.$

The upper bound on the expected loan rate guarantees that, in case the banks did possess the monitoring technology themselves, the optimal monitoring effort is less than 1 (see Proposition 1 later). Whereas the lower bound implies that the expected repaid amount must cover the cost of capital plus the maximum monitoring cost, i.e., $C(1)$.[12] We normalize the outside option of the loan officer to 0.

We assume that the individual monitoring efforts m_i and m_j can be verified by the banks, and hence, can be contracted upon. In many situations, a bank's internal rules specify tasks for the loan officers which are directly correlated with their monitoring effort. Analyzing the empirical relationship between borrower collateral and monitoring for a large Swedish bank, Cerqueiro, Ongena, and Roszbach (2016) measure bank monitoring by the intervals of reviews of borrower quality and collateral. Gustafson, Ivanov, and Meisenzahl (2020), in a study of the U.S. syndicated loan market, measure monitoring

[9] An alternative interpretation is the following (see Boot and Thakor, 2000). The payoff structure of the project when the borrower shirks may be thought of the one under *transaction lending*. By contrast, *relationship lending* (when the borrower is monitored at a strictly positive intensity π) enhances the probability of success relative to transaction lending, i.e., $\pi p > 0$ which reaches its full potential p when $\pi = 1$.

[10] All our results hold good in the context of costly state verification in which monitoring is of ex-post nature.

[11] The assumption that banks do not own a monitoring technology may mean that bank managers, who oversee many departments including a bank's lending division, do not often possess specialized skills (such as the knowledge of regulation, passing licensing tests, etc.) that the loan officers do.

[12] A necessary condition for the inequalities in Assumption 1 to hold simultaneously is that $c > 2$.

frequency by the maximum number of times a given loan is invigilated within a year [during the life of the loan]. Thus, the empirical measures of monitoring not only suggest that bank monitoring many a time is verifiable and measurable, but also is an activity that is performed in the interim stage of a lending relationship. Although monitoring *level* is assumed to be verifiable, and hence, can be specified by the bank, the *outcome* of the monitoring process, i.e., whether it has been successful or not is not verifiable by the banks. By 'successful monitoring' we mean that a loan officer has been able to detect borrower (mis)behavior, and is in a position to take appropriate actions to make her behave diligently.[13] Non-observability of the monitoring outcome opens up the possibility of collusion between the concerned loan officer and the firm. When a loan officer collude with the borrowing firm, he lets the borrower shirk (and enjoy the private benefit) in exchange of a bribe, and reports to his employer that he has not learnt anything about the actions of the firm. This is typical incentive problem that stems from delegation contracts (between a bank and its loan officers). Formally, a contingent contract for loan officer i specifies the required monitoring level $m_i \in [0, 1]$ as well as a share $s_i \in [0, 1]$ of pr, the amount repaid to the bank. We assume that the delegation contract is exclusive, i.e., the banks hire a loan officer apiece, and each loan officer can work for only one bank.[14] Note that in the present framework, monitoring is a public good, in that if one loan officer is successful in detecting borrower (and decides not to collude), then the borrower behaves diligently so that the other bank benefits as well. The possibility of collusion between the loan officer(s) and the borrower and its effect on monitoring will be the central issue of our paper. Because monitoring effort is contractible, the sole objective of offering the incentive contract is to deter collusion. One of the main focuses of our paper is to study the strategic effects of incentive contracts on bank monitoring under multiple-bank lending, and hence, we first analyze the collusion-free contracts. In Section 6.1, we examine if restricting attention to collusion-free equilibria is without loss of generality or not.

The timing of events is as follows. At $t = 1$, each bank lends \$1 to the firm. At date 2, the banks employ loan officers, with banks i and j offering contracts (m_i, s_i) and (m_j, s_j) to loan officers i and j, respectively. At $t = 3$, the loan officers monitor at the level specified in the contract. At $t = 4$, the firm and the successful loan officers, if any, decide whether to collude. In case at least one loan officer is successful and decides not to collude, the borrower is diligent. Otherwise, the borrower misbehaves. At date 5, the cash flow from the project is realized and the agreed upon payments are made.

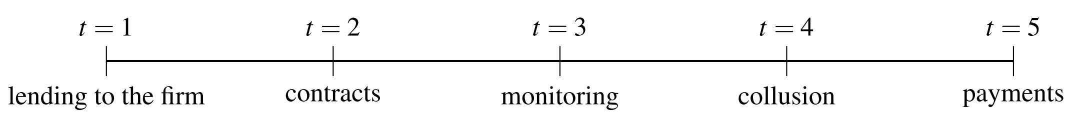

Figure 1: *The timing of events.*

The information structure at date 4 is important. Collusion between the borrower and a loan officer is feasible only if the concerned loan officer has been successful in monitoring. We assume that the outcome of the individual monitoring effort is private information, i.e., one loan officer does not observe whether the other one has succeeded or failed in detecting firm behavior. There are the following three

[13]In Appendix B, we analyze an extension where the monitoring level is not contractible (as in Mookherjee and Png, 1995).

[14]A loan officer is typically an employee of a bank, and is not allowed to moonlight in other lending companies. Regulation in many U.S. states prohibits mortgage loan officers (MLOs) to work for more than one mortgage banker or broker (see, for example, Section 8 of the MLO licensing guidelines of the State of New York, available at `https://www.dfs.ny.gov/apps_and_licensing/mortgage_companies/mlo_guide`).

possible scenarios:

1) *Neither loan officer succeeds:* The borrower misbehaves and consumes $2v$, and collusion between the borrower and loan officer(s) does not take place;

2) *Only one of the two loan officers succeeds:* This loan officer, say i, can either take actions in order to make the firm being diligent or collude with the borrower. In case the borrower being diligent, she obtains $2p(y-r)$ and the concerned loan officer receives prs_i. On the other hand, in the case of collusion, the borrower enjoys private benefit v, and the loan officer receives bribe b_i;

3) *Both loan officers succeed:* There are two possibilities—(a) either one or both loan officers behave honestly, and the firm behaves diligently, (b) or both loan officers collude with the borrower, let her enjoy private benefit $2v$, and they receive bribes b_i and b_j from her.

4 Optimal loan officer incentives and monitoring

4.1 Single-bank lending

We first consider the case when there is only one bank that lends $1 to the firm.

4.1.1 Direct bank monitoring without delegation

As a benchmark, we first analyze the optimal monitoring if the bank had access to the same monitoring technology $C(\cdot)$ as the loan officer, and had chosen to directly monitor the firm. The optimal monitoring level solves

$$\max_{m\in[0,1]} B(m) \equiv mpr - \frac{1}{2}cm^2 - 1. \tag{$\mathscr{M}^*$}$$

The solution to the above maximization problem (derived from its first-order condition) is characterized as follows:

Proposition 1 *When the single bank lends to the borrower and can monitor her directly, the optimal monitoring intensity is given by:*

$$m_1^*(r) = \frac{pr}{c}.$$

Clearly, $m_1^*(r)$ is less than 1 by Assumption 1. The optimal level of monitoring is independent of firm quality z. This is because in the absence of any delegation, the bank does not face any incentive problem arising from collusive possibilities. As we shall later find, collusion incentives do however depend on firm quality.

4.1.2 Delegated monitoring

We now turn to the case where the bank has no access to the monitoring technology, and must delegate monitoring duties to a loan officer. Let m and s denote respectively the monitoring effort and share of repayment. As is usual, we solve this game backwards.

The bribery subgame: incentives to deter collusion At date 4, collusion between the firm and the loan officer is feasible only if the loan officer is successful in detecting borrower behavior. At this stage, the loan officer can either choose to report truthfully, or he can collude with the borrower in exchange for a bribe b and report to the bank that he has not learnt anything. We follow much of the literature on collusion (e.g. Kofman and Lawarée, 1993) in abstracting from the minutiae of any particular bargaining protocol involved in the collusion process, and instead focus on deriving a no-collusion constraint that ensure that collusion is not possible under any such protocol. To that end, we derive a condition that the aggregate surplus for the loan officer and the firm is lower in case there is collusion, so that one cannot sustain a positive amount of bribe, whatever may be the bargaining protocol or level of bribe.

Let $F(\sigma)$ and $M(\sigma)$ respectively be the payoffs of the firm and the monitor where $\sigma \in \{0, 1\}$ with 0 representing *collusion* and 1 representing *no-collusion*. The payoffs are given by:

$$F(0) = v - b, \quad \text{and} \quad M(0) = b - C(m);$$
$$F(1) = p(y - r), \quad \text{and} \quad M(1) = prs - C(m).$$

Thus, collusion is not feasible if and only if

$$F(1) + M(1) \geq F(0) + M(0) \quad \Longleftrightarrow \quad prs \geq pr - z. \tag{NC}$$

The above *no-collusion constraint*, (NC) takes a simple form—the (expected) incentive pay of the loan officer, prs must exceed the stake of collusion, $v - p(y - r) = pr - z$. Clearly, in a collusion-free equilibrium, the optimal contract must satisfy the no-collusion constraint, which we would refer to as the *collusion-free contracts*.

Optimal loan officer incentives and monitoring At $t = 2$, the bank would offer a contract (m, s) which, apart from satisfying (NC), must also satisfy (i) the *participation constraint* of the loan officer

$$mprs - \frac{1}{2}cm^2 \geq 0, \tag{PC}$$

and (ii) the *feasibility constraint*

$$(m, s) \in [0, 1] \times [0, 1]. \tag{F}$$

Note that the monitor's contract (m, s) is subject to *limited liability* as well, in that he obtains a payment only if the project yields a verifiable cash flow. We denote by $\mathscr{F}$ the set of *feasible contracts*, i.e., the set of contracts that satisfies (PC), (NC) and (F). The bank solves the following maximization problem:

$$\max_{(m,s)\in\mathscr{F}} B(m, s) \equiv mpr(1 - s) - 1. \tag{$\mathscr{M}$}$$

The optimal monitoring intensity and share of repayment for the loan officer are characterized as follows.

Proposition 2 *There are threshold values of firm quality, z_1^{min}, z_1^0 and $\bar{z}_1$, with $0 < z_1^{min} < z_1^0 < \bar{z}_1 < pr$, where z_1^{min} is such that bank-lending is not feasible for any $z \leq z_1^{min}$.*

(a) *The optimal monitoring intensity is given by:*

$$m_1(z, r) = \begin{cases} 1 & \textit{for } z \in [z_1^{min}, z_1^0], \\ \frac{2(pr-z)}{c} & \textit{for } z \in (z_1^0, \bar{z}_1], \\ \frac{pr}{c} & \textit{for } z \in (\bar{z}_1, pr]. \end{cases}$$

Thus, whenever firm quality is low, i.e., $z_1^{min} \leq z \leq \bar{z}_1$, the equilibrium outcome involves over-monitoring relative to the non-delegated level of monitoring $m_1^*(r)$. Whereas the monitoring level is at the non-delegated level of monitoring $m_1^*(r)$ whenever firm quality is high, i.e., $z > \bar{z}_1$.

(b) *The optimal share of repayment for the loan officer, on the other hand, is given by:*

$$s_1(z,r) = \begin{cases} 1 - \frac{z}{pr} & \text{for } z \in [z_1^{min}, \bar{z}_1], \\ \frac{1}{2} & \text{for } z \in (\bar{z}_1, pr]. \end{cases}$$

The preceding results are depicted in Figure 2.

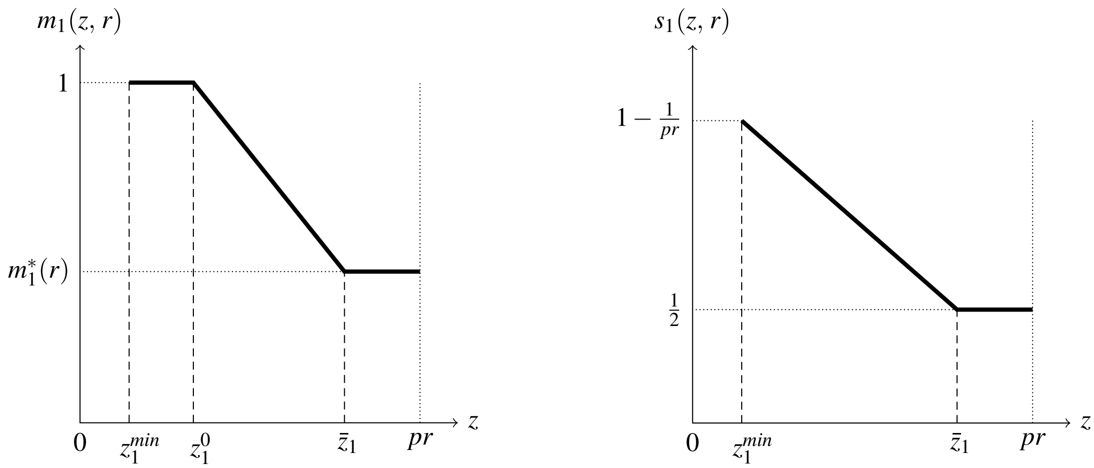

Figure 2: *Equilibrium monitoring and share of the loan officer as functions of firm quality. For $z < z_1^{min}$, bank-lending is not feasible. For $z \in [z_1^{min}, \bar{z}_1]$, the no-collusion constraint binds, and there is over-monitoring relative to the non-delegated level $m_1^*(r)$. For For $z > \bar{z}_1$, the no-collusion constraint does not bind and monitoring is at the non-delegated level.*

Proposition 2 is intuitive. In an optimal contracting model under limited liability such as ours, the typical trade-off a bank faces is between incentive provision (setting a high share of repayment in order to deter collusion) and rent extraction from the loan officer (setting a low share as long as the contract is individually rational). Thus, (PC) as well as (NC) together play a crucial role in determining the optimal contract. Note that, because the loan rate is exogenously fixed, the only instrument the bank has to provide incentives is the share of repayment. Consider first high values of z which means low stake of collusion, $pr - z$. As a result, the threat of collusion is low, and the no-collusion constraint does not bind. The bank can thus keep on lowering s without affecting incentives until the entire rent of the loan officer has been extracted, i.e., the participation constraint binds. Therefore, in the absence of collusion incentive problem, the non-delegated level of monitoring, $m_1^*(r)$ can be implemented via a low incentive pay.

Next, consider the situation when borrower quality is low, and hence, the stake of collusion is high. in this case, the no-collusion constraint bind. The only way to deter collusion is to raise the incentive pay through s. The bank can continue extracting the entire rent of the loan officer by specifying larger m (as the binding participation constraint implies a positive association between m and s). Both m and s increases as z decreases, and the monitoring effort is higher than that under non-delegation implying an over-monitoring. Therefore, as in Mookherjee and Png (1995), we find that the threat of collusion turns out to be an inefficient instrument to enhance monitoring. When z is lowered further, due to very high stake of collusion, the agency problem is so severe that the loan officer must be given incentives in the form of additional rent, i.e., (PC) does not bind anymore (the loan officer earns *efficiency wage*). The monitoring level is optimally set at its highest possible level, 1.

The *collusion-proofness principle* (e.g. Tirole, 1986; Laffont and Rochet, 1997; Laffont and Martimort, 2000) asserts that in a principal-supervisor-agent hierarchy under the possibility of vertical collusion [between the supervisor and the agent], there is no loss of generality in restricting attention to the collusion-free contracts. Under single-bank lending, the aforementioned principle clearly holds because if the contract does not aim at deterring collusion, the loan officer surely colludes in equilibrium, and the bank never breaks even because it does not receive any repayment. In other words, it is never optimal for the bank to tolerate graft.

4.2 Two-bank lending

When there are two banks in the market, the firm receives \$1 to invest from each bank, and hence, the aggregate loan amount is \$2. Given the individual monitoring efforts $m_i, m_j \in [0, 1]$, the aggregate monitoring intensity is given by (1). We continue assuming that the loan rate is exogenously given, and is the same under both lending structures, which is equal to r.

4.2.1 Direct bank monitoring without delegation

As in the previous subsection, we first analyze the situation where each bank can monitor the borrower directly. Bank i solves

$$\max_{m_i \in [0,1]} B_i(m_i, m_j) \equiv \pi(m_i, m_j)pr - \frac{1}{2}cm_i^2 - 1. \tag{$\mathscr{M}_i^*$}$$

The first-order conditions of the maximization problems of banks i and j yield the following best reply functions:

$$cm_i = pr(1 - m_j), \tag{BR$_i$}$$

$$cm_j = pr(1 - m_i). \tag{BR$_j$}$$

These reaction functions have a unique and symmetric solution $m_2^*(r)$. The aggregate monitoring intensity is given by $\pi^*(r) = \pi(m_2^*(r), m_2^*(r))$.

Proposition 3 *When two banks lend to the borrower, and they can monitor her directly, the optimal monitoring effort and aggregate monitoring intensity are respectively given by:*

$$m_2^*(r) = \frac{pr}{c + pr},$$

$$\pi^*(r) = 1 - [1 - m_2^*(r)]^2 = 1 - \left(\frac{c}{c + pr}\right)^2.$$

As in Proposition 1, the individual monitoring effort, and thus the aggregate monitoring intensity, are independent of the level of borrower quality because there is no collusion under non-delegation.

4.2.2 Delegated monitoring

The bribery subgame We analyze the collusion possibility between the firm and loan officer i. Recall that for loan officer i to be able to collude with the firm he must succeed in detecting borrower (mis)behavior. There are two possible cases following the monitoring success of loan officer i. First, consider the case when loan officer j has not succeeded. Let $F(\sigma)$ and $M_i(\sigma)$ respectively be the payoffs of the firm and monitor i where $\sigma \in \{0, 1\}$ with 0 representing *collusion* and 1 representing *no-collusion*. The payoffs are given by:

$$F(0) = 2v - b_i, \quad \text{and} \quad M_i(0) = b_i - C(m_i);$$
$$F(1) = 2p(y-r), \quad \text{and} \quad M_i(1) = prs_i - C(m_i).$$

Thus, collusion is not feasible if and only if

$$F(1) + M_i(1) \geq F(0) + M_i(0) \quad \Longleftrightarrow \quad prs_i \geq 2(pr - z). \tag{NC$_i$}$$

Similarly, the no-collusion constraint for loan officer j (when he succeeds in monitoring, but loan officer i does not) is given by:

$$prs_j \geq 2(pr - z). \tag{NC$_j$}$$

Now consider the second scenario where both i and j have successfully detected borrower behavior. In this case, the no-collusion constraint the bank must take into account is slightly different from (NC$_i$) and (NC$_j$). Let 0 represent collusion between among the borrower and two loan officers, and 1 represent no-collusion, i.e., at least one loan officer decided not to collude with the borrower. The payoffs in this case are given by:

$$F(0) = 2v - (b_i + b_j), \quad M_i(0) = b_i - C(m_i) \quad \text{and} \quad M_j(0) = b_j - C(m_j);$$
$$F(1) = 2p(y-r), \quad M_i(1) = prs_i - C(m_i) \quad \text{and} \quad M_j(1) = prs_j - C(m_j).$$

Thus, collusion is not feasible if and only if

$$F(1) + M_i(1) + M_j(1) \geq F(0) + M_i(0) + M_j(0) \quad \Longleftrightarrow \quad pr(s_i + s_i) \geq 2(pr - z). \tag{NC$_{ij}$}$$

Clearly, (NC$_i$) and (NC$_j$) together imply (NC$_{ij}$). The above no-collusion constraints assert that the (expected) incentive pay of each loan officer must exceed the stake of collusion under two-bank lending, $2(pr - z)$.

