Race to collusion: Monitoring and incentives under multiple-bank lending*

Kaniṣka Dam† Prabal Roy Chowdhury‡

Abstract

We examine interactions between multiple bank-loan officer-borrower hierarchies. Possibility of collusion between an informationally opaque borrower and loan officers in charge of monitoring shapes incentives for the loan officers. When 'borrower quality is low', collusive threats induce over-monitoring in a collusion-free equilibrium, whereas for high borrower quality, monitoring is at its non-delegation level—an outcome akin to vertical integration. Under two-bank lending, delegation contracts may solve the free-riding problem in monitoring, and lead to higher levels of monitoring relative to single-bank lending. This is because collusive threats make monitoring efforts strategic complements because of a novel 'race-to-collusion' effect—a hitherto unexplored effect of multiple-bank lending. Two-bank lending, as opposed to single-bank lending, may also entail under-monitoring relative to the social optimum. We further show that the collusion-proofness principle may not hold under two-bank lending, which gives rise to a different sort of free-riding problem.

JEL Codes: D23, D86, G21, L13.
Keywords: Multiple-bank lending; vertical collusion; monitoring.

1 Introduction

Many organizations share two important characteristics—that they have a hierarchical structure, and that they interact with other organizations at various levels, which gives rise to diverse issues of interest. The hierarchical structure suggests collusion may be a problem, whereas the fact that organizations interact leads to possible contractual externality. In the interest of imposing some structure on the problem, we examine these issues in a specific context, one where there are two banks, both of which lends to a single firm, and employs a loan officer apiece to monitor the borrowing firm.[1]

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, in a scenario where informationally opaque borrowers, i.e., borrowers who require intense monitoring and due diligence, lending relationships with

*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.

†Corresponding author. 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] In Section 7 we briefly explore some other contexts where the framework and insights developed in this paper can be useful.

multiple investors may lead to lower level of monitoring because of the free-riding problem and costly monitoring duplication (e.g. Sufi, 2007).

We seek 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 borrowing firm and multiple lenders in which misbehaving by the firm (borrower moral hazard) can only be mitigated by costly monitoring. The novel feature of our model is that (unlike e.g. Winton, 1995) banks delegate monitoring activity to loan officers. Delegation in turn opens up the possibility of collusion between the borrower and corruptible loan officers.[2] We establish that the likelihood of such vertical collusion generates an hitherto unexplored effect of borrowing from multiple lenders, what we term the *race-to-collusion* effect. This solves the free-riding problem is monitoring, and as a consequence, per-bank monitoring under multiple-bank lending may exceed that under the single-bank lending mode.

We analyze a banking market where a firm borrows from two banks to invest in a project. Diligent behavior (high effort) by the borrowing firm generates a stochastic but verifiable cash flow, whereas if the firm misbehaves (low effort), then the project yields only a non-verifiable private benefit. Non-monitored lending is not feasible because in this case the firm necessarily misbehaves, and the banks fail to break-even. Banks must delegate the task of monitoring 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 "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 they may collude with the borrowing firm, and misreport (e.g. Stein, 2002; Liberti and Mian, 2009). While the monitoring level is observable, and hence, contractible,[3] its outcome, i.e., whether monitoring has been successful in detecting borrower misbehavior is not verifiable. Therefore, delegation gives rise to conflict of interests between each bank and its loan officer (collusion incentive problem), and the optimal contract thus involves each bank designing an incentive scheme for its loan officer which consists of a stipulated monitoring effort and a share of repayment. In case of two-bank lending the optimal contract between one bank and its loan officer depends on that of the other, i.e., contracts are interdependent. Using a parsimonious model involving competing bank-loan officer-borrower hierarchies, 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? Further, what are the welfare properties of two-bank lending?

(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?

Under both lending modes, contracts are critically affected by the presence of collusion possibilities. We first restrict attention to *collusion-free contracts*, i.e., contracts that are so designed that loan officers have no incentive to collude with the borrower. In other words, the contracts adhere to to a *no-collusion constraint* which takes a very simple form—the incentive pay of each loan officer must exceed the stake in collusion. The

[2] There is a plethora of evidence on collusion among loan officers and borrowers. 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. Uchida, Udell, and Yamori (2012) find similar evidence for Japanese banks in relation to SME lending. Goldmann (2010) documents many recent cases of loan officer frauds in the U.S.