Optimal loan officer incentives and monitoring At date $t = 2$, each bank offers a contract to its loan officer, taking the contract offer of the rival bank as given. Apart from satisfying the no-collusion constraint, the contract (m_i, s_i) must satisfy the *participation constraint* of loan officer i

$$\pi(m_i, m_j)prs_i - \frac{1}{2}cm_i^2 \geq 0,, \tag{PC$_i$}$$

and the *feasibility constraint*

$$(m_i, s_i) \in [0, 1] \times [0, 1]. \tag{F$_i$}$$

Similarly, in addition to the no-collusion constraint, (m_j, s_j) must satisfy the participation constraint of loan officer j

$$\pi(m_i, m_j)prs_j - \frac{1}{2}cm_j^2 \geq 0, \tag{PC$_j$}$$

and the feasibility constraint

$$(m_j, s_j) \in [0, 1] \times [0, 1]. \tag{F$_j$}$$

We denote by $\mathscr{F}_i$ the set of *feasible contracts* for bank i, i.e., the set of contracts that satisfies (PC$_i$), (NC$_i$) and (F$_i$). We define $\mathscr{F}_j$ for bank j in an analogous manner. It is worth noting that the feasible set $\mathscr{F}_i$ of bank i depends on the contract offered by bank j and vice versa. In other words, contracts for the loan officers are subject to externality. Bank i solves the following maximization problem:

$$\max_{(m_i, s_i) \in \mathscr{F}_i} B_i(m_i, m_j, s_i) \equiv \pi(m_i, m_j)pr(1 - s_i) - 1. \tag{$\mathscr{M}_i$}$$

We analyze an equilibrium in which the two no-collusion constraints either bind simultaneously or none of them binds.[15] Let $m_i(m_j)$ and $m_j(m_i)$ denote the best reply functions derived by solving the maximization problems of the banks. Lemma 1 derives properties of the best reply functions $m_i(m_j)$ and $m_j(m_i)$ that will be critical to whether individual monitoring level under single-bank lending is lower relative to that under two-bank lending or not (see the discussion following Proposition 5).

Lemma 1 *Under two-bank lending, the participation constraint of each loan officer binds. Moreover,*

(a) *when neither of the two no-collusion constraints binds, the individual monitoring efforts are strategic substitutes, i.e., the best reply functions $m_i(m_j)$ and $m_j(m_i)$ are negatively sloped;*

(b) *When both the no-collusion constraints bind, the individual monitoring efforts are strategic complements, i.e., the best reply functions $m_i(m_j)$ and $m_j(m_i)$ are positively sloped.*

Unlike the single-bank lending mode, the participation constraints of each loan officer binds at the optimum. When neither of the two no-collusion constraints binds, the agency problem does not affect banks' choice of monitoring levels relative to direct bank monitoring. In this case, the best reply functions are given by (BR$_i$) and (BR$_j$). Thus, the fact that $\pi_{ij}(m_i, m_j) < 0$ ensures that individual monitoring efforts are strategic substitutes, i.e., the best reply functions are negatively sloped. Next, suppose that both the no-collusion constraints bind. We can use the binding participation constraints to substitute for s_i (resp. s_j) into (NC$_i$) (resp. (NC$_j$)) to obtain the best replies $m_i(m_j)$ and $m_j(m_i)$:

$$cm_i^2 = 4(pr - z)\pi(m_i, m_j), \tag{BR$'_i$}$$

$$cm_j^2 = 4(pr - z)\pi(m_i, m_j). \tag{BR$'_j$}$$

The above best reply functions are upward-sloping. This strategic complementarity of monitoring efforts emerges from an interesting interaction between the two participation constraints. Intuitively, the trade-off each bank faces is either to provide incentives to deter collusion or to extract rent from its loan officer.

[15] There is an asymmetric equilibrium where only one of the two no-collusion constraints (NC$_i$) and (NC$_j$) binds. See Proof of Proposition 4 in Appendix A for details.

Consider the optimal contract for loan officer i. With an increase in m_j, the participation constraint of loan officer i gets relaxed because $\pi_j(m_i, m_j) > 0$. However, in order to extract this additional rent, bank i cannot lower the loan officer's share because s_i is fixed given the binding no-collusion constraint. Therefore, the only option bank i is left with is to increase m_i in order to make the participation constraint bind. Moreover, given that bank i is constrained by (NC_i), it has already been operating at a sub-optimal level of m_i. Thus, an increase in m_i in this case increases the bank's profits.[16]

The optimal monitoring and share of repayment for each loan officer are characterized in Proposition 4 below. Under this lending mode, as in the case of single-bank lending, collusive threats exacerbate incentives for over-monitoring, and hence, banks are required to offer stronger incentives to deter collusion.

Proposition 4 *There are threshold values of firm quality, z_2^0 and $\bar{z}_2$ with $0 < z_2^0 < \bar{z}_2 < pr$ such that*

(a) *The symmetric equilibrium monitoring effort is given by:*

$$m_2(z, r) = \begin{cases} \frac{8(pr-z)}{c+4(pr-z)} & \text{for } z \in (z_2^0, \bar{z}_2], \\ \frac{pr}{c+pr} & \text{for } z \in (\bar{z}_2, pr]. \end{cases}$$

The aggregate monitoring intensity is given by:

$$\pi(z, r) = \begin{cases} 1 - \left(\frac{c-4(pr-z)}{c+4(pr-z)}\right)^2 & \text{for } z \in [z_2^0, \bar{z}_2], \\ 1 - \left(\frac{c}{c+pr}\right)^2 & \text{for } z \in (\bar{z}_2, pr]. \end{cases}$$

Thus, whenever borrower quality is low, i.e., $z \leq \bar{z}_2$, the outcome involves over-monitoring relative to the non-delegated level of monitoring, i.e., $\pi^*(r)$. Whereas for all $z > \bar{z}_2$, so that firm quality is high, the monitoring level is at the non-delegated level of monitoring.

(b) *The optimal share of repayment for the loan officer, on the other hand, is given by:*

$$s_2(z, r) = \begin{cases} 2\left(1 - \frac{z}{pr}\right) & \text{for } z \in [z_2^0, \bar{z}_2], \\ \frac{c}{2(2c+pr)} & \text{for } z \in (\bar{z}_2, pr]. \end{cases}$$

Proposition 4 is depicted in Figure 3.

Note that when the collusion incentive problem is important (binding no-collusion constraints), there emerges a positive externality between the banks. Because binding (NC_i) and (NC_j) fix both s_i and s_j at a minimum, any strategic interaction between the lenders that may exist must be through m_i and m_j. Note that for any bank, say bank i, the 'relative cost' $C(m_i)/\pi(m_i, m_j)$ of monitoring m_i is decreasing

[16] In the absence of the incentive problems with respect to collusion, wasteful duplication of monitoring effort implied by the technology $\pi(m_i, m_j) = m_i + m_j - m_i m_j$ results in the strategic substitutability of monitoring efforts m_i and m_j. If we have assumed instead that $\pi(m_i, m_j) = m_i + m_j + \delta m_i m_j$ with $\delta > 0$, i.e. there is no wasteful monitoring duplication, then monitoring efforts would have been strategic complements because the marginal benefit of monitoring by one loan officer is increasing in that by the other. By contrast, when the no-collusion constraints of both loan officers bind, even under no wasteful monitoring duplication, individual monitoring efforts remain to be strategic complements because $\delta > 0$ reinforces the effect of the trade-off between rent extraction and incentive provision.

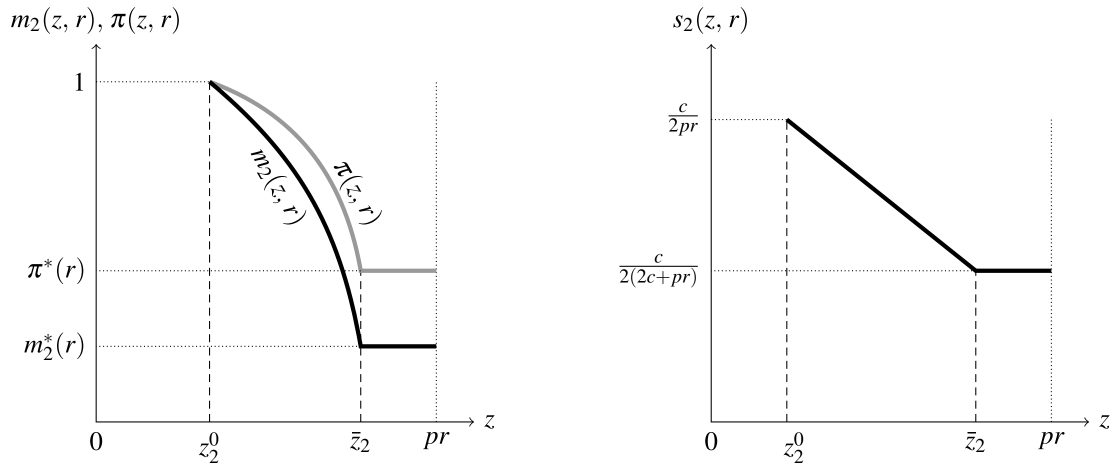

Figure 3: *The equilibrium individual monitoring effort, aggregate monitoring intensity and share of the loan officer as functions of firm quality. For $z \leq \bar{z}_2$, the no-collusion constraints bind, and there is over-monitoring relative to the non-delegated level.*

in monitoring by the other bank, m_j. Thus, bank i can reduce this relative cost of increasing m_i as it knows there will be a concomitant increase in m_j. In equilibrium, there cannot remain such a scope for deviation, which leads to over monitoring. The aforementioned externality is more severe for lower z—if z is very small (very high stake of collusion), s_j is very high due to binding (NC$_j$). Therefore, the incentive to increase m_i is also high because m_j can be increased a lot.[17]

The equilibrium association between monitoring and borrower quality described in Propositions 2 and 4 are similar because they share a common general intuition.[18] Under both lending structures, low values of z is equivalent to high stake of collusion because $pr - z$ is high, which exacerbates the collusion incentive problem. Moreover, in both cases the share of repayment is the only instrument at the banks' disposal address collusion concerns because they are not able to reduce the size of the stake of collusion given that the loan rate is exogenously given (an endogenous loan rate, which is analyzed in Appendix B, is an additional instrument for the banks which can be utilized to reduce the stake). Therefore, we obtain a positive association between the equilibrium share and the stake of collusion. Similar intuition is present in earlier works on vertical collusion (e.g. Tirole, 1986; Laffont and Rochet, 1997). Notably, although the monitoring level is contractible, it is not an instrument to provide incentives, rather each bank control this variable in order to extract rent from the loan officers.

It is also worth noting that we restrict attention to $z > z_2^0$ because in any symmetric equilibrium, we

[17]Consider $m_i(m_j)$ defined by (BR$'_i$). For a given $\bar{m}_j$, differentiating (BR$'_i$) we obtain

$$\frac{dm_i}{dz} = -\frac{2\pi(m_i, \bar{m}_j)}{cm_i - 2(pr-z)(1-\bar{m}_j)}.$$

Because for any $\bar{m}_j \in [0, 1]$ the denominator of the above expression is strictly positive, we have $dm_i/dz < 0$ implying that, for a given $\bar{m}_j$, m_i increases as z decreases, i.e. the best reply function $m_i(m_j)$ shifts outward for any $m_j \in [0, 1]$ as borrower quality lowers. Same argument goes through for the best reply $m_j(m_i)$ defined by (BR$'_j$).

[18]The concavity of $m_2(z,r)$ and $\pi(z,r)$ with respect to z, as opposed to the linearity of $m_1(z,r)$, arises because the individual monitoring efforts are substitutes in determining the aggregate monitoring intensity. In the case when $\pi(m_i, m_j) = m_i + m_j + \delta m_i m_j$, the functions $m_2(z,r)$ and $\pi(z,r)$ would have been strictly decreasing and convex on $(z_2^0, \bar{z}_2]$.

must have $m_2(z, r) < 1$ (which is equivalent to $z > z_2^0$). If one of the two banks, say bank j chooses $m_j = 1$, then the aggregate monitoring intensity is given by $\pi(m_i, 1) = 1$ for any $m_i \geq 0$. Thus, bank i has no incentive to induce any monitoring at all, let alone $m_i = 1$. Moreover, it is easy to show that at $z = z_2^0$, each bank earns strictly positive expected profit.[19] Finally, note that our focus is on analyzing the strategic interaction between the banks when they enforce collusion-free contracts. Thus, in the baseline model, we abstract from the issue of whether the collusion-proofness principle holds or not. For completeness however, we take up this issue in Section 6.1, where we show that there exist asymmetric equilibria for values of $z \leq z_2^0$ in which only one bank implements the collusion-free contract.

5 Comparison of the two lending modes

5.1 The free-riding and race-to-collusion effects of two-bank lending

We now turn to the central question of this paper—a comparison of the equilibrium monitoring, as well as the incentives offered to the loan officer(s) under the single- and two-bank lending structures.

Proposition 5 *Two bank lending implies larger collusive threats, i.e., the no-collusion constraint of each loan officer binds over a larger range of firm quality ($\bar{z}_2 > \bar{z}_1$). Hence, there is inefficient over-monitoring for a larger range of firm quality z. Moreover, for a given loan rate r, there are unique threshold values of firm quality, $z_m, z_s \in (z_2^0, \bar{z}_2)$ such that:*

(a) *Equilibrium monitoring effort under two-bank lending is higher relative to single-bank lending, i.e., $m_2(z, r) \geq m_1(z, r)$ if and only if $z \leq z_m$;*

(b) *Incentives for the loan officers are stronger under two-bank lending, i.e., $s_2(z, r) \geq s_1(z, r)$ if and only if $z \leq z_s$.*

The results in Propositions 5 are depicted in Figure 4.

Proposition 5 conveys the central message of our paper—when firm quality is high, two-bank lending elicits lower monitoring effort. Moreover, the banks provides *weaker* incentives under two-bank lending to *each* loan officer. By contrast, when firm quality is low, two-bank lending implies more intense monitoring and stronger incentives to each loan officer. These results are driven by the two countervailing channels unearthed in Lemma 1 earlier. The first one is the strategic substitutability of individual monitoring efforts which gives rise to *free-riding*, and the second one is the strategic complementarity that leads to the *race-to-collusion effect*. Clearly, the two effects point in opposite directions, and hence, the effect of two-bank lending on individual monitoring effort and loan officer incentives is ambiguous.

Which one of these two effects dominates depends on the level of borrower quality, i.e., z. When firm quality is high, there is no feasible bribe that makes collusion profitable, and hence, loan officer incentives are perfectly aligned with those of the banks, and the race-to-collusion effect plays no role. Only the free-riding effect is at play here, leading to lower monitoring effort for each monitor under two-bank lending. One would have obtained the same result under a common loan rate r had the banks been

[19] Let z_2^{min} denote the level of borrower quality at which both banks earn zero expected payoffs. It is easy to show that $z_2^{min} < z_2^0$.

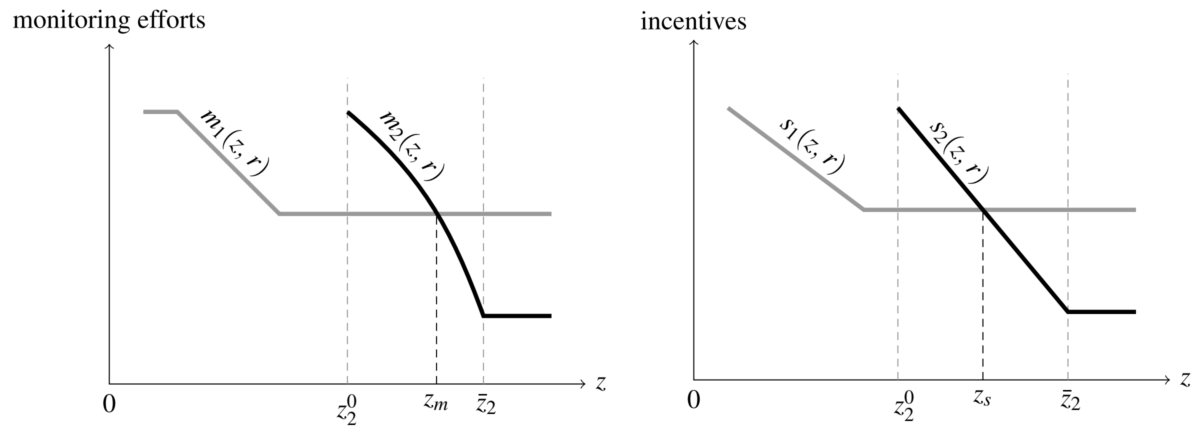

Figure 4: *The equilibrium individual monitoring efforts and incentives under the single- and two-bank lending structures. Under two-bank lending, each bank elicits higher monitoring effort and provides stronger incentives if and only if firm quality is low.*

able to monitor the borrower directly, which has been shown by Carletti (2004) and Khalil et al. (2007). On the other hand, when borrower quality is low, the no-collusion constraints bind. As we have argued earlier, this leads to over-monitoring under single-bank lending as well as two-bank lending. However, under two-bank lending, an additional effect comes into play—namely, the race-to-collusion effect, that amplifies the effect on m_i arising from a larger m_j. Thus for low levels of z, *individual* monitoring effort under two-bank lending exceeds that under single-bank lending. Consequently, the share of repayment must also be higher in an effort to deter collusion. Moreover, as we have noted earlier that a lower z shifts both the best reply functions defined by (BR_i') and (BR_j') outward, the race-to-collusion effect is stronger for low values of firm quality causing even a higher equilibrium monitoring effort.

We next turn to comparing the monitoring intensity under single-bank lending with the *aggregate* monitoring intensity under two-bank lending.

Proposition 6 *Assume $c \leq 8.47$. Then, there is a unique threshold value of firm quality, $z_\pi \in (z_2^0, \bar{z}_2)$ such that the aggregate monitoring intensity under two-bank lending is higher than that under single-bank lending, i.e. $\pi(z,r) > m_1(z,r)$ if and only if $z < z_\pi$.*

While comparing the aggregate monitoring intensities under single- and two-bank lending, a comparison of the individual monitoring levels alone provides a partial picture. There is an additional effect in play here in that monitoring is a public good, so that the presence of an additional monitor ensures that $\pi(m_i, m_j)$ is higher than both m_i and m_j. When borrower quality is low, the race-to-collusion effect ensures that individual monitoring levels are higher under two-bank lending, and hence, given the public good effect, so is the aggregate monitoring intensity. By contrast, for high values of z, the presence of free-riding makes the public good effect more important. On the one hand, low individual monitoring effort under two-bank lending translates into low aggregate monitoring. On the other, the public good effect may imply higher aggregate monitoring even if the individual monitoring effects are low. It turns out that when the marginal cost of monitoring is not too high, i.e., $c \leq 8.47$, the first effect dominates the second one, and hence for high values of z, $\pi \leq m_1$.

5.2 Disentangling the two countervailing effects

We have demonstrated that a move from one-bank to two-bank lending has ambiguous consequences for monitoring as well as loan officer incentives. We have argued that this ambiguity emerges because of the interaction between two countervailing channels—namely, the free-riding and the race-to-collusion effects of two-bank lending. In a bid to disentangle the two effects, in what follows we carry out a purely conceptual exercise. In particular, we shut down one channel at a time, and examine how the closing down of a particular channel affects monitoring and incentives under two-bank lending.

5.2.1 Strategic versus merged banks: the free-riding effect

We first compare the baseline two-bank lending framework analyzed in Section 4.2 with one where the free-riding channel is shutdown. To that end we examine a scenario where the two banks merge, with the merged bank maximizing joint profits, but still employing two independent loan officers. The fact that the banks coordinate on monitoring ensures that there is no free-riding. However, the race-to-collusion effect is still present because the merged entity employs two loan officers. In order to ensure that there is no confounding effect from loan size, we assume that the merged bank continues lending \$2 to the firm. Formally, the merged bank solves the following maximization problem:

$$\max_{\{(m_i,s_i)\in\mathscr{F}_i,(m_j,s_j)\in\mathscr{F}_j\}} B_i(m_i,m_j,s_i)+B_j(m_i,m_j,s_j)=\pi(m_i,m_j)pr(2-s_i-s_j)-2, \qquad (\mathscr{M}_{ij})$$

Note that the merged entity faces the same participation and no-collusion constraints as in Section 4.2. We compare the optimal solution of the above maximization problem with that described in Proposition 4. The following proposition analyzes the role of the free-riding effect in determining the optimal monitoring and incentives.

Proposition 7 *The individual monitoring effort is higher and loan officer incentives are stronger when the two banks act as a merged entity (so that there is no free-riding), rather than when they maximize their own profits independently (when free-riding is present).*

These results are graphically illustrated in Figure 5 where we depict the symmetric equilibrium monitoring efforts (in the left panel) and equilibrium share of repayment (in the right panel). Intuitively, irrespective of whether the banks act as a merged entity or not, they face the same no-collusion constraints (NC_i) and (NC_j). Thus, when the no-collusion constraints bind, the optimal contracts under strategic and merged banks coincide, leading to identical aggregate monitoring as well. Whereas if the no-collusion constraints do not bind, merged banks are able to offer stronger incentives to their loan officers since the free-riding effect is absent, and hence, aggregate as well as per bank monitoring are higher. Consequently, the merged entity elicits higher monitoring, i.e., the equilibrium monitoring effort function shifts up from $m_2(z,r)$ to $m_{mb}(z,r)$.

Proposition 7 finds resonance in the literature on strategic managerial incentives in oligopolistic markets (e.g. Martin, 1993; Golan, Parlour, and Rajan, 2015; Dam and Robinson-Cortés, 2020), where strategic interaction among firms results in weaker incentives for managers, generating efficiency losses. In these models firms, who are inefficient to begin with, hire a manager apiece whose principal task is to exert R&D effort so as to improve productive efficiency. The cost realization of each firm is independent of those of its rivals, and hence, there is no contract externality. When there are more firms in the market,

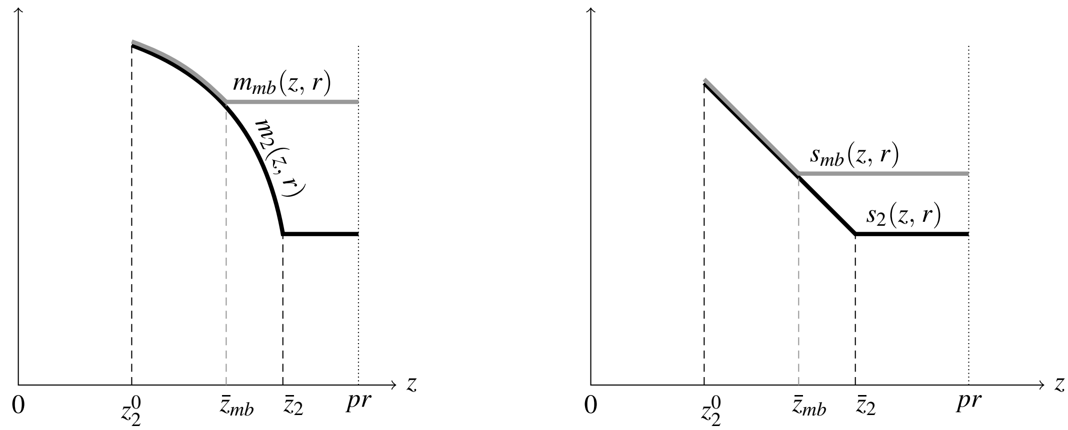

Figure 5: *The equilibrium monitoring effort and share under strategic and merged banks. Subscript 'mb' denotes the variables under merged banks.*

output, and hence, profits per capita declines. This results in a relatively lower marginal value of cost reduction, and thus lower marginal benefit of managerial effort. Consequently, as the level of strategic interaction intensifies with a growing number of firms, each firm finds it optimal to weaken managerial incentives that, in turn, elicit lower managerial efforts in equilibrium.

5.2.2 Competing versus cooperating loan officers: the race-to-collusion effect

Next, we analyze the role played by the race-to-collusion effect, comparing the baseline framework in Section 4.2 with one where the race-to-collusion effect is absent. To be precise we examine a scenario where loan officers behave cooperatively in the sense that if at least one of them succeeds in monitoring, then they share this information among themselves. Further, they jointly collect the bribe in case they decide to collude with the firm.[20]

Suppose at least one of the loan officers has succeeded, and the loan officers together collude with the borrower. The no-collusion constraint in this case is given by (NC_{ij}), i.e.,

$$pr(s_i + s_j) \geq 2(pr - z),$$

i.e., the aggregate surplus from truth telling exceeds that under collusion. Each bank i would thus choose (m_i, s_i) to maximize the objective function in ($\mathscr{M}_i$), subject to (PC_i), (F_i), and the (joint) no-collusion constraint (NC_{ij}). Comparing the optimal contract in this situation with that described in Proposition 4, we obtain the impact of the race-to-collusion effect on the optimal aggregate monitoring and loan officer incentives.

Proposition 8 *Both individual monitoring effort and aggregate monitoring intensity are higher and loan officer incentives are stronger when the loan officers compete with each other, so that there is a race to collude, rather than behave cooperatively, when there is no such race.*

[20] An alternative way to shut down the race-to-collusion effect would be to assume that there is a single loan officer who is jointly employed by the two firms. However, given that the monitoring cost function is super-additive, i.e. $C(m_i) + C(m_j) > C(m_i + m_j)$, following that approach would introduce an additional confounding effect, and hence we do not adopt it.

Proposition 8 is illustrated in Figure 6. Intuitively, when the loan officers cooperate instead of competing with each other for bribe, lower shares are required to deter collusion. When the monitors compete, the minimum share in each bank-loan officer relationship that is required to deter collusion is given by $2(1-z/pr)$. By contrast, when they cooperate, from (NC_{ij}) it follows that the minimum *aggregate* collusion-deterring share $s_i + s_j$ is $2(1-z/pr)$. Thus, the race-to-collusion effect disappears, and the inefficiency due to incentives for over-monitoring reduces because the equilibrium monitoring effort function shifts down from $m_2(z, r)$ to $m_{cp}(z, r)$. Consequently, the required incentives are weaker, i.e., the optimal share reduces from $s_2(z, r)$ to $s_{cp}(z, r)$.

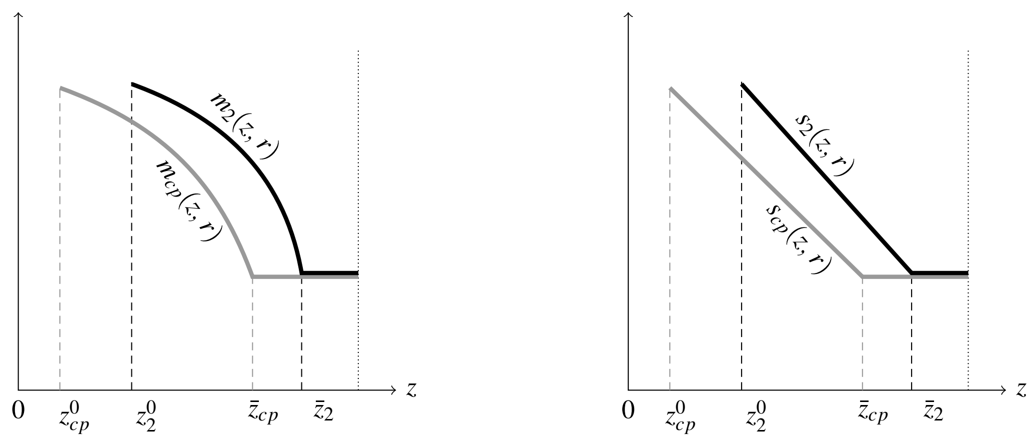

Figure 6: *The equilibrium monitoring effort and share under competing and cooperating loan officers. Subscript 'cp' denotes the variables under cooperating loan officers.*

Proposition 8 is related to the literature on optimal incentive contracts in a multi-agent situation. In analyzing an organization with a single principal and two agents, Ramakrishnan and Thakor (1991) assert that the principal may prefer that the agents cooperate, rather than compete. Similarly, Macho-Stadler and Pérez-Castrillo (1993) demonstrate that an increase in the capacity of the agents to cooperate leads to more efficient outcomes. Proposition 8 points in a similar direction. If banks could have enforced cooperation between the loan officers, the loss in efficiency due to incentives for over-monitoring would have been minimized. However, in our framework, banks cannot ensure cooperation between the loan officers because collusion is not publicly verifiable.

Remark 1 The intuitions developed in Propositions 7 and 8 are quite robust. Proposition 7, for example, goes through if, under both strategic and merged banks, loan officers behave cooperatively, rather than competitively. The results can be summarized in a figure that is very similar to Figure 5. Similarly, Proposition 8 goes through if, under both competitive and cooperative loan officers, we assume that the banks are merged, rather than strategic. The results can be summarized in a figure similar to Figure 6.[21]

5.3 Welfare analysis

Comparison of welfare under the single- and two-bank lending structures is difficult because of the differences in the amounts lent and private benefits, and because the monitoring cost functions are convex. We instead compare welfare under two-bank lending with (a) lending under merged banks, and (b) lending

[21]The proofs are available from the authors upon request.

under cooperating loan officers. The first exercise aims at understanding the importance of the free-riding effect with respect to welfare, and the second one analyzes the role played by the race-to-collusion effect.

Denote by $W(m_i, m_j)$ be the welfare under two-bank lending at the monitoring levels (m_i, m_j), which is given by:

$$W(m_i, m_j) = 2[\pi(m_i, m_j)py + (1-\pi(m_i, m_j))v - 1] - \frac{1}{2}\left(cm_i^2 + cm_j^2\right).$$

The above expression is maximized at $m_i = m_j = \frac{2z}{c+2z}$. Therefore, none of $m_2(z, r)$, $m_{mb}(z, r)$ and $m_{cp}(z, r)$ is at the efficient level, i.e., the one which maximizes welfare. Further, let $W(m) \equiv W(m, m)$ be the welfare under symmetric monitoring efforts $m_i = m_j = m$. The following proposition states the main results with respect to welfare.

Proposition 9 *Consider the symmetric equilibrium monitoring efforts under two-bank lending.*

(a) *There is a unique $\tilde{z} \in (\bar{z}_{mb}, \bar{z}_2)$ such that $W(m_2(z, r)) \geq W(m_{mb}(z, r))$ for all $z \in [z_2^0, \tilde{z}]$, and $W(m_2(z, r)) < W(m_{mb}(z, r))$ for all $z \in (\tilde{z}, pr]$;*

(b) *On the other hand, there is a unique $\hat{z} \in (\bar{z}_{cp}, \bar{z}_2)$ such that $W(m_2(z, r)) < W(m_{cp}(z, r))$ for all $z \in [z_2^0, \hat{z})$, and $W(m_2(z, r)) \geq W(m_{cp}(z, r))$ for all $z \in [\hat{z}, pr]$.*

Consider first the case of merged versus strategic banks. The difference in monitoring levels under these two lending modes emerges for $z > \bar{z}_{mb}$ as shown in Figure 5. Thus, merged and strategic banks entail the same welfare for all $z \leq \bar{z}_{mb}$. For low values of z close to $\bar{z}_{mb}$, strategic banks face collusive threats, and hence, the efficiency enhancing effect of merged banks is not yet important. By contrast, in the absence of the collusion incentive problems for high values of z, the free-riding problem comes to be significant. Thus, the role of merged banks in enhancing efficiency becomes consequential. Therefore, for low values of z strategic banks entail higher welfare, whereas welfare is higher under merged banks for high values of firm quality. The intuition behind part (b) is very similar. For low values of z, cooperating loan officers add more to efficiency because collusive threats are severe. By contrast, as the collusion incentive problem dies out for high values of z, there is hardly any difference if the loan officers compete or cooperate. Therefore, for high borrower quality, competition between the monitors implies higher welfare.

6 Extensions

In the preceding analysis we have kept the framework parsimonious in an effort to focus on the essential trade-offs between the race-to-collusion and the free-riding effect arising from strategic interaction. We now extend the analysis in several directions.[22]

6.1 Is it always optimal to deter collusion?

We have discussed earlier that the collusion-proofness principle always holds under single-bank lending. By contrast, under two-bank lending, if one bank enforces a collusion-free contract, then the other bank

[22] In Appendix B we analyze two more extension—(a) when the monitoring efforts are not contractible, and (b) multiple-bank lending when there are $n \geq 2$ banks, each of them lending \$1 to the firm.

has incentives to free-ride on the (non-colluding) bank, thereby saving on incentive costs itself. Thus, it is not clear that the collusion-proofness principle necessarily holds. We demonstrate that if firm quality z is sufficiently large, then for both banks it is optimal to offer collusion-free contracts to their loan officers. However, if z is small, then optimally while one bank ensures that there is no collusion by its own loan officer, the other bank does not, free-riding on the monitoring being done by its rival bank. For intermediate values of z multiple equilibria exist, so that while there is one equilibrium where both banks ensure non-collusion, there is another equilibrium where one bank allows collusion, while the other bank does not.

For ease of exposition, we shall classify all possible equilibria under three headings:

(1) An equilibrium strategy profile is said to be (NC, NC) if both banks implement collusion-free contracts so that neither loan officer colludes;

(2) An equilibrium strategy profile is said to be (C, NC) if one bank does not implement a collusion-free contract, whereas the other bank does;

(3) An equilibrium strategy profile is said to be (C, C) if neither of them implements a collusion-free contract.

Clearly, one cannot have an equilibrium of type (C, C) because neither bank would break even, making a loss of -1. Further, recall that equilibria of type (NC, NC) have been analyzed in Section 4.2. So, in what follows we analyze equilibria of type (C, NC). Without any loss of generality, we assume that bank i does not implement a collusion-free contract, but bank j does in an equilibrium (C, NC). Note that for high values of firm quality $z > \bar{z}_2$ the no-collusion constraint of neither bank binds [cf. Proposition 4], and hence, the equilibrium (C, NC) can only occur for some values of $z \leq \bar{z}_2$. We first state the main result of this subsection.

Proposition 10 *There are threshold values z_j^{min}, $\bar{z}_j$, z_2^C and $\hat{z}_2$ of firm quality, with $z_j^{min} < \bar{z}_j < z_2^C < \hat{z}_2 \leq \bar{z}_2$, such that*

(a) *for any $z \geq z_2^C$, an equilibrium of type (NC, NC) exists, whereas*

(b) *for any $z \leq \hat{z}_2$, an equilibrium of type (C, NC) exists, where bank i sets $m_i = 0$, and only loan officer j monitors which is given by:*

$$m_j(z, r) = \begin{cases} 1 & \text{for } z \in [z_j^{min}, z_2^0], \\ \frac{4(pr-z)}{c} & \text{for } z \in (z_2^0, \bar{z}_j], \\ \frac{pr}{c} & \text{for } z \in (\bar{z}_j, \hat{z}_2]; \end{cases}$$

(c) *The equilibrium individual monitoring effort under two-bank lending is higher than that under single-bank lending if $z \leq \hat{z}_2$. By contrast, the equilibrium individual monitoring effort under two-bank lending is lower than that under single-bank lending if $z \geq z_2^C$.*

Note that for any $z \geq \hat{z}_2$, the unique equilibrium involves (NC, NC), so that the collusion-proofness principle holds; whereas for each $z \in [z_2^C, \hat{z}_2]$, the collusion-proofness principle holds weakly in that there exist one equilibrium of type (NC, NC), and another of type (C, NC). For $z < z_2^C$, all equilibria are

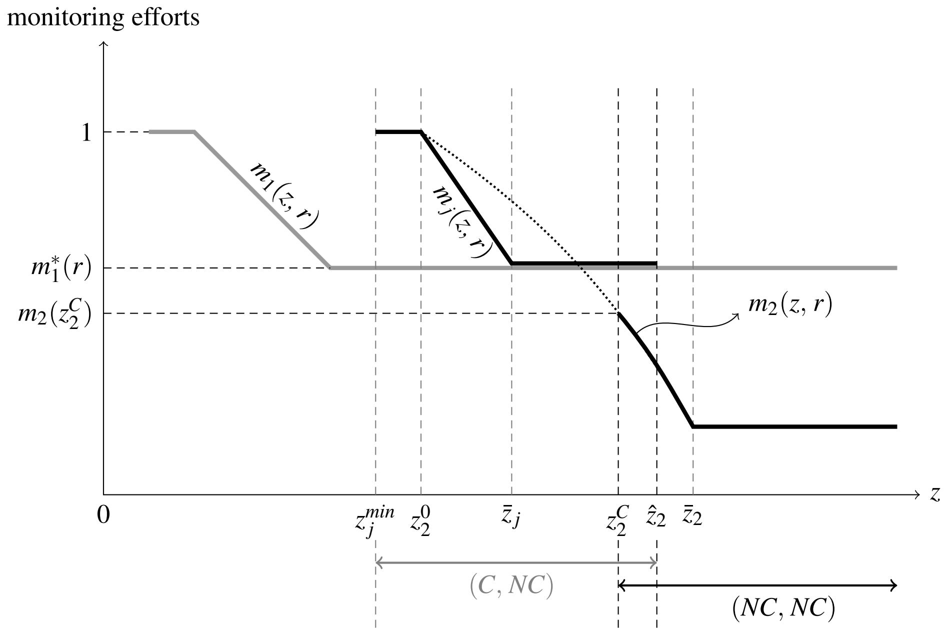

Figure 7: *The equilibrium monitoring efforts under two-bank lending has two segments—$m_j(z, r)$ corresponds to (C, NC) equilibrium, and $m_2(z, r)$ corresponds to (NC, NC) equilibrium. While the segment $m_j(z, r)$ lies above $m_1(z, r)$, the monitoring under single-bank lending, the segment $m_2(z, r)$ lies below $m_1(z, r)$. Also, for $z_2^C \leq z \leq \hat{z}_2$ there are multiple equilibria.*

of type (C, NC), so that the collusion-proofness principle does not hold. The above results are depicted in Figure 7.

We postpone the detailed technical analysis of the aforementioned equilibria under two-bank lending until Appendix A [cf. Proof of Proposition 10]. In a (C, NC) equilibrium, only loan officer j monitors, and hence, the equilibrium monitoring function looks similar to that under single-bank lending. However, $m_j(z, r)$ is higher that $m_1(z, r)$ because with \$2 being lent to the firm, the aggregate surplus from collusion is now doubled, so that the no-collusion constraint is more difficult to satisfy. From Proposition 10, under two-bank lending, a collusion-free equilibrium does not exist for any $z \leq z_2^C$. However, a (C, NC) equilibrium exists for all $z \in [z_j^{min}, z_2^0]$, with bank j earning non-negative profits if $z \geq z_j^{min}$. Moreover, bank j optimally sets $m_j = 1$ when z is close to z_j^{min}. Notice also that $m_2(z, r)$ is always higher than $m_j(z, r)$ for low values of z (as shown by the dotted curve in Figure 7). This means that for low values of z, the (C, NC) equilibrium reduces over-monitoring, thereby generating an efficiency gain over the (NC, NC) equilibrium.