[3] In many situations, a bank's internal rules specify tasks for the loan officers which are directly correlated with their monitoring effort. Ono and Uesugi (2009), who analyze a detailed dataset of the Japanese main bank system, use the frequency of document submissions to the borrower's main bank as a proxy for monitoring. 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 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 suggest that delegated bank monitoring many a time is verifiable and measurable.

incentive pay plays a dual role here. An increase in incentive pay not only relaxes the participation constraint of each loan officer, thus allowing the bank to mandate higher levels of monitoring, it also makes collusion less attractive. We first show that, under both lending structures, when borrower quality is low or equivalently, the stake in collusion is high, the no-collusion constraint binds implying a severe incentive problem. Thus, a high share of repayment for the loan officer is required to deter collusion. As a result, there is over-monitoring relative to the non-delegation solution, i.e., the one where the banks could monitor the borrower directly. By contrast, for high borrower quality, the collusion incentive problem becomes insignificant as the no-collusion constraint does not bind, and hence, the non-delegation level of monitoring is implemented.

Our main focus is the comparison of the outcomes of the single- and two-bank lending modes. 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, this result is reversed—two-bank lending induces higher monitoring efforts, as well as stronger incentives.

First consider the case when firm quality is high, so that the no-collusion constraint of neither loan officer binds. Moreover, reports by loan officers are public information, allowing each bank to free-ride on the monitoring done by the other bank. Consequently, 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.

Next, consider the case when borrower quality is low. Interestingly, in this case the monitoring efforts become strategic complements, which we call the race-to-collusion effect. 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 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, we 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 instead of compete with each other. 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 next analyze a situation when under two-bank lending, the banks are not constrained to offer collusion-free contracts. We show that the *collusion-proofness principle* (e.g. Tirole, 1986; Laffont and Rochet, 1997) may fail to hold. In particular, there is an equilibrium in which one bank does not employ a loan officer and free-rides on the information gathered by the other bank, whereas the other bank implements the collusion-free contract. The reason is that for low firm quality, one loan officer exerting high monitoring effort enhances the aggregate probability of detecting borrower misbehavior. This leaves little incentive for the other bank to employ a loan officer, thereby saving on the cost of incentive provision. As a result, only one bank monitors in equilibrium, and serves as an intermediary between the other bank and the borrower.

2 Related literature and our contribution

Our paper is related and contributes to two strands of the literature: multiple-bank lending and bank monitoring, and collusion in hierarchical organizations.

Multiple-bank lending and bank monitoring. The papers closest to ours are by Carletti, Cerasi, and Daltung (2007), Carletti (2004) and Khalil et al. (2007). Carletti et al. (2007) analyze a model with multiple lenders. There is two-sided moral hazard, in the effort exerted by the borrowers, as well as 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.

Both Carletti (2004) and Khalil et al. (2007) analyze the effect of multiple-bank lending on monitoring incentives.[4] In Carletti (2004) the central issue is of *interim* moral hazard, with monitoring aimed at preventing borrower misbehavior, whereas in Khalil et al. (2007) the central problem is of costly state verification (ex-post moral hazard). 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.[5] The present paper however differs from both the aforementioned works in that we explicitly take into account the agency problem emerging from collusive threats in bank-monitor-borrower hierarchies. This induces an interplay between the free-riding effect and collusion incentives for the loan officers. Interestingly, Khalil et al. (2007) also demonstrate that if principals coordinate on monitoring levels, but not on the transfers, then there may be over-monitoring relative to the case when the principals merge. By contrast, our results are driven by the possibility of vertical collusion. Thus, by adding loan officers and collusion possibility to bank-borrower relationships, our paper delivers a key relevant feature that is absent in the extant literature on the effect of multiple-bank lending on bank monitoring, which leads to novel results.

The extant theoretical literature has unearthed diverse channels via which multiple-bank lending can have a positive effect on firm quality – other than its effect on bank monitoring – by limiting the ability of any single lender to holdup (e.g. von Thadden, 1992; Padilla and Pagano, 1997), by ameliorating the soft budget constraint problem by making re-financing of unprofitable projects more complicated (e.g. Dewatripont and Maskin, 1995), by reducing strategic default by making debt renegotiation more difficult (e.g. Bolton and Scharfstein, 1996), etc. While, like this literature, the present paper also finds that multiple-bank lending can improve firm performance in the presence of borrower moral hazard, it differs in several respects. First, the channel via which this operates, namely the race-to-collusion effect, is novel in 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 with the aforementioned literature, where multiple-bank lending decreases the fundamental inefficiency, e.g. re-financing of unprofitable projects

[4] 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).

[5] 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).

(as in Dewatripont and Maskin, 1995).

Collusion in organizations. The analysis of vertical collusion in principal-supervisor-agent hierarchies goes back to Tirole's (1986) seminal paper where a supervisor is defined as an imperfect technology that produces verifiable information about the productivity of the agent (adverse selection). He proves a *collusion-proofness principle* which asserts that there is no loss of generality in restricting attention to collusion-free contracts. The principal can deter collusion by setting a high enough incentive pay for the supervisor, which must exceed the supervisor's stake