Intuitively, in an equilibrium (C, NC) where bank i allows for collusion, it optimally sets $s_i = 0$. Moreover, given that loan officer i colludes, no bank gains from any monitoring this loan officer might do, so that in this case we have $\pi(m_i, m_j) = m_j$. This situation is thus equivalent to the scenario in which bank i does not employ any loan officer, thereby saving on the cost of incentive provision, and free rides on the monitoring effort exerted by the loan officer of the other bank.[23] From Figure 7, it is also clear

[23]In the Appendix we show that in any (C, NC) equilibrium, bank i can employ a loan officer by setting $s_i = 0$ and $m_i > 0$

that both banks implementing the collusion-free contracts is an equilibrium strategy only if they face a high-quality borrower, i.e., $z \geq z_2^C$. The reason is that, under two-bank lending, the cost of providing incentives to deter collusion is low when firm quality is high (it is zero if firm quality is sufficiently high, i.e., $z > \bar{z}_2$) as higher values of z imply lower incentive for vertical collusion.

The last part of Proposition 10 asserts that the results in Proposition 5 are reasonably robust in the sense that when the borrower is more informationally opaque (low z), two-bank lending entails higher per-bank monitoring relative to single-bank lending even if the banks are not constrained to offer collusion-free contracts. Note that $m_j(z, r)$ is decreasing in z, and reaches a minimum of $m_1^*(r) = pr/c$. Thus, monitoring by bank j in the (C, NC) equilibrium is higher than the monitoring level under single-bank lending for $z_2^C \geq z \geq z_j^{min}$. By contrast, the individual monitoring effort $m_2(z, r)$ in the (NC, NC) equilibrium takes a maximum value at z_2^C, and it turns out that $m_2\left(z_2^C, r\right) < m_1^*(r)$. Thus for high values of firm quality, two-bank lending entails lower monitoring effort.

Tressel and Verdier (2011) analyze the failure of collusion-proofness principle under multiple bank lending. In their model, the banking sector consists of many banks and borrowers, and banks, having an advantage in monitoring borrowers, may collude with the latter. Their main finding is that foreign capital flows via FDI exacerbates collusion incentive problem of banks, and that a positive measure of domestic banks collude with domestic firms as FDI makes the domestic bank capital relatively scarce. However, Tressel and Verdier (2011) do not consider the possibility of vertical collusion in hierarchical bank-monitor-borrower relationships, and they do not contemplate the situation in which many banks lend to a single firm.

6.2 Endogenous loan rate

In the baseline model we have not only assumed the loan rate to be exogenous, but also that the rates are the same under the single- and two-bank lending modes in order to make the two lending structures comparable. However, in reality, rates are endogenous and vary according to market structure. In what follows, we endogenize the loan rates under both lending modes. We assume that there is a market for funds in which banks are price takers, and hence, equilibrium loan rates are determined by the zero profit conditions of the banks, rather than being strategic variables. We show that the results presented in Proposition 5 remain valid.

6.2.1 Single-bank lending

We first analyze the equilibrium loan rate under single-bank lending with delegated monitoring. Denote by $R \equiv pr$ the loan rate under single-bank lending. At any R, let $B_1(z, R)$ denote the bank's equilibrium expected profit. The equilibrium loan rate is determined from the zero-profit condition of the bank, which is given by:

$$R_1(z) = \begin{cases} z + \frac{c}{2z} & \text{for } z \in [z_1^0, \bar{z}_1], \\ \sqrt{2c} & \text{for } z \in (\bar{z}_1, z_1^{max}]. \end{cases}$$

The threshold values of firm quality, z_1^0, $\bar{z}_1$ and z_1^{max}, which are functions of c, are derived in the Appendix (see the proof of Proposition 11). Note that $R_1(z)$ is strictly decreasing and convex on $[z_1^0, \bar{z}_1]$, and

as long as m_i not too high so that loan officer i's participation constraint is satisfied.

independent of borrower quality thereafter. Substituting the above expression into those of optimal monitoring and share of repayment described in Proposition 2, we obtain the optimal monitoring intensity and incentives under single-bank lending.

6.2.2 Two-bank lending

Under two-bank lending with delegated monitoring and the aggregate loan amount \$2, consider a time line where the banks first post r_i and r_j, with the subsequent game following the action sequence described in Figure 1. Let $R_i \equiv pr_i$ and $R_j \equiv pr_j$. In order to derive the no-collusion constraint of loan officer i, note that the payoff to the firm from colluding with loan officer i is given by

$$2v - p(2y - r_i - r_j) = R_i + R_j - 2z,$$

which is the maximum stake of collusion for this loan officer. Thus, the no-collusion constraints of loan officers i and j are respectively given by:

$$R_i s_i \geq R_i + R_j - 2z, \tag{NC$_i^0$}$$

$$R_j s_j \geq R_i + R_j - 2z. \tag{NC$_j^0$}$$

The participation constraints are given by:

$$\pi(m_i, m_j)R_i s_i - \frac{1}{2}cm_i^2 \geq 0, \tag{PC$_i^0$}$$

$$\pi(m_i, m_j)R_j s_j - \frac{1}{2}cm_j^2 \geq 0. \tag{PC$_j^0$}$$

Thus, given R_i and R_j, at the optimal contracting stage bank i solves

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)R_i(1 - s_i) - 1, \tag{$\mathscr{M}_i^0$}$$

subject to (PC$_i^0$), (NC$_i^0$) and (F$_i$).

We look at a symmetric situation that each set of constraints — namely, (PC$_i^0$) and (PC$_j^0$), (NC$_i^0$) and (NC$_j^0$), and (F$_i$) and (F$_j$), either bind simultaneously, or are slack together. We analyze a symmetric equilibrium whenever it exists. When the no-collusion constraints bind for each bank, the equilibrium loan rates, derived from the banks' zero-profit conditions, are symmetric which are given by:

$$R_2(z) = \frac{2z(2+3c) - c\left(1 - \sqrt{4z(z-1) - c}\right)}{4(1+c)} \quad \text{for } z \in (z_2^0, \bar{z}_2].$$

Note also that $R_2(z)$ is strictly decreasing and convex in z. By contrast, when the no-collusion constraints are slack, there does not exist any symmetric loan rates in any pure strategy equilibrium, but there is a set of asymmetric loan rates which are given by:

$$\bar{R}_i = \frac{c\left(4(1+c) - \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2} \text{ and } \bar{R}_j = \frac{c\left(4(1+c) + \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}$$

for $z \geq \bar{z}_2$. The above rates are independent of firm quality.[24] Because $\bar{R}_i < \bar{R}_j$, the equilibrium monitoring efforts m_i and m_j, as well as the equilibrium shares s_i and s_j are also asymmetric. The threshold values of firm quality, z_2^0 and $\bar{z}_2$ are derived in the Appendix.

[24] For the solutions to be real, we must assume that $2 < c \leq 1 + \sqrt{5}$. See the Appendix for details.

6.2.3 Comparison of the two lending modes under endogenous loan rates

We compare the equilibrium loan rates and loan officer contracts under the two lending structures when loan rates are endogenously determined.

Proposition 11 *Under endogenous loan rates, two-bank lending exacerbates collusive threats, i.e., $\bar{z}_1 < \bar{z}_2$. Moreover,*

(a) *there is a unique threshold value of firm quality, $z_R \in (z_2^0, \bar{z}_2)$ such that the equilibrium loan rates under two-bank lending are higher than that under single-bank lending if and only if $z < z_R$;*

(b) *There are unique threshold value of firm quality, $z_m, z_\pi, z_s \in (z_2^0, \bar{z}_2)$ such that the equilibrium individual monitoring efforts and aggregate monitoring intensity are higher, and incentives for loan officers are stronger under two-bank lending if and only if $z \leq z_m$, $z \leq z_\pi$ and $z \leq z_s$, respectively.*

The above proposition shows that the result of Proposition 5 is robust even if the loan rates are endogenous. Because each loan officer receives a fraction of the repayments made by the firm, under endogenous loan rates, each bank has an additional instrument to incentivize its loan officer because the incentive pays are sR under single-bank lending, and s_iR_i and s_jR_j under two-bank lending. In the presence of the collusion incentive problem, stronger incentive pays (both loan rate and share) are required to deter collusion under two-bank lending because of the race-to-collusion effect. By contrast, when such incentive problems are absent, under two-bank lending, a lower monitoring is elicited by a weaker incentive pay. The comparison between the equilibrium loan rates can be depicted graphically in a figure similar to Figure 4.

7 Concluding Remarks

Multiple-bank lending is in general viewed as detrimental to efficiency in the sense that when borrower output can only be verified through costly monitoring by lenders, such financial arrangement leads to a free-riding problem in monitoring (e.g. Khalil et al., 2007), thereby lowering the monitoring level of each lender. The present paper identifies an additional source of inefficiency under multiple-bank lending if monitoring activities must be delegated and there is a possibility of vertical collusion between a loan officer and the borrower, namely a countervailing race-to-collusion effect which leads to over-monitoring. Moreover, under multiple-bank lending, the race-to-collusion effect gets exacerbated, so that incentives offered to each monitor are stronger in order to deter collusion, leading to more intense monitoring. The results are shown to be robust even when the loan rates charged to the borrower are endogenously determined, and they differ across lending modes.

Our results have important testable implications both with respect to production of soft or hard information and multiple-bank lending. The critical role that loan officers play in performance monitoring has been recognized in the empirical literature on relationship lending (e.g. Hertzberg et al., 2010; Uchida et al., 2012). In particular, Hertzberg et al. (2010), who analyze the lending relationships of a large multinational U.S. bank in Argentina, show that under a policy whereby loan officers assigned to a particular borrowing firm are rotated frequently, the incumbent officer tends to report more accurately because if his successor produces different but verifiable information, then this may hurt his reputation. Our framework with multiple loan officers under the possibility of vertical collusion resembles that of Hertzberg et al. (2010), although we do not consider a dynamic long-term lender-borrower relationship.

The main objective of our paper has been to focus on the strategic effects of incentive contracts under multiple bank lending, and hence, we have confined attention to situations in which banks are constrained to offer collusion-free contracts to the loan officers. Our assumption that all banks are engaged in monitoring is an example of *consortium lending* when a group of banks lend to a single borrower, each of which independently appraises and monitors the borrower. We have also shown that when banks are not constrained to offer collusion-free contracts to their loan officers, there is an equilibrium where at least one bank finds it optimal not to employ a loan officer, and tends to rely on the information gathered by the loan officer of the other bank. Such scenario resembles a *lending syndicate* in which bank plays the role of a *lead arranger* to whom the responsibility of monitoring is delegated. Gangopadhyay and Mukhopadhyay (2002) show that syndicated lending emerges when banks have asymmetric monitoring costs. By contrast, we provide an equilibrium argument in favor of syndicated lending even when the lenders are identical with respect to monitoring costs, and show that when the borrower is more informationally opaque, lending syndicates are the preferred lending arrangement over consortium lending.

Appendices

A Proofs

Proof of Proposition 2

Consider the maximization problem ($\mathscr{M}$). We first argue that at the optimum neither $m = 0$ nor $s = 1$ is an optimal solution. Both at $m = 0$ or at $s = 1$, $B(m, s) = -1 < 0$, and hence, the bank is better-off by not lending. Moreover, $s = 0$ cannot be optimal too because this would violate (NC) for any $z \leq pr$. Thus, the only relevant feasibility constraint we are left with is $m \leq 1$. The Lagrangean is given by:

$$\mathscr{L} = \underbrace{mpr(1-s)-1}_{B(m,s)} + \mu_P \underbrace{\left(mprs - \frac{1}{2}cm^2\right)}_{g^P(m,s)} + \mu_N \underbrace{(prs-(pr-z))}_{g^N(m,s)} + \mu_F \underbrace{(1-m)}_{g^F(m,s)},$$

where μ_P, μ_N and μ_F are the associated Lagrange multipliers. The Karush-Kuhn-Tucker first-order conditions are given by:

$$\frac{\partial \mathscr{L}}{\partial m} = pr(1-s) + \mu_P(prs - cm) - \mu_F = 0, \tag{2}$$

$$\frac{\partial \mathscr{L}}{\partial s} = m(\mu_P - 1) + \mu_N = 0, \tag{3}$$

$$\mu_k g^k(m, s) = 0 \quad \text{for } k = P, N, F, \tag{4}$$

$$g^k(m, s) \geq 0 \quad \text{for } k = P, N, F, \tag{5}$$

$$\mu_k \geq 0 \quad \text{for } k = P, N, F. \tag{6}$$

First, we consider the case when (F) binds, i.e., $\mu_F > 0$ and $m = 1$. We argue that, in this case, (NC) must bind. Suppose not, i.e., $\mu_N = 0$. Substituting, $m = 1$ and $\mu_N = 0$ into (3), we obtain that $\mu_P = 1$ because

$pr > 0$. Substituting $\mu_P = 0$ into (2), we obtain $pr - c = \mu_F > 0$ which contradicts our assumption that $pr \leq c$. Given that $m = 1$ and $prs = pr - z$, the multipliers μ_P, μ_N and μ_F are determined from the two first-order conditions (2) and (3), and the complementary slackness condition associated with the participation constraint, i.e.,

$$z + \mu_P(pr - z - c) = \mu_F, \tag{7}$$

$$\mu_P + \mu_N = 1, \tag{8}$$

$$\mu_P\left(pr - z - \frac{c}{2}\right) = 0. \tag{9}$$

The unique set of solutions to the above system is given by $\{\mu_P = 0,\, \mu_N = 1,\, \mu_F = z\}$. Because $\mu_P = 0$, the participation constraint, (PC) holds, which also implies that

$$pr - z - \frac{c}{2} \geq 0 \iff z \leq pr - \frac{c}{2} \equiv z_1^0.$$

We now prove that $\mu_F = z \geq 1$. Suppose not, i.e., $z < 1$; then the bank's payoff is given by $B_1(z, r) = mpr(1-s) - 1 = z - 1 < 0$. This is not possible, and hence, $z \geq 1 \equiv z_1^{min}$.

Next, consider the case when (F) does not bind, i.e., $m < 1$. In this case, we have $\mu_F = 0$. We argue that, in this case, (PC) must be binding. If, on the contrary, (PC) does not bind, then $\mu_P = 0$, and hence, (2) implies $s = 1$, and $B(m, s) = -1 < 0$. Thus, the bank is better-off by not lending. Given that the participation constraint binds at the optimum, there are two sub-cases.

(a) First, suppose that (NC) binds, i.e., $prs = pr - z$. Substituting the value of s into the binding (PC), we obtain

$$m = \frac{2(pr - z)}{c}.$$

Because (F) is slack, we have

$$m = \frac{2(pr - z)}{c} < 1 \iff z > z_1^0 = pr - \frac{c}{2}.$$

Substituting $m = 2(pr - z)/c$, $prs = pr - z$ and $\mu_F = 0$ into (2) and (3), we obtain

$$\mu_P = \frac{z}{pr - z} \quad \text{and} \quad \mu_N = \frac{2(pr - 2z)}{c}.$$

Now, $\mu_N \geq 0$ implies $z \leq pr/2 \equiv \bar{z}_1$.

(b) Finally, consider the sub-case when (NC) does not bind at the optimum, i.e., $\mu_N = 0$. Thus, substituting $\mu_N = \mu_F = 0$ into (2) and (3), we obtain

$$pr(1 - s) + \mu_P(prs - cm) = 0, \tag{10}$$

$$m(\mu_P - 1) = 0. \tag{11}$$

From (11) it follows that either $\mu_P = 1$ or $m = 0$. The solution $m = 0$ is not feasible because $B(0, s) = -1 < 0$, and hence $\mu_P = 1$. Substituting $\mu_P = 1$ into (10), we obtain $m = pr/c$. From binding (PC) we get $s = 1/2$. Because (NC) is slack, we have $pr/2 > pr - z \iff z > \bar{z}_1$.

Proof of Lemma 1

Consider the maximization problem of Bank i. Note first that $s_i = 0$ cannot be a solution to the maximization problem because this would violate (NC$_i$). On the other hand, $s_i = 1$ implies that the bank cannot break even, and hence, this cannot be an optimal solution. Note also that $m_i = 0$ is not an optimal solution because, given that $s_i > 0$, the participation constraint is slack to begin with, so that any $m_j > 0$ that satisfies the participation constraint (PC$_i$) would improve bank i's expected payoff. Thus, the only relevant feasibility constraint is given by:

$$m_i \leq 1. \tag{F$'_i$}$$

Therefore the Lagrangean of bank i is given by:

$$\mathscr{L}_i = \underbrace{\pi(m_i, m_j)pr(1-s_i) - 1}_{B_i(m_i,m_j,s_i)} + \mu_P^i \underbrace{\left(\pi(m_i, m_j)prs_i - \frac{1}{2}cm_i^2\right)}_{g_i^P(m_i,m_j,s_i)} + \mu_N^i \underbrace{(prs_i - 2(pr-z))}_{g_i^N(m_i,m_j,s_i)} + \mu_F^i \underbrace{(1-m_i)}_{g_i^F(m_i,m_j,s_i)},$$

where μ_P^i, μ_N^i and μ_F^i are the associated Lagrange multipliers. The Karush-Kuhn-Tucker first-order conditions (for bank i) are given by:

$$\frac{\partial \mathscr{L}_i}{\partial m_i} = (1-m_j)pr(1-s_i) + \mu_P^i\left[(1-m_j)prs_i - cm_i\right] - \mu_F^i = 0, \tag{12}$$

$$\frac{\partial \mathscr{L}_i}{\partial s_i} = \pi(m_i, m_j)(\mu_P^i - 1) + \mu_N^i = 0, \tag{13}$$

$$\mu_k^i g_i^k(m_i, m_j, s_i) = 0 \quad \text{for } k = P, N, F, \tag{14}$$

$$g_i^k(m_i, m_j, s_i) \geq 0 \quad \text{for } k = P, N, F, \tag{15}$$

$$\mu_k^i \geq 0 \quad \text{for } k = P, N, F. \tag{16}$$

Bank j has similar optimality conditions. We first show that both $m_i < 1$ and $m_j < 1$ at the optimum. Suppose to the contrary that $m_j = 1$. Then, (12) reduces to

$$-\mu_P^i \cdot cm_i = \mu_N^i.$$

Given that $m_i > 0$, the above equation implies that μ_P^i and μ_N^i are of opposite signs which contradicts the non-negativity conditions in (16). Therefore $m_j < 1$ and $\mu_F^j = 0$. Similarly, $\mu_F^i = 0$.

Next, we show that both the participation constraint must bind at the optimum. Suppose at least one of them, say (PC$_i$), is slack, and hence, $\mu_P^i = 0$. Given that $\mu_F^i = 0$, (12) reduces to

$$(1-m_j)pr(1-s_i) = 0.$$

Given that $m_j < 1$, we must have $s_i = 1$; but this implies $B_i(\cdot) = -1$ which is not feasible. Therefore the participation constraints bind.

Next the binding participation constraint of loan officer i yields

$$prs_i = \frac{cm_i^2}{2\pi(m_i, m_j)}. \tag{17}$$

Substituting for s_i into the objective function ($\mathscr{M}_i$) and the no-collusion constraint (NC_i), the above maximization problem of bank i reduces to:

$$\max_{m_i} \ \pi(m_i, m_j)pr - \frac{1}{2}\,cm_i^2 - 1, \tag{$\mathscr{M}_i'$}$$

$$\text{subject to} \quad \frac{cm_i^2}{2\pi(m_i, m_j)} \geq 2(pr - z). \tag{NC_i'}$$

Given that $\partial\pi/\partial m_i = 1 - m_j$, when the no-collusion constraint of neither loan officer binds at the optimum, the first-order conditions of the maximization problems of banks i and j yield the best reply functions $m_i(m_j)$ and $m_j(m_i)$ that solve the following two equations, respectively:

$$pr(1 - m_j) - cm_i = 0, \tag{BR_i}$$

$$pr(1 - m_i) - cm_j = 0. \tag{BR_j}$$

Note that $m_i'(m_j) = m_j'(m_i) = -pr/c < 0$, and hence, m_i and m_j are strategic substitutes.

Next we prove that, when both the no-collusion constraints bind, m_i and m_j are strategic complements for $m_i, m_j \in [0,\, 1]$. In this case the best reply functions $m_i(m_j)$ and $m_j(m_i)$ solve the following two equations, respectively.

$$cm_i^2 = 2a\pi(m_i, m_j), \tag{BR_i'}$$

$$cm_j^2 = 2a\pi(m_i, m_j), \tag{BR_j'}$$

where $a \equiv 2(pr - z) \geq 0$. Analyzing the behavior of $m_i(m_j)$ suffices to prove the assertion as the behavior of $m_j(m_i)$ is symmetric. Note first that (BR_i') can be written as

$$\frac{m_i^2}{m_i + m_j - m_i m_j} = \frac{2a}{c}$$

The left-hand-side of the above equation is always less than 1 because $m_i^2 \leq m_i + m_j - m_i m_j$ is equivalent to $(1 - m_i)(m_i + m_j) \geq 0$. Therefore, in equilibrium we must have

$$2a \leq c. \tag{18}$$

Solving for m_i from (BR_i') we get

$$m_i(m_j) = \underbrace{\frac{a(1 - m_j) - \sqrt{a^2(1 - m_j)^2 + 2acm_j}}{c}}_{m_i^-(m_j)},\ \underbrace{\frac{a(1 - m_j) + \sqrt{a^2(1 - m_j)^2 + 2acm_j}}{c}}_{m_i^+(m_j)}.$$

Note that for any $m_j \geq 0$, $m_i^-(m_j) \leq 0$, and hence, we discard this root. On the other hand, $m_i^+(m_j) \geq 0$ for all $m_j \in [0,\, 1]$. From the expression of $m_i^+(m_j)$ we get

$$m_i'(m_j) = \frac{a}{c}\left(\frac{c - a(1 - m_j)}{\sqrt{a^2(1 - m_j)^2 + 2acm_j}} - 1\right)$$

The above expression is positive if and only if $2a \geq c$, which holds from (18).

Proof of Proposition 4

The proof is very similar to that of Proposition 2, and hence, we would omit the details. Instead, we will describe the threshold values z_2^0 and $\bar{z}_2$, and the optimal values of (m_i, m_j), (s_i, s_j) and $\pi(m_i, m_j)$. The equilibrium is symmetric, and hence, $m_i = m_j = m_2$ and $s_i = s_j = s_2$. We have already shown that the participation constraints (PC$_i$) and (PC$_j$) are both binding. We have two cases—(a) (NC$_i$) and (NC$_j$) both bind, and (b) neither of them binds. First, consider the case when the no-collusion constraints bind. The optimal m_i and m_j are solutions to the system (BR$'_i$) and (BR$'_j$), which has two symmetric solutions

$$(0, 0) \quad \text{and} \quad \left(\frac{8(pr-z)}{c+4(pr-z)}, \frac{8(pr-z)}{c+4(pr-z)}\right),$$

and no asymmetric solutions. At $m_i = m_j = 0$, each bank's expected profit equals -1, and hence, this solution is not optimal. The optimal monitoring efforts m_i and m_j are thus given by the other symmetric solution, and hence,

$$\pi(m_i, m_j) = 1 - (1-m_2)^2 = 1 - \left(\frac{c-4(pr-z)}{c+4(pr-z)}\right)^2.$$

Using (17), we get the optimal share which is given by:

$$s_i = s_j = 2\left(1 - \frac{z}{pr}\right).$$

Note that because the feasibility constraints must be slack, we require

$$m_2(z, r) = \frac{8(pr-z)}{c+4(pr-z)} < 1 \quad \iff \quad z > pr - \frac{c}{4} \equiv z_2^0.$$

Let $B_2(z, r) \equiv \pi(z, r)pr[1 - s_2(z, r)] - 1$ denote each bank's expected payoff. Note that $B_2(z_2^0, r) = pr - c/2 - 1$ which is positive, given Assumption 1.

Next, consider the case when neither (NC$_i$) nor (NC$_j$) binds. Because both $m_i(m_j)$ and $m_j(m_i)$ defined by (BR$_i$) and (BR$_j$) are linear and downward-sloping, there is a unique solution to the system of equations, which is also symmetric. This is given by:

$$m_i = m_j = m = \frac{pr}{c+pr}.$$

Thus,

$$\pi(m_i, m_j) = 1 - (1-m_2)^2 = 1 - \left(\frac{c}{c+pr}\right)^2.$$

Using (17), we get the optimal share which is given by:

$$s_i = s_j = s = \frac{c}{2(2c+pr)}.$$

The non-binding no-collusion constraints imply

$$\frac{c}{2(2c+pr)} > 2(pr-z) \iff z > \frac{pr(7c+4pr)}{4(2c+pr)} \equiv \bar{z}_2.$$

This completes the proof of the proposition.

Proof of Proposition 5

We first show that $\bar{z}_2 > \bar{z}_1$ which is equivalent to $3c + 2pr > 0$. The last inequality always holds because both pr and c are strictly positive. To be consistent with Figure 4, it is also necessary to show that $\bar{z}_1 < z_2^0$ which is equivalent to $pr > c/2$. The last inequality holds because of Assumption 1.

To show Part (a), note that because $\bar{z}_1 < z_2^0$, we have $m_1(z_2^0, r) = pr/c$. Thus, $m_2(z_2^0, r) = 1 \geq pr/c = m_1(z_2^0, r)$. On the other hand, $m_2(\bar{z}_2, r) = pr/(c+pr) < pr/c = m_1(\bar{z}_2, r)$. Because $m_2(z, r)$ is strictly decreasing on $[z_2^0, \bar{z}_2]$, there is a unique $z_m \in (z_2^0, \bar{0}_2)$ such that $m_2(z, r) > m_1(z, r)$ if and only if $z < z_m$.

For Part (b), note that $s_2(z_2^0, r) = c/2pr \geq 1/2 = s_1(z_2^0, r)$ because $pr \leq c$, and $s_2(\bar{z}_2, r) = (c/2)(2c + pr) < 1/2 = s_1(\bar{z}_2, r) \iff c + pr > 0$. As $s_2(z, r)$ is strictly decreasing on $[z_2^0, \bar{z}_2]$, there exists a unique $z_s \in (z_2^0, \bar{z}_2)$ such that $s_2(z, r) > s_1(z, r)$ if and only if $z < z_m$.

Proof of Proposition 6

Note that $\pi(z_2^0) = 1 > pr/c = m_1(z_2^0)$. On the other hand,

$$\pi(\bar{z}_2, r) < m_1(\bar{z}_2, r) \iff \frac{pr(2c+pr)}{(c+pr)^2} < \frac{pr}{c} \quad \iff \quad p^2r^2 + cpr - c^2 > 0.$$

The above inequality holds for two sets of parameter values—(i) $2 < c \leq 2(2+\sqrt{5}) \approx 8.47$ and $1 + \frac{c}{2} \leq pr \leq c$; (ii) $c > 8.47$ and $0.62c \leq pr \leq c$. It is also worth noting that the above inequality is reversed, i.e., $\pi(z, r) \geq m_1(z, r)$ for all $z \in [z_2^0, pr]$ if $c > 8.47$ and $1 + \frac{c}{2} \leq pr \leq 0.62c$.

Proof of Proposition 7

The merged bank maximizes joint profits, but offer individualized contracts (m_i, s_i) and (m_j, s_j) to loan officers i and j, respectively. Therefore, we can solve the optimal contract to each loan officer separately. Therefore, the optimal contract (m_i, s_i) between loan officer i and the merged entity solves

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)pr(2 - s_i - s_j) - 2, \qquad (\mathscr{M}_{ij})$$
$$\text{subject to (PC}_i\text{), (NC}_i\text{), (F}_i\text{).}$$

Denote the threshold values of firm quality and the equilibrium variables with subscript '*mb*'. In a symmetric equilibrium, as in Proposition 4, when both no-collusion constraints (NC_i) and (NC_j) bind, the optimal contracts coincide with those under two bank lending. Therefore, we have $z_{mb}^0 = z_2^0$, $m_{mb}(z, r) = m_2(z, r)$, $\pi_{mb}(z, r) = \pi(z, r)$ and $s_{mb}(z, r) = s_2(z, r)$. This is because the race-to-collusion effect has the same strength under both lending structures. By contrast, when none of the no-collusion constraints binds, the equilibrium contracts under merged banks differ from those under two-bank lending (strategic banks). The reason is simple. When the no-collusion constraints are slack, under both lending modes, we can ignore both the no-collusion and feasibility constraints, and use the binding participation constraints to write

$$\pi(m_i, m_j)prs_i = \frac{1}{2}cm_i^2, \quad \text{and} \quad \pi(m_i, m_j)prs_j = \frac{1}{2}cm_j^2.$$

When banks choose the contracts independently, substituting the above expressions into the objective functions, the payoffs of banks i and j respectively reduce to:

$$B_i(m_i, m_j) \equiv \pi(m_i, m_j)pr - \frac{1}{2}cm_i^2 - 1, \quad \text{and} \quad B_j(m_i, m_j) \equiv \pi(m_i, m_j)pr - \frac{1}{2}cm_j^2 - 1.$$

On the other hand, the objective function of the merged entity [under binding participation constraints] becomes

$$B(m_i, m_j) \equiv 2\pi(m_i, m_j)pr - \frac{1}{2}cm_i^2 - \frac{1}{2}cm_j^2 - 2.$$

Note that, under strategic banks, each bank maximizes the net surplus of the bank-monitor relationship, i.e., bank's expected revenue minus the sum of monitoring cost and the opportunity cost of capital. By contrast, under merged banks, they maximize the joint expected revenue minus the aggregate monitoring and opportunity costs. Clearly under merged banks, the aggregate surplus is higher due to the absence of the free-riding problem, and hence, monitoring efforts, shares and the aggregate monitoring intensity are higher than those under strategic banks. Thus, under merged banks, in the absence of collusion incentive problem, we have

$$m_{mb}(z, r) = \frac{2pr}{c + 2pr}, \; \pi_{mb}(z, r) = 1 - \left(\frac{c}{c + 2pr}\right)^2, \text{ and } s_{mb}(z, r) = \frac{c}{2(c + pr)}.$$

Each of the above expressions are strictly higher than that under two-bank lending. It is also the case that $\bar{z}_{mb} < \bar{z}_2$, i.e., incentives for over-monitoring are lower under merged banks which implies an efficiency gain.

Proof of Proposition 8

When the banks choose contracts independently but the loan officers behave cooperatively, each bank i chooses (m_i, s_i) to solve

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)pr(1 - s_i) - 1, \qquad (\mathscr{M}_i)$$

subject to (PC_i), (NC_{ij}), (F_i).

We analyze a symmetric equilibrium, i.e., $m_i = m_j$ and $s_i = s_j$. When the no-collusion constraint (NC_{ij}) is slack, the equilibrium contracts coincide with those under two-bank lending. When (NC_{ij}) binds, we have $s_i = s_j = 1 - z/pr$. It follows from the binding participation constraints that

$$m_i = m_j = m_{cp}(z, r) = \frac{4(pr - z)}{c + 2(pr - z)} \quad \text{and} \quad \pi_{cp}(z, r) = 1 - \left(\frac{c - 2(pr - z)}{c + 2(pr - z)}\right)^2.$$

We must have $m_{cp}(z, r) \leq 1$ which is equivalent to $z \geq pr - c/2 \equiv z_{cp}^0 = z_1^0 < z_2^0$. The threshold value of borrower quality $\bar{z}_{cp}$ solves

$$\frac{c}{2(2c + pr)} = 1 - \frac{z}{pr},$$

whereas $\bar{z}_2$ solves

$$\frac{c}{2(2c + pr)} = 2\left(1 - \frac{z}{pr}\right).$$

Thus, $\bar{z}_{cp} < \bar{z}_2$, i.e., two-bank lending under cooperative loan officers ameliorates the collusion incentive problem relative to two-bank lending with competing monitors. As a result, there is a gain in efficiency under cooperative loan officers. It is immediate to show that $m_{cp}(z, r) < m_2(z, r)$ and $\pi_{cp}(z, r) < \pi(z, r)$ given that $c > 0$. Clearly, $s_{cp}(z, r) = 1 - z/pr < 2(1 - z/pr) = s_2(z, r)$.

Proof of Proposition 9

We first prove part (a). Note first that $W(m_{mb}(z, r))$ is strictly increasing in z for $z \in [z_2^0, \bar{z}_{mb}]$, and is constant on $[\bar{z}_{mb}, pr]$ because m_{mb} is independent of z (the function is strictly decreasing for $z \geq \bar{z}_{mb}$; however, this part lies strictly below the constant function). Call the constant part $\overline{W}_{mb}$. On the other hand, $W(m_2(z, r))$ is strictly concave on $[z_2^0, \bar{z}_2]$ which reaches its maximum at a unique $\tilde{z}_2 \in (\bar{z}_{mb}, \bar{z}_2)$, and is constant on $[\bar{z}_2, pr]$. First, $W(m_{mb}(z, r)) = W(m_2(z, r))$ for all $z \in [z_2^0, \bar{z}_{mb}]$ because $m_{mb}(z, r) = m_2(z, r)$ for these values of firm quality. It is easy to show that $W(m_2(\tilde{z}_2, r)) > \overline{W}_{mb}$ and $W(m_2(\bar{z}, r)) < \overline{W}_{mb}$. The last two inequalities imply (i) $W(m_2(z, r)) > W(m_{mb}(z, r))$ on $[\bar{z}_{mb}, \tilde{z}_2]$, and (ii) there is a unique $\tilde{z} \in (\tilde{z}_2, \bar{z}_2)$ such that $W(m_2(z, r)) > W(m_{mb}(z, r))$ if and only if $z \in [\tilde{z}_2, \tilde{z})$ because $W(m_2(z, r))$ is strictly decreasing on $[\tilde{z}_2, \bar{z}_2]$. This completes the proof of part (a).

Recall that the relevant range of values of z for the welfare comparison is $[z_2^0, pr]$. As $W(m_{mb}(z, r)$, $W(m_{cp}(z, r)$ is strictly increasing in z for $z \in [z_2^0, \bar{z}_{cp}]$, and is constant on $[\bar{z}_{cp}, pr]$. On the other hand, $W(m_2(z, r))$ is strictly concave on $[z_2^0, \bar{z}_2]$ which reaches its maximum at a unique $\tilde{z}_2 \in (\bar{z}_{cp}, \bar{z}_2)$, and is constant on $[\bar{z}_2, pr]$. The rest of the proof is very similar to that of part (a) with the only exception that for small values of z, $W(m_2(z, r)) < W(m_{cp}(z, r))$, and $W(m_2(z, r)) = W(m_{cp}(z, r))$ for all $z \geq \bar{z}_2$.

Proof of Proposition 10

As discussed earlier, the outcome where both loan officers allow collusion is not an equilibrium because then the banks will not receive any repayments, and they would lose their entire investment. Next, consider any $z > \bar{z}_2$. For these values of firm quality z, in the (NC, NC) equilibrium described in Proposition 4, neither (NC_i) nor (NC_j) binds, so that the banks do not worry about the incentive problems emerging from collusion possibilities. Therefore the equilibrium in Proposition 4 is robust even if banks can write contracts that allow collusion. Therefore, we concentrate on firm quality $z \leq \bar{z}_2$ in order to analyze the equilibrium. We proceed via several lemmas.

Lemma 2 *There is a unique threshold value of firm quality $z_2^C < \bar{z}_2$ such that for any $z \geq z_2^C$ there is an equilibrium of type (NC, NC).*

Proof Consider an (NC, NC) equilibrium and we examine if bank i, say, wants to allow collusion or not. Recall that a collusion possibility emerges if loan officer i is successful in monitoring, whereas loan officer j is not. Thus if the equilibrium involves officer i colluding, then collusion occurs with probability $m_i(1 - m_j)$. The Nash bargaining solution for the optimal bribe, for a given distribution $(1 - \beta, \beta)$ of bargaining power between the firm and loan officer i, is given by:

$$b_i^*(z) = \max\{0, 2\beta(pr - z) + (1 - \beta)prs_i\}.$$

Thus, the participation constraint of loan officer i is given by:

$$[m_j + (1-\beta)m_i(1-m_j)]prs_i + 2\beta m_i(1-m_j)(pr-z) - \frac{1}{2}cm_i^2 \geq 0. \tag{PC$_i^C$}$$

Given that bank i does not want to implement a collusion-free contract it would optimally would choose $s_i = 0$. From the perspective of bank i, the probability of success is now given by m_j because it does not rely anymore on the information gathered by loan officer i. Thus, bank i solves

$$\max_{m_i\in[0,1]} m_j pr - 1, \tag{$\mathscr{M}_i^C$}$$
$$\text{subject to (PC}_i^C\text{)}.$$

In what follows we ignore the participation constraint, since the payoff of bank i does not depend on m_i. Recall that in the (NC, NC) equilibrium both banks monitor at the level $m_2(z, r) = 8(pr-z)/(c+4(pr-z))$ [cf. Proposition 4]. Thus the deviation payoff for bank i is computed by substituting $m_j = m_2(z, r) = 8(pr-z)/(c+4(pr-z)$, which is given by:

$$\tilde{B}_i(z, r) = \frac{8pr(pr-z)}{c+4(pr-z)} - 1.$$

Therefore, for bank i, it is not profitable to deviate from the (NC, NC) equilibrium in Proposition 4 and allow for collusion if and only if

$$B_2(z, r) \geq \tilde{B}_i \iff z \geq \frac{pr(3c+4pr)}{4(c+pr)} \equiv z_2^C,$$

where note that we ignore the participation constraint (PC$_i^C$).[25]

To summarize, for all $z \geq z_2^C$, the equilibrium denoted in Proposition 4 continues to be one, so that the collusion-proofness principle holds in the sense that there is one equilibrium where imposing the NC constraints are without loss of generality. It is immediate to show that $z_2^C < \bar{z}_2$. In fact, $z_2^C = \bar{z}_{mb}$ [cf. Figure 5] at which $B_2(z)$ is maximized. $\square$

We next establish that there exists a cutoff value of firm quality, $\hat{z}_2$, such that for any $z \leq \hat{z}_2$, there exists an equilibrium where one bank monitors, whereas the other bank free rides. Furthermore, we show that for intermediate values of z multiple equilibria exist.

Lemma 3 *There is a unique threshold value of firm quality $\hat{z}_2$, with $z_2^C < \hat{z}_2 \leq \bar{z}_2$, such that for any $z \leq \hat{z}_2$ there is an equilibrium of type (C, NC).*

Proof For any $z \leq z_2^C$, there cannot be an equilibrium of type (NC, NC) described in Proposition 4. Thus the only possible equilibria are of type (C, NC). Clearly, it suffices to analyze an equilibrium where

[25] The participation constraint (PC$_i^C$) holds at $m_j = m_2(z, r)$, i.e.,

$$m_i \leq \frac{4\beta(pr-z)(c-4(pr-z))}{c(c+4(pr-z))} \equiv \bar{m}_i.$$

Note that $\bar{m}_i > 0$ if and only if $c - 4(pr-z) > 0$ i.e., $z > z_2^0$. If $z \leq z_2^0$, then it is optimal for bank i to set $m_i = 0$, so that the bank does not employ a loan officer at all.

bank i does not intend to deter collusion, but bank j offers a collusion-proof contract. We next turn to characterizing such a candidate equilibrium. In such a situation, no bank gains from monitoring by loan officer i, and hence, $\pi(m_i, m_j) = m_j$. Thus, bank j solves

$$\max_{(m_j, s_j)\in[0,1]^2} m_j pr(1-s_j) - 1, \tag{$\mathscr{M}_j^C$}$$

$$\text{subject to } m_j prs_j - \frac{1}{2}cm_j^2 \geq 0, \tag{PC$_j^C$}$$

$$prs_j \geq 2(pr-z). \tag{NC$_j$}$$

The above program is similar to the one of single-bank lending with the only modification that the no-collusion constraint is (NC$_j$) instead of (NC). The optimal monitoring effort of loan officer j is given by:

$$m_j^C(z, r) = \begin{cases} 1 & \text{for } z_j^{min} \leq z < z_2^0, \\ \frac{4(pr-z)}{c} & \text{for } z_2^0 \leq z < \bar{z}_j, \\ \frac{pr}{c} & \text{for } \bar{z}_j \leq z \leq \bar{z}_2, \end{cases}$$

where $z_j^{min} = (pr+1)/2$ such that $B_j(m_j^C(z, r)) \geq 0$ if and only if $z \geq z_j^{min}$, and $\bar{z}_j = 3pr/4 < \bar{z}_2$. As bank i allows collusion in this candidate equilibrium, it optimally sets $s_i = 0$, and hence, its expected payoff is given by $m_j pr - 1$. Thus, the expected payoff of bank i is given by:

$$B_i^C(z, r) \equiv B_i(m_j^C(z, r)) = \begin{cases} pr - 1 & \text{for } z_j^{min} \leq z < z_2^0, \\ \frac{4pr(pr-z)}{c} - 1 & \text{for } z_2^0 \leq z < \bar{z}_j, \\ \frac{p^2r^2}{c} - 1 & \text{for } \bar{z}_j \leq z \leq \bar{z}_2. \end{cases}$$

In order to establish that such an equilibrium exists, we have to check that, given $m_j^C(z, r)$, bank i has no incentive to deviate to a collusion free contract, i.e. one that imposes an NC constraint. We first calculate bank i's payoff from such a deviation when it solves the collusion-proof maximization problem ($\mathscr{M}_i$) taking into account that $m_j = m_j^C(z)$.

First, consider the case when (NC$_i$) binds. In this case, the best reply $m_i(m_j^C(z))$ solves (BR$_i$), which is given by:

$$\tilde{m}_i(z, r) = \begin{cases} m_i(1) & \text{for } z_j^{min} \leq z < z_2^0, \\ m_i\left(\frac{4(pr-z)}{c}\right) & \text{for } z_2^0 \leq z < \bar{z}_j, \\ m_i\left(\frac{pr}{c}\right) & \text{for } \bar{z}_j \leq z \leq \bar{z}_2. \end{cases}$$

Next, consider the case when (NC$_i$) does not bind. In this case, the best reply $m_i(m_j^C(z))$ solves (BR$_i'$), which is given by:

$$\bar{m}_i(z, r) = \begin{cases} m_i(1) = 0 & \text{for } z_j^{min} \leq z < z_2^0, \\ m_i\left(\frac{4(pr-z)}{c}\right) = pr\left(1 - \frac{4(pr-z)}{c}\right) & \text{for } z_2^0 \leq z < \bar{z}_j, \\ m_i\left(\frac{pr}{c}\right) = pr\left(1 - \frac{pr}{c}\right) & \text{for } \bar{z}_j \leq z \leq \bar{z}_2. \end{cases}$$

One can show that $\tilde{m}_i(z, r)$ and $\bar{m}_i(z, r)$ intersect each other once at $z = \bar{z}_i > \bar{z}_2$.[26] Thus, for any $z \in [z_j^{min} < \bar{z}_2]$, $\tilde{m}_i(z, r)$ lies strictly above $\bar{m}_i(z, r)$. Note that if $\bar{m}_i(z, r) < \tilde{m}_i(z, r)$, then the optimal

[26]The computation of the best replies is cumbersome. We have used *Mathematica* to analyze the deviation on behalf of bank i, whose codes are available from the authors upon request.

monitoring level is given by $\tilde{m}_i(z,r)$ becuase for any lower level of monitoring the no-collusion constraint would be violated (as it binds at $\tilde{m}_i(z,r)$ and there is a monotonic relationship between m_i and s_i). Whereas if $\bar{m}_i(z,r) > \tilde{m}_i(z,r)$, then the optimal monitoring level is given by $\bar{m}_i(z,r)$ because $\bar{m}_i(z,r)$ is the unconstrained solution and the NC constraint does not bind. Hence, the monitoring level set by bank i in the above deviation strategy is given by $m_i^{dev}(z,r) \equiv \max\{\bar{m}_i(z,r), \tilde{m}_i(z,r)\} = \tilde{m}_i(z,r)$, which is depicted in the left panel of Figure 8 (the dark solid curve).

We next turn to comparing bank i's payoff from allowing collusion, i.e. $B_i^C(z,r)$, with $B_i^{dev}(z,r)$, its payoff from the deviation strategy $m_i^{dev}(z,r)$. Note that $B_i^C(z,r)$ is decreasing in z. However, because the bank's payoff is strictly decreasing in monitoring, $B_i^{dev}(z,r)$ is strictly increasing in z for all $z \le \bar{z}_i$, and constant with respect to z for all $z > \bar{z}_i$ (since for $z > \bar{z}_i$, we have that $z > \bar{z}_2$, and thus $\bar{m}_i(z,r)$ is constant in z). It is easy to show that $B_i^{dev}(z_2^C, r) < B_i^C(z_2^C, r) = (p^2r^2/c) - 1$. Hence there is a unique $z^{dev} > z_2^C$ such that $B_i^C(z,r) \ge B_i^{dev}(z,r)$ if and only if $z \le z^{dev}$.

Finally define $\hat{z}_2 = \min\{z^{dev}, \bar{z}_2\}$. Clearly, for all $z \le \hat{z}_2$ there is an equilibrium of type (C, NC). Therefore, (a) for all $z \le z_2^C$, (C, NC) is the unique equilibrium , and (b) for any $z \in [z_2^C, \hat{z}_2]$, (C, NC) is also an equilibrium along with (NC, NC). The above analysis is depicted in the right panel of Figure 8 in the case when $\hat{z}_2 = z^{dev}$. $\square$

Combining the preceding two lemmas, we have Proposition 10.

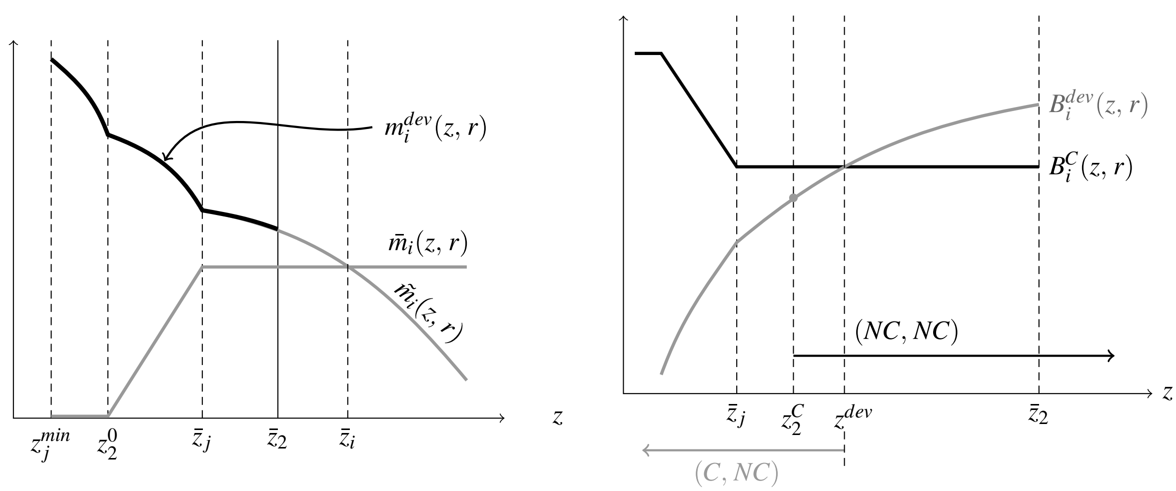

Figure 8: *In the left panel, the dark solid curve represents $m_i^{dev}(z,r)$. The right panel depicts $B_i^C(z,r)$ and $B_i^{dev}(z,r)$. For high values of firm quality, i.e., $z > z^{dev}$ both banks offer collusion-free contracts, whereas for all $z < z_2^C$ at least one bank does not offer collusion-free contract. For any $z \in [z_2^C, z^{dev}]$, there are two equilibria—namely, (NC, NC) and (C, NC).*

Proof of Proposition 11

Equilibrium loan rate under single-bank lending. The bank's expected profit under single-bank lending is given by $B_1(z,R) = m_1(z,R)R(1 - s_1(z,R)) - 1$. We must consider three cases in solving the optimal bank-monitor contract. First, when the feasibility constraint $m \le 1$ binds, i.e., $m_1(z,R) = 1$, the no-collusion constraint binds, and the participation constraint does not. In this case, $s_1(z,R) = 1 - z/R$.

Thus, the bank's payoff reduces to

$$B_1(z,R) = z - 1.$$

Thus, the bank earn zero profits at $z = 1 = z_1^{min}$. Clearly, we must set $z_1^{min} = z_1^0$. Next, consider the case when $m_1(z,R) \leq 1$. We first analyze the sub-case when both the participation and no-collusion constraints bind. Thus, $s_1(z,R) = 1 - z/R$ and $m_1(z,R) = 2(R-z)/c$, given which, the bank's payoff is given by:

$$B_1(z,R) = \frac{2z(R-z)}{c} - 1.$$

Setting the above equal to zero we obtain

$$R_1(z) = z + \frac{c}{2z},\ m_1(z) \equiv m_1(z, R_1(z)) = \frac{1}{z} \quad \text{and} \quad s_1(z) \equiv s_1(z, R_1(z)) = \frac{c}{c+2z^2}.$$

Finally, consider the sub-case (for $m \leq 1$) when the participation constraint binds, but the no-collusion constraint does not. In this case, $m_1(z,R) = R/c$ and $s_1(z,R) = 1/2$, and hence, the bank's profit in this case are given by:

$$B_1(z,R) = \frac{R^2}{2c} - 1.$$

Thus, zero-profit condition yield

$$\bar{R}_1 = \sqrt{2c},\ \bar{m}_1 \equiv m_1(z, \bar{R}_1) = \sqrt{\frac{2}{c}} \quad \text{and} \quad \bar{s}_1 = \frac{1}{2}.$$

The other threshold value of firm quality, $\bar{z}_1$ is determined by solving

$$m_1(z) = \bar{m}_1 \quad \Longleftrightarrow \quad \frac{1}{z} = \sqrt{\frac{2}{c}} \quad \Longleftrightarrow \quad \bar{z}_1 = \sqrt{\frac{c}{2}}.$$

Note that $c > 2$ implies that $z_1^0 < \bar{z}_1$.

Equilibrium loan rates under two-bank lending. Given the expected loan rates R_i and R_j, in the contracting stage, we consider two situations. First, consider the case when both the participation and no-collusion constraints of both loan officers bind, i.e.,

$$\begin{aligned}
&\pi(m_i,m_j)R_i s_i = \frac{1}{2}cm_i^2 \quad \text{and} \quad \pi(m_i,m_j)R_j s_j = \frac{1}{2}cm_j^2,\\
&R_i s_i = R_i + R_j - 2z = R_j s_j,
\end{aligned}$$

which together yield

$$cm_i^2 = 2\pi(m_i,m_j)(R_i+R_j-2z) = cm_j^2.$$

The above system has one pair of non-zero symmetric solutions, which is given by:

$$m_i(z,R_i,R_j) = m_j(z,R_i,R_j) = \frac{4(R_i+R_j-2z)}{c+2(R_i+R_j-2z)},$$

$$\text{and hence,}\ \ \pi(z,R_i,R_j) = \frac{8c(R_i+R_j-2z)}{(c+2(R_i+R_j-2z))^2}.$$

The zero-profit conditions of banks i and j are respectively given by:

$$
\begin{aligned}
B_i(z, R_i, R_j) &= \pi(z, R_i, R_j)R_i(1 - s_i(z, R_i, R_j)) - 1 = 0, \\
B_j(z, R_i, R_j) &= \pi(z, R_i, R_j)R_j(1 - s_j(z, R_{i,R_j})) - 1 = 0.
\end{aligned}
$$

The above system has two pairs of symmetric solutions which are given by:

$$
\begin{aligned}
R_i(z) = R_j(z) &= \frac{2z(2+3c) - c\left(1 + \sqrt{4z(z-1)-c}\right)}{4(1+c)}, \\
R_i(z) = R_j(z) &= \frac{2z(2+3c) - c\left(1 - \sqrt{4z(z-1)-c}\right)}{4(1+c)}.
\end{aligned}
$$

We consider the larger loan rate. Call it $R_2(z)$. The equilibrium individual monitoring efforts and the aggregate monitoring intensity are given by:

$$
\begin{aligned}
m_i(z) = m_j(z) = m_2(z) &= \frac{2}{2z - \sqrt{4z(z-1)-c}}, \\
\pi(z) &= \frac{4\left(c + 2z(2+c) + c\sqrt{4z(z-1)-c}\right)}{(c+4z)^2}, \\
s_i(z) = s_j(z) = s_2(z) &= s_i(z, R_i(z), R_j(z)).
\end{aligned}
$$

Note that $m_2(z) \leq 1$ implies that $z \geq 1 + (c/4) \equiv z_2^0$. For the above solutions to be real, we also require that $4z(z-1) - c \geq 0$, i.e., $z \geq \frac{1}{2} + \frac{1}{2}\sqrt{1+c}$. Because $\frac{1}{2} + \frac{1}{2}\sqrt{1+c} < z_2^0$ for any $c > 2$, the equilibrium values of loan rates, individual monitoring effort, aggregate monitoring intensity and share for each loan officer are real.

Next, consider the case when the participation constraints of both loan officers bind, but both the no-collusion constraints are slack. Substituting for s_i from binding (PC_i^0), and s_j from binding (PC_j^0), the objective functions of the banks reduce to:

$$
\begin{aligned}
B_i &= \pi(m_i, m_j)R_i - \frac{1}{2}cm_i^2 - 1, \\
B_j &= \pi(m_i, m_j)R_j - \frac{1}{2}cm_j^2 - 1.
\end{aligned}
$$

The first-order conditions of the banks' maximization problems yield

$$
cm_i = R_i(1 - m_j) \quad \text{and} \quad cm_j = R_j(1 - m_i),
$$

from which we obtain

$$
\begin{aligned}
m_i(R_i, R_j) &= \frac{R_i(c - R_j)}{c^2 - R_iR_j} \quad \text{and} \quad m_j(R_i, R_j) = \frac{R_j(c - R_i)}{c^2 - R_iR_j}; \\
\pi(R_i, R_j) &= \frac{R_i^2R_j^2 + c^2(c(R_i + R_j) - 3R_iR_j)}{(c^2 - R_iR_j)^2}.
\end{aligned}
$$

The equilibrium values of R_i and R_j are determined from the zero-profit conditions of the banks, which are given by:

$$B_i(R_i, R_j) = \pi(R_i, R_j)R_i - \frac{1}{2}c(m_i(R_i, R_j))^2 - 1 = 0,$$
$$B_j(R_i, R_j) = \pi(R_i, R_j)R_j - \frac{1}{2}c(m_j(R_i, R_j))^2 - 1 = 0.$$

The above system has (a) two pairs of asymmetric and real solutions, (b) one pair of symmetric, but complex solutions, and (c) two pairs of asymmetric, but complex solutions.[27] The pairs of real solutions are given by:

$$\left(\bar{R}_i = \frac{c\left(4(1+c) - \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2},\ \bar{R}_j = \frac{c\left(4(1+c) + \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}\right),$$
$$\left(\bar{R}_i = \frac{c\left(4(1+c) + \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2},\ \bar{R}_j = \frac{c\left(4(1+c) - \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}\right).$$

For the above solutions to be real, we require that $4+2c-c^2 \geq 0$ which holds for $c \leq 1+\sqrt{5}$. Without any loss of generality, we consider only the first set as the equilibrium values of R_i and R_j, which yield

$$\bar{m}_i = \frac{2c - \sqrt{2c(4+2c-c^2)}}{c(2+c)},\ \bar{m}_j = \frac{2c + \sqrt{2c(4+2c-c^2)}}{c(2+c)} \quad \text{and} \quad \bar{\pi} = \pi(\bar{R}_i, \bar{R}_j) = \frac{2}{c},$$
$$\bar{s}_i = \frac{4 - \sqrt{2c(4+2c-c^2)}}{4(2+c)} \quad \text{and} \quad \bar{s}_j = \frac{4 + \sqrt{2c(4+2c-c^2)}}{4(2+c)}.$$

The threshold value of firm quality, $\bar{z}_2$ is determined as follows. Bank i offers a smaller share of repayment to its loan officer, and charges a smaller loan rate to the firm. Hence, if (NC_i^0) holds, then (NC_j^0) must also hold. Note that (NC_i^0) is slack at $(\bar{R}_i, \bar{R}_j, \bar{s}_i)$, i.e.,

$$\bar{R}_i\bar{s}_i > \bar{R}_i + \bar{R}_j - 2z \quad \Longleftrightarrow \quad z > \frac{c(12+12c+c^2) + 2c\sqrt{2c(4+2c-c^2)}}{4(2+c)^2} \equiv \bar{z}_2.$$

It is easy to show $z_2^0 < \bar{z}_2$.

Comparison of the two lending modes. For any $c > 2$, we have $\bar{z}_1 < z_2^0$, and hence, $\bar{z}_1 < \bar{z}_2$, i.e., two-bank lending (under endogenous loan rates) exacerbates collusive threats. Because $\bar{z}_1 < z_2^0$, the relevant values of firm quality are $z \geq z_2^0$. We first compare the equilibrium loan rates under the two financing modes. Note that $R_1(z_2^0) = \bar{R}_1 = \sqrt{2c}$. Next,

$$R_2(z_2^0) = \frac{2+c(3+c)}{2(1+c)}.$$

For any $c > 2$ we have $R_2(z_2^0) > R_1(z_2^0)$. On the other hand, for any $c > 2$ we have $\bar{R}_j < \bar{R}_1$, and hence, $\bar{R}_i < \bar{R}_j < \bar{R}_1$. Because $R_2(z)$ is strictly decreasing on $[z_2^0, \bar{z}_2]$, there is a unique $z_R \in (z_2^0, \bar{z}_2)$ such that the loan rates charged are higher under two-bank lending if and only if $z \leq z_R$.

[27] It is easy to see that at the symmetric solution $R_i = R_j$, both B_i and B_j are strictly positive, and hence, the symmetric solution is not real. Solving the system of equations, $B_i(R_i, R_j) = B_j(R_i, R_j) = 0$ analytically is not an easy task. Hence, we used Mathematica to solve them, which gives two complex and two real asymmetric solutions.

The comparison of the individual monitoring efforts, the aggregate monitoring intensity, and shares of repayment under the two lending modes are similar as above given that

$$m_2(z_2^0) = \pi(z_2^0) = 1 \quad \text{and} \quad s_2(z_2^0) = \frac{c}{2+c}.$$

Thus, we prove the existence of unique thresholds z_m, z_π and z_s.

B Further extensions

B.1 Constant loan size

We analyze a situation where the amount lent to the firms remains constant at \$1 across the two lending modes. Under two-bank lending, each bank invests \$0.50, and gets back $r/2$ in the event of success. Because the amount lent does not change, the stake of collusion also remains constant at $pr-z$. Thus, the no-collusion constraint of loan officer i is given by:

$$\frac{1}{2}prs_i \geq pr - z \iff prs_i \geq 2(pr-z). \tag{NC$_i$}$$

The participation constraint of loan officer i is given by:

$$\frac{1}{2}\pi(m_i, m_j)prs_i - \frac{1}{2}cm_i^2 \geq 0. \tag{PC$'_i$}$$

Thus, bank i maximizes

$$\frac{1}{2}\pi(m_i, m_j)pr(1-s_i) - \frac{1}{2}$$

subject to (PC$'_i$), (NC$_i$) and (F$_i$). We omit the detailed analysis as the derivation of the optimal contract is similar to the two-bank lending case when the aggregate lending is \$2. The symmetric equilibrium monitoring effort is given by:

$$\hat{m}_2(z, r) = \begin{cases} \frac{4(pr-z)}{c+2(pr-z)} & \text{for } z \in (\hat{z}_2^0, \hat{z}_2], \\ \frac{pr}{2c+pr} & \text{for } z \in (\hat{z}_2, pr], \end{cases}$$

where $\hat{z}_2^0$ solves $\hat{B}_2(z) = 0$, $\hat{B}_2(z)$ being the utility of each bank in the symmetric equilibrium, and $\hat{z}_2$ solves $\frac{4(pr-z)}{c+2(pr-z)} = \frac{pr}{2c+pr}$, and is given by:

$$\hat{z}_2 = \frac{pr(7c+2pr)}{2(4c+pr)}.$$

In the following proposition we compare $m_1(z, r)$ and $\hat{m}_2(z, r)$.

Proposition 12 *Two bank lending implies larger collusive threats, i.e., the no-collusion constraint of each loan officer binds over a larger range of firm quality ($\hat{z}_2 > \bar{z}_1$). Hence, there is inefficient over-monitoring for a larger range of firm quality z. Moreover, for a given loan rate r, if $pr > 4.38$ and $\frac{pr}{2}\left(pr - \sqrt{pr(pr-4)}\right) < c$, then there are unique threshold values of firm quality, $\hat{z}_m \in (z_2^0, \hat{z}_2)$ such that the equilibrium monitoring effort under two-bank lending is higher relative to single-bank lending, i.e., $\hat{m}_2(z, r) \geq m_1(z, r)$ if and only if $z \leq \hat{z}_m$.*

Proof The proof is very similar to that of Proposition 5. Note first that $\hat{m}_2(z,r) = 1$ at $z = z_1^0$. However, $B_2(z_1^0, r) < 0$. So, for the optimal monitoring effort, we need to guarantee that $B_2(z,r) \geq 0$ which is the case for $z \geq \hat{z}_2^0$ where $\hat{z}_2^0$ is the smaller root of z that solves $B_2(z,r) = 0$. It turns out that

$$\hat{m}_2(\hat{z}_2^0) = \frac{1}{pr - \sqrt{p^2r^2 - pr - c}}.$$

It is easy to show that $\hat{m}_2(\hat{z}_2^0) > pr/c$ for $pr > 4.38$ and $\frac{pr}{2}\left(pr - \sqrt{pr(pr-4)}\right) < c$. On the other hand, $\frac{pr}{2c+pr} < \frac{pr}{c}$. The proof follows immediately. $\square$

Why do we require to the above restrictions on pr and c to obtain the proposition? Note that, under constant loan size, the no-collusion constraints of the two loan officers are unchanged. However, each bank has only half of the revenue, $r/2$ out of which it incentivizes the loan officer. If r is low, the corresponding incentive pays, prs_i and prs_j are low too which are not sufficient to elicit high monitoring effort when the no-collusion constraints bind. In other words, high r (the corresponding c must also be high because $c \geq pr$) is able to induce high monitoring effort when borrower quality is low.

B.2 Non-contractible monitoring effort

The main objective of our paper is to highlight the role of the loan officers' incentives to collude with the borrower under relationship lending. Therefore, we have assumed that monitoring efforts are contractible, and focused on the aforementioned incentive problem in bank-monitor-borrower hierarchies. If monitoring efforts were non-contractible, then each bank would face an additional *effort incentive constraint* (moral hazard). We assume the following.

Assumption 2 $2\sqrt{c} \leq pr \leq c$, *and* $4 < c < 10$.

Single-bank lending: Under single-bank lending, the maximization problem ($\mathscr{M}$) would be subject to an additional incentive constraint which is given by:

$$m = \underset{\hat{m}}{\operatorname{argmax}} \left\{ \hat{m}pry - \frac{1}{2}c\hat{m}^2 \right\} = \frac{prs}{c}. \tag{ICM}$$

Substituting for s, which is given by (ICM), into the payoff functions of the bank and the loan officer, the bank's maximization problem boils down to:

$$\max_m B^{MH}(m) \equiv mpr - cm^2 - 1, \tag{$\mathscr{M}^{MH}$}$$

$$\text{subject to } cm \geq pr - z. \tag{NC^{MH}}$$

It is easy to show that, under (ICM), the participation constraint of the loan officer does not bind, and hence it is omitted from the above maximization problem. As before, there will be two cases—(a) when the no-collusion constraint binds at the optimum, and (b) when it is slack. Without going into much details of algebra, we describe the optimal monitoring effort and share for the loan officer, repspectively.

$$m_1^{MH}(z) = \begin{cases} \frac{pr-z}{c} & \text{for } z \in [\underline{z}_1, \tilde{z}_1], \\ \frac{pr}{2c} & \text{for } z \in (\tilde{z}_1, z_1^{max}], \end{cases}$$

$$s_1^{MH}(z) = \begin{cases} 1 - \frac{z}{pr} & \text{for } z \in [\underline{z}_1, \tilde{z}_1], \\ \frac{1}{2} & \text{for } z \in (\tilde{z}_1, z_1^{max}], \end{cases}$$

where $\underline{z}_1$ and z_1^{max} are such that $B_1^{MH}(z)$, the equilibrium payoff of the bank is non-negative if and only if $z \in [\underline{z}_1, z_1^{max}]$, and $\tilde{z}_1 = pr/2$. Note that $m_1^{MH}(z)$ is strictly decreasing on $[\underline{z}_1, \bar{z}_1]$, and constant on $(\tilde{z}_1, z_1^{max}]$. Thus,

$$\max_z \left\{ m_1^{MH}(z) \right\} = m_1^{MH}(\underline{z}_1) = \frac{pr + \sqrt{p^2r^2 - 4c}}{2c} < \frac{pr}{c} = m_1^*(r),$$

i.e., under moral hazard, there is always under-monitoring relative to the non-delegation level.

Two-bank lending: Under two-bank lending, on the other hand, the effort incentive constraints of loan officers i and j are given by:

$$m_i = \operatorname*{argmax}_{\hat{m}_i} \left\{ \pi(\hat{m}_i, m_j) prs_i - \frac{1}{2} c\hat{m}_i^2 \right\} = \frac{(1-m_j)prs_i}{c}, \tag{ICM$_i$}$$

$$m_j = \operatorname*{argmax}_{\hat{m}_j} \left\{ \pi(m_i, \hat{m}_j) prs_j - \frac{1}{2} c\hat{m}_j^2 \right\} = \frac{(1-m_i)prs_j}{c}. \tag{ICM$_j$}$$

Thus, bank i solves the maximization problem ($\mathscr{M}_i$) subject to (PC$_i$), (ICM$_i$) and (NC$_i$).[28] Bank j solves a similar program. Under (ICM$_i$), the participation constraint of loan officer i always holds, and hence, we disregard this constraint. Substituting for s_i and s_j from the above two effort incentive constraints, the maximization problem of bank i reduces to:

$$\max_{m_i} B_i^{MH}(m_i, m_j) \equiv \pi(m_i, m_j)pr - \frac{cm_im_j}{1-m_j} - cm_i^2 - 1, \tag{$\mathscr{M}_i^{MH}$}$$

$$\text{subject to } \frac{cm_i}{1-m_j} \geq 2(pr - z). \tag{NC$_i^{MH}$}$$

When both the no-collusion constraints bind, the best replies $m_i(m_j)$ and $m_j(m_i)$ are solutions to the following first-order conditions:

$$cm_i = 2(pr - z)(1 - m_j), \tag{BR$'_i$}$$

$$cm_j = 2(pr - z)(1 - m_i). \tag{BR$'_j$}$$

Notice that, unlike the case of contractible monitoring efforts, the best replies are downward sloping, i.e., m_i and m_j are strategic substitutes. If (NC$_i$) binds, then (ICM$_i$) reduces to $cm_i = 2(pr - z)(1 - m_j)$. As m_j increases, the effort incentive constraint of loan officer i becomes slack. Because each loan officer must receive additional rent (non-binding participation constraint), the effort incentive constraint must bind, and hence, m_i decreases. When none of the no-collusion constraints binds at the optimum, the best replies $m_i(m_j)$ and $m_j(m_i)$ are solutions to the following first-order conditions:

$$2cm_i = pr(1 - m_j) - \frac{cm_j}{1-m_j}, \tag{BR$_i$}$$

$$2cm_j = pr(1 - m_i) - \frac{cm_i}{1-m_i}. \tag{BR$_j$}$$

[28] Under both lending modes, the equilibrium monitoring effort is strictly less than 1, and hence, we ignore the feasibility constraints.

It is also the case that the best replies are downward sloping. There is a unique symmetric equilibrium which is described as follows. The symmetric equilibrium monitoring efforts are given by:

$$m_2^{MH}(z) = \begin{cases} \frac{2(pr-z)}{c+2(pr-z)} & \text{for } z \in [\underline{z}_2, \tilde{z}_2], \\ \frac{3c+2pr-\sqrt{9c^2+4cpr}}{2(2c+pr)} & \text{for } z \in (\tilde{z}_2, z_2^{max}]. \end{cases}$$

The equilibrium aggregate monitoring intensity is given by:

$$\pi^{MH}(z) = \begin{cases} \frac{4(pr-z)(c+pr-z)}{(c+2(pr-z))^2} & \text{for } z \in [\underline{z}_2, \tilde{z}_2], \\ \frac{3c^2+6cpr+2p^2r^2-c\sqrt{9c^2+4cpr}}{2(2c+pr)^2} & \text{for } z \in (\tilde{z}_2, z_2^{max}]. \end{cases}$$

Finally the symmetric equilibrium shares of repayment are given by:

$$s_2^{MH}(z) = \begin{cases} 2(1-z/pr) & \text{for } z \in [\underline{z}_2, \tilde{z}_2], \\ \frac{\sqrt{9c^2+4cpr}-3c}{2pr} & \text{for } z \in (\tilde{z}_2, z_2^{max}]. \end{cases}$$

The thresholds $\underline{z}_2$ and z_2^{max} are such that $B_2^{MH}(z)$, the equilibrium payoff of the bank is non-negative if and only if $z \in [\underline{z}_2, z_2^{max}]$, and $\tilde{z}_2$ is such that both the no-collusion constraints bind for for $z \leq \tilde{z}_2$, which is given by:

$$\tilde{z}_2 = \frac{3c+4pr-\sqrt{9c^2+4cpr}}{4}.$$

It is also easy to show that there is always under-monitoring relative to the non-delegation monitoring level.

Comparison of the two lending modes: In the following proposition, we compare the equilibrium monitoring efforts, aggregate monitoring intensity, and repayment shares across the two lending modes.

Proposition 13 *The symmetric equilibrium individual monitoring effort under two-bank lending is always lower than that under single-bank lending, i.e., $m_2^{MH}(z) < m_1^{MH}(z)$ for all $z \in [\underline{z}_2, z_2^{max}]$. There are unique threshold values of firm quality $\tilde{z}_\pi, \tilde{z}_s \in (\underline{z}_2, \tilde{z}_2)$ such that aggregate monitoring intensity is higher and incentives are stronger under two-bank lending, i.e., $\pi^{MH}(z) > m_1^{MH}(z)$ and $s_2^{MH}(z) > s_1^{MH}(z)$ if and only if $z < \tilde{z}_\pi$ and $z < \tilde{z}_s$ respectively.*

Proof In order to compare the individual monitoring efforts under the two lending modes, we first show that $\tilde{z}_1 < \underline{z}_2$. It is not an easy task to find the exact expression for $\underline{z}_2$. Note that the symmetric equilibrium payoff of each bank (when the no-collusion constraints bind) are given by:

$$B_2^{MH}(z) = \frac{4(pr-z)(c+pr-z)(2z-pr)}{(c+2(pr-z))^2} - 1.$$

Hence, $\underline{z}_2$ solves $B_2^{MH}(z) = 0$. An analytical solution is difficult to get. However, we know that $B_2^{MH}(z)$ is strictly increasing in z up to some z_2^+ at which $B_2^{MH}(z)$ is maximized. Note that $B_2^{MH}(\tilde{z}_1) = -1 < 0 = B_2^{MH}(\underline{z}_2)$, and hence, $\tilde{z}_1 < \underline{z}_2$. Thus, as in the case of Proposition 5, the relevant region of comparison of the two lending mode is $[\underline{z}_2, z_2^{max}]$. Given Assumption 2, it is always the case that

$$m_2^{MH}(\tilde{z}_2) = \frac{3c+2pr-\sqrt{9c^2+4cpr}}{2(2c+pr)} < \frac{pr}{2c} = m_1^{MH}(\tilde{z}_2) = m_1^{MH}(\underline{z}_2).$$

Because $m_2^{MH}(z)$ is strictly decreasing on $[\underline{z}_2, \tilde{z}_2]$, the individual monitoring effort reaches its maximum at $z = \underline{z}_2$. We will show that $m_2^{MH}(\underline{z}_2) < m_1^{MH}(\underline{z}_2)$. Suppose on the contrary that $m_2^{MH}(\underline{z}_2) \geq m_1^{MH}(\underline{z}_2)$. Then, there must be a unique $\tilde{z}_m \in (\underline{z}_2, \tilde{z}_2)$ such that $m_2^{MH}(z) = m_1^{MH}(\underline{z}_2) = pr/2c$, which is given by:

$$\tilde{z}_m = pr - \frac{cpr}{2(2c - pr)}.$$

It is easy to see that $B_2^{MH}(\tilde{z}_m) < 0$ under Assumption 2. Therefore, it must be the case that $\tilde{z}_m < \underline{z}_2$, and hence, $m_1^{MH}(\underline{z}_2) > m_2^{MH}(\underline{z}_2)$.

The proof of the second part follows from a similar logic as above. Let $\tilde{z}_\pi$ and $\tilde{z}_s$ are respectively given by:

$$\frac{3c^2 + 6cpr + 2p^2r^2 - c\sqrt{9c^2 + 4cpr}}{2(2c + pr)^2} = \frac{pr}{c} \quad \text{and} \quad 2\left(1 - \frac{z}{pr}\right) = \frac{1}{2}.$$

It is easy to show that, under Assumption 2, $B_2^{MH}(\tilde{z}_\pi) > 0 = B_2^{MH}(\underline{z}_2)$ and $B_2^{MH}(\tilde{z}_s) > 0 = B_2^{MH}(\underline{z}_2)$. Therefore, it must be the case that $\underline{z}_2$ is lower than both $\tilde{z}_\pi$ and $\tilde{z}_s$. On the other hand, it is also true that $\pi^{MH}(\tilde{z}_2) < m_1^{MH}(\tilde{z}_2)$ and $s_2^{MH}(\tilde{z}_2) < s_1^{MH}(\tilde{z}_2)$, and hence, both $\tilde{z}_\pi$ and $\tilde{z}_s$ are lower than $\tilde{z}_2$. □

Two-bank lending implies lower monitoring because of the strategic substitutability of m_i and m_j. When monitoring efforts are contractible, interaction between the participation constraints of the loan officers induces the monitoring efforts to be strategic complements (the race-to-collusion effect). Under non-contractible monitoring effort, such strategic interaction emerges through the effort incentive constraints. However, for low values of z, the aggregate monitoring intensity may be higher under two-bank lending even if the individual monitoring effort is lower. Finally, it is not very surprising that the banks provide stronger incentives for low values of firm quality because the share of repayment is the only instrument that takes care of the double incentive problems (collusion and effort), and for low borrower quality, both incentive problems are strong.

One way to restore strategic complementarity between the individual monitoring efforts is to publicly verifiable signals. In particular, we consider the following model. Let m_i and m_j be the efforts chosen by loan officers i and j respectively, whereas $\hat{m}_i$ and $\hat{m}_j$ are monitoring efforts stipulated in the contracts. Let λ be the exogenously given probability that the banks know the monitoring effort chosen by the loan officers. In case it is detected that $m_i < \hat{m}_i$, then loan officer i pays a penalty

$$\frac{1}{2}\lambda\tau(\hat{m}_i - m_i)^2,$$

where $\tau > 0$ is the penalty rate. We prove the following result:

Proposition 14 *When the no-collusion constraints of both loan officers bind, the monitoring efforts are strategic complements (substitutes) if $\lambda\tau$ is high (low).*

Proof When the no-collusion constraints bind, the optimal shares are given by $s_i = s_j = 2(1 - z/pr) \equiv s^0$. Thus, the maximization problem of bank i is given by:

$$\max_{\hat{m}_i} \; \pi(m_i, m_j)pr(1 - s^0) + \frac{1}{2}\lambda\tau(\hat{m}_i - m_i)^2 - 1,$$

$$\text{subject to } \pi(m_i, m_j)prs^0 - \frac{1}{2}cm_i^2 - \frac{1}{2}\lambda\tau(\hat{m}_i - m_i)^2 \geq 0, \qquad (\text{PC}_i^{MH})$$

$$m_i = \underset{m}{\operatorname{argmax}} \left\{\pi(m, m_j)prs^0 - \frac{1}{2}cm^2 - \frac{1}{2}\lambda\tau(\hat{m}_i - m)^2\right\} = \frac{prs^0(1 - m_j) + \lambda\tau\hat{m}_i}{c + \lambda\tau}. \qquad (\text{ICM}_i')$$

Constraint (ICM'_i) is the effort incentive constraint of loan officer i. Note that when $\lambda\tau = 0$, (ICM'_i) is equivalent to (ICM_i). Solving for m_i and m_j from the effort incentive constraints of the two loan officers, we obtain

$$m_i(\hat{m}_i, \hat{m}_j) = \frac{\lambda\tau(c+\lambda\tau)\hat{m}_i + prs^0(c - prs^0 + \lambda\tau(1-\hat{m}_j))}{(c - prs^0 + \lambda\tau)(c + prs^0 + \lambda\tau)},$$
$$m_j(\hat{m}_i, \hat{m}_j) = \frac{\lambda\tau(c+\lambda\tau)\hat{m}_j + prs^0(c - prs^0 + \lambda\tau(1-\hat{m}_i))}{(c - prs^0 + \lambda\tau)(c + prs^0 + \lambda\tau)}.$$

If $\lambda\tau \to \infty$, then we have $m_i \to \hat{m}_i$ and $m_j \to \hat{m}_j$. Therefore, in the limit the above models reduces to the two-bank lending mode with contractible monitoring efforts. It follows from the binding participation constraints of the loan officers that individual monitoring efforts are strategic complements. $\square$

Strategic complementarity of monitoring efforts implies that for low values of z the symmetric equilibrium monitoring effort under two-bank lending is higher than that under the single-bank lending mode even if monitoring effort is not contractible, a result similar to Proposition 5.

B.3 Lending with many banks

We next extend the model to allow for n banks, where $n \geq 2$, with each bank employing a single loan officer. Each bank invests \$1 in the firm, and hence, the aggregate loan is \$$n$. We continue assuming that r is exogenously given. Let $\mathbf{m} = (m_1, \ldots, m_n)$ be the vector of monitoring efforts chosen by the n loan officers. The aggregate monitoring intensity is thus given by:

$$\pi(\mathbf{m}) = 1 - \prod_{i=1}^{n}(1 - m_i). \tag{1'}$$

The no-collusion constraint for loan officer i is given by:

$$prs_i \geq n(pr - z). \tag{NC$_i^n$}$$

The participation constraint of the loan officer at each bank $i \in \{1, \ldots, n\}$ is given by:

$$\pi(\mathbf{m})prs_i - \frac{1}{2}cm_i^2 \geq 0, \tag{PC$_i^n$}$$

which together with the feasibility constraint (F_i) and the no-collusion constraint (NC_i^n) constitute the feasible set $\mathscr{F}_i^n$ of contracts (m_i, s_i) offered by bank i. The feasible set $\mathscr{F}_i^n$ of bank i depends on the contract offers of all the rival banks. Thus, each bank i solves

$$\max_{(m_i, s_i)\in\mathscr{F}_i^n} B_i(\mathbf{m}, s_i) \equiv \pi(\mathbf{m})pr(1 - s_i) - 1, \tag{$\mathscr{M}_i^n$}$$

We look at the symmetric equilibrium in the choice of monitoring efforts and incentives, i.e. $m_i = m$ and $s_i = s$ for all $i \in \{1, \ldots, n\}$.[29] As before, the participation constraint (PC_i^n) of each loan officer i binds at the optimum. There are two situations. When none of the no-collusion constraints bind, the symmetric equilibrium monitoring efforts are given by the following first-order condition:

$$pr(1 - m)^{n-1} = cm. \tag{$\star$}$$

[29] It is easy to show that there is a unique equilibrium which is symmetric.

The above equilibrium condition yields the monitoring effort which is same as the non-delegated level of monitoring effort. By contrast, when all the no-collusion constraints bind, the symmetric equilibrium share is given by $s = n(1 - z/pr)$, and the equilibrium monitoring effort solves

$$n(pr - z)[1 - (1 - m)^n] = \frac{1}{2}cm^2. \tag{$\star\star$}$$

In the following proposition, we compare equilibrium monitoring effort, incentives and the aggregate monitoring intensity under two distinct lending modes — namely, n-bank lending and n'-bank lending with $n' > n \geq 2$.

Proposition 15 *Let $m(z, n)$, $s(z, n)$ and $\pi(z, n)$ respectively denote the monitoring effort, share of repayment, and the aggregate monitoring intensity in the symmetric equilibrium.*

(a) *For each $n \geq 2$, there are threshold values $z^0(n)$ and $\bar{z}(n)$ of firm quality with $z^0(n) < \bar{z}(n)$ such that the equilibrium outcome involves over-monitoring for all $z \in [z^0(n), \bar{z}(n)]$. Moreover, the threshold $\bar{z}(n)$ is strictly increasing in n with $\lim_{n\to\infty} \bar{z}(n) = pr$;*

(b) *For any two n and n' with $n' > n \geq 2$, there are $z_m, z_s \in (\max\{\bar{z}(n), z^0(n')\}, \bar{z}(n'))$ such that equilibrium monitoring effort is lower under n-bank lending, i.e. $m(z, n) < m(z, n')$ if and only is $z < z_m$, and equilibrium incentives are weaker under n-bank lending, i.e. $s(z, n) < s(z, n')$ if and only is $z < z_s$;*

(c) *For any two n and n' with $n' > n \geq 2$, equilibrium aggregate monitoring intensity is always lower under n-bank lending, i.e. $\pi(z, n) < \pi(z, n')$ for all $z \geq z^0(n')$.*

Proof The proof is similar to those of Propositions 2 and 4, and hence, many details will be omitted. The binding participation constraint of loan officer i defines the optimal share as a function of the vector of efforts, i.e. $s_i(\mathbf{m})$. Thus, the maximization problem of bank i boils down to:

$$\max_{m_i} \ \pi(\mathbf{m})pr - \frac{1}{2}cm_i^2 - 1, \tag{$\mathscr{M}_i^n$}$$

$$\text{subject to} \quad prs_i(\mathbf{m}) \geq n(pr - z). \tag{NC$_i^n$}$$

Let $\mathbf{m}_{-i} = (m_i, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n)$. Note that

$$\pi_i(\mathbf{m}_{-i}) \equiv \frac{\partial \pi(\mathbf{m})}{\partial m_i} = \prod_{j \neq i}^{n} (1 - m_j).$$

In a symmetric equilibrium, given $m_j = m$ for all $j \neq i$, m_i must be the best reply against m. When (NC$_i^n$) does not bind, the best reply of loan officer i solves

$$pr\,\pi_i(\mathbf{m}_{-i}) = cm_i.$$

Thus, using symmetry, the above first-order condition reduces to ($\star$). Denote by $m(n)$ the symmetric equilibrium monitoring effort. We first show that $m'(n) < 0$. Write ($\star$) as $h(m, n) = \gamma(r, c)m$ where $h(m, n) \equiv (1 - m)^{n-1}$ and $\gamma(r, c) = c/pr \geq 1$. Note that $h(m, n)$ is strictly decreasing and convex in m for any $n > 2$, and strictly decreasing in n because $1 - m < 1$. Moreover, $h(m, 2)$ is linearly decreasing in m with slope -1. On the other hand, $\gamma(r, c)m$ is an increasing line passing through the origin with slope greater than 1. Because $m(n)$ is determined by the intersection of $h(m,n)$ and $\gamma(r,c)m$, the intersection point moves to the left as n grows, which implies that $m'(n) < 0$. Moreover, because $\gamma(r,c) \geq 1$, the intersection between $h(m,2)$ and $\gamma(r)m$ is always to the left of a point where $m = 1/2$ which implies that $m(n) < 1/2$ for all $n > 2$. The aggregate monitoring intensity is given by $\pi(n) = 1-(1-m(n))^n = 1-\gamma(r,c)m(n)(1-m(n))$. The last equality follows from ($\star$). Thus,

$$\pi'(n) = -\gamma(r,c)(1-2m(n))m'(n) > 0$$

because $m(n) \leq 1/2$ for all $n \geq 2$, and $m'(n) < 0$. The equilibrium incentives are determined from the binding participation constraint of each loan officer, i.e.

$$s(n) = \frac{\gamma(r,c)}{2} \cdot \frac{[m(n)]^2}{\pi(n)}$$

Because $\pi'(n) > 0$ and $m'(n) < 0$, we have $s'(n) < 0$. Also, at the above optimum, the no-collusion constraint must hold with strict inequality, i.e.

$$prs(n) > n(pr-z) \quad \Longleftrightarrow \quad z > pr\left(1-\frac{s(n)}{n}\right) \equiv \bar{z}(n).$$

Note that $s(n)/n$ is strictly decreasing in n, and hence, $\bar{z}(n)$ is strictly increasing in n. Because $s(n)$ is bounded above by 1, $\lim_{n\to\infty}[s(n)/n] = 0$ which implies that $\lim_{n\to\infty} \bar{z}(n) = pr$. This completes the proof of part (a).

Next, consider the case when each no-collusion constraint binds. In the symmetric equilibrium, the shares are given by:

$$s(z,n) = n\left(1-\frac{z}{pr}\right),$$

which is strictly increasing in n for each z. Using the fact that $\pi = 1-(1-m)^n$, it follows from the binding binding participation constraints that

$$[1-(1-m)^n]prs(\lambda,n) = \frac{1}{2}cm^2 \quad \Longleftrightarrow \quad \underbrace{2[1-(1-m)^n]n\left(1-\frac{z}{pr}\right)}_{H(m,n)} = \underbrace{\gamma(r,c)m^2}_{G(m)}.$$

The solution to the above equation gives the equilibrium monitoring effort. Note that $G(m)$ is strictly increasing and strictly convex in m with $G(0) = 0$ and $G(1) = \gamma(r,c)$. On the other hand, $H(m,n)$ is strictly increasing and strictly concave in m with $H(0,n) = 0$ for all $n > 0$ and $H(1,n) = 2n(1-z/pr)$. Note that, whenever $m = \pi = 1$, from the binding participation constraint we have $s = c/2pr$. At this optimal share, the no-collusion constraint does not hold, i.e. $c/2pr < n(1-z/pr)$ if and only if $z < pr(1-(c/2npr)) \equiv z^0(n)$. Thus, bank lending is feasible only if $z \geq z^0(n)$ which is equivalent to $H(1,n) = 2n(1-z/pr) \leq \gamma(r,c) = G(1)$. Thus, the intersection of $H(m,n)$ and $G(m)$ is unique for each n, which gives the symmetric equilibrium monitoring effort $m(z,n)$ which is unique for each n. Moreover, $H(m,n)$ is strictly increasing in n which implies that $m(z,n)$ is strictly increasing for each $z \in [z^0(n), \bar{z}(n)]$. The aggregate monitoring intensity is given by $\pi(z,n) = 1-[1-m(z,n)]^n$. Because $m(z,n)$ is strictly increasing in n, we have that $\pi(z,n)$ is strictly increasing in n for all $z \in [z^0(n), \bar{z}(n)]$. Part (b) of the proposition follows from the fact that both $m(z,n)$ and $s(z,n)$ are strictly increasing in n, and both $m(n)$ and $s(n)$ are strictly decreasing in n. In the proof of Proposition 6 we have shown that

$\bar{z}_1 \equiv \bar{z}(1) < z_2^0 \equiv z^0(2)$. But for any n and n' with $n' > n \geq 2$, it is not easy to show that $\bar{z}(n) < z^0(n')$. But the result holds irrespective of whether the last inequality is satisfied or not. Therefore, we have taken $\max\{\bar{z}(n), z^0(n')\}$ as the infimum of both z_m and z_s. On the other hand, the fact that both $\pi(z,n)$ and $\pi(n)$ are strictly increasing in n implies part (c). $\square$

The above proposition is a generalization of Propositions 5 and 6 when the number of banks exceeds 2. Recall that over-monitoring emerges when the no-collusion constraint of each loan officer binds. As the number of bank grows, (NC_i^n) of each loan officer becomes more stringent, i.e., collusive threats become larger. Therefore, with larger n, over-monitoring becomes more likely. Although we consider a finite number of banks in a lending coalition, the collusion incentive problem becomes the only determining factor of the optimal loan officer contracts as the number of banks grows very large. In other words, at the limit, the non-delegation outcome can no more be implemented. The second part of the proposition is a ramification of the interaction between the free-riding and race-to-collusion effects. When none of the no-collusion constraints binds, more banks amplifies the free-riding effect, and hence, individual monitoring effort is lower and incentives are weaker as the number of banks grow. By contrast, when the no-collusion constraints bind, more banks exacerbates the race-to-collusion effect, and the aforementioned result flips.

The effect of an increase in the number of banks on the aggregate monitoring intensity is two-fold. Recall that, at a common effort level $m_i = m$ for all i, the aggregate monitoring intensity is given by $\pi = 1-(1-m)^n$. Because of the public good characteristics of the aggregate monitoring intensity function, a growing number of banks has a direct effect on the aggregate monitoring intensity. Because $m \leq 1$, a higher n decreases $(1-m)^n$, and hence, π increases. An increased number of banks also has an indirect effect on the aggregate monitoring intensity which works through a change in m. When the no-collusion constraints bind, individual monitoring effort is strictly increasing in n, and hence, the overall monitoring also increases with the number of banks via this indirect effect. As both these effects point in the same direction, a higher n implies greater aggregate monitoring when the no-collusion constraints bind. By contrast, when the no-collusion constraints are non-binding, monitoring effort m is a decreasing function of n, and hence, growing number of banks has an adverse effect on the aggregate monitoring intensity. Thus, the aggregate effect of the number of banks on overall monitoring depends on which of the two countervailing effects is stronger. When one compares the equilibrium aggregate monitoring intensities under single- and two-bank lending (cf. Proposition 6), the direct effect is not always dominant, and hence, the aggregate monitoring intensity may be lower under two-bank lending for high values of firm quality. It turns out that, for $n \geq 2$, the direct effect of growing n dominates the indirect one. As a consequence, higher n boosts the aggregate monitoring up even when there is no collusion incentive problem.

References

Aoki, Masahiko (1994), “Monitoring Characteristics of the Main Bank Sysytem: An Analytical and Developmental View.” In *The Japanese Main Bank System: Its Relevance for Developing and Transforming Economies* (Masahiko Aoki and Hugh Patrick, eds.), Oxford University Press.

Berger, Allen and Gregory Udell (2002), “Small Business Credit Availability and Relationship Lending: The Importance of Bank Organizational Structure.” *The Economic Journal*, 112, F32–F53.

Bolton, Patrick and David Scharfstein (1996), "Optimal Debt Structure and the Number of Creditors." *The Journal of Political Economy*, 104, 1–25.

Boot, Arnoud and Anjan Thakor (2000), "Can Relationship Banking Survive Competition?" *The Journal of Finance*, 55, 679–713.

Carletti, Elena (2004), "The Structure of Bank Relationships, Endogenous Monitoring, and Loan Rates." *Journal of Financial Intermediation*, 13, 58–86.

Carletti, Elena, Vittoria Cerasi, and Sonja Daltung (2007), "Multiple-Bank Lending: Diversification and Free-riding in Monitoring." *Journal of Financial Intermediation*, 16, 425–451.

Cerqueiro, Geraldo, Steven Ongena, and Kasper Roszbach (2016), "Collateralization, Bank Loan Rates, and Monitoring." *The Journal of Finance*, 71, 1295–1322.

Dam, Kaniṣka and Alejandro Robinson-Cortés (2020), "Executive Compensation and Competitive Pressure in the Product Market: How Does Firm Entry Shape Managerial Incentives?" Forthcoming, *Mathematical Social Sciences*.

Detragiache, Enrica, Paolo Garella, and Luigi Guiso (2000), "Multiple vs. Single Banking Relationships: Theory and Evidence." *The Journal of Finance*, 55, 1133–1161.

Dewatripont, Mathias and Eric Maskin (1995), "Credit and Efficiency in Centralized and Decentralized Economies." *The Review of Economic Studies*, 62, 541–555.

Diamond, Douglas (1984), "Financial Intermediation and Delegated Monitoring." *The Review of Economic Studies*, 51, 393–414.

Farinha, Luísa and João Santos (2002), "Switching from Single to Multiple Bank Lending Relationships: Determinants and Implications." *Journal of Financial Intermediation*, 11, 124–151.

Gangopadhyay, Shubhashis and Bappaditya Mukhopadhyay (2002), "Multiple Bank Lending and Seniority in Claims." *Journal of Economics and Business*, 54, 7–30.

Golan, Limor, Christine Parlour, and Uday Rajan (2015), "Competition, Managerial Slack, and Corporate Governance." *The Review of Corporate Finance Studies*, 54, 43–68.

Gustafson, Matthew, Ivan Ivanov, and Ralf Meisenzahl (2020), "Bank Monitoring: Evidence from Syndicated Loans." Forthcoming *Journal of Financial Economics*.

Hertzberg, Andrew, José Maria Liberti, and Daniel Paravisini (2010), "Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation." *The Journal of Finance*, 65, 795–828.

Hölmstrom, Bengt and Jean Tirole (1997), "Financial Intermediation, Loanable Funds, and the Real Sector." *The Quarterly Journal of Economics*, 112, 663–691.

Kahn, Charles and Dilip Mookherjee (1998), "Competition and Incentives with Nonexclusive Contracts." *The Rand Journal of Economics*, 29, 443–465.

Khalil, Fahad, David Martimort, and Bruno Parigi (2007), "Monitoring a Common Agent: Implications for Financial Contracting." *Journal of Economic Theory*, 135, 35–67.

Kofman, Fred and Jacques Lawarée (1993), "Collusion in Hierarchical Agency." *Econometrica*, 61, 629–656.

Laffont, Jean-Jacques and David Martimort (2000), "Mechanism Design with Collusion and Correlation." *Econometrica*, 68, 309–342.

Laffont, Jean-Jacques and Jean-Charles Rochet (1997), "Collusion in Organizations." *Scandinavian Journal of Economics*, 99, 485–495.

Liberti, José Maria and Atif Mian (2009), "Estimating the Effect of Hierarchies on Information Use." *The Review of Financial Studies*, 22, 4057–4090.

Macho-Stadler, Inés and David Pérez-Castrillo (1993), "Moral Hazard with Several Agents: The Gains From Cooperation." *International Journal of Industrial Organization*, 11, 73–100.

Martin, Stephen (1993), "Endogenous Firm Efficiency in a Cournot Principal-Agent Model." *Journal of Economic Theory*, 4, 445–450.

Mookherjee, Dilip and Ivan Png (1995), "Corruptible Law Enforcers: How Should They be Compensated?" *The Economic Journal*, 105, 145–159.

Ongena, Steven and David Smith (2000), "What Determines the Number of Bank Relationships? Cross-Country Evidence." *Journal of Financial Intermediation*, 9, 26–56.

Padilla, Jorge and Marco Pagano (1997), "Endogenous Communication among Lenders and Entrepreneurial Incentives." *The Review of Financial Studies*, 10, 205–236.

Parlour, Christine and Uday Rajan (2001), "Competition in Loan Contracts." *The American Economic Review*, 91, 187–226.

Ramakrishnan, Ram and Anjan Thakor (1991), "Cooperation versus Competition in Agency." *Journal of Law, Economics and Organization*, 7, 248–283.

Repullo, Rafael and Javier Suárez (2000), "Entrepreneurial Moral Hazard and Bank Monitoring: A Model of the Credit Channel." *European Economic Review*, 44, 1931–1950.

Stein, Jeremy (2002), "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms." *The Journal of Finance*, 57, 1891–1921.

Sufi, Amir (2007), "Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans." *The Journal of Finance*, 62, 629–668.

Tirole, Jean (1986), "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations." *Journal of Law, Economics and Organization*, 2, 181–214.

Tressel, Thierry and Thierry Verdier (2011), "Financial Globalization and the Governance of Domestic Financial Intermediaries." *Journal of the European Economic Association*, 9, 130–175.

Uchida, Hirofumi, Gregory Udell, and Nobuyoshi Yamori (2012), "Loan Officers and Relationship Lending to SMEs." *Journal of Financial Intermediation*, 21, 97–122.

von Thadden, Ernst-Ludwig (1992), “The Commitment of Finance, Duplicated Monitoring and the Investment Horizon.” CEPR discussion paper No. 27.

Winton, Andrew (1995), “Costly State Verification and Multiple Investors: The Role of Seniority.” *The Review of Financial Studies*, 8, 91–123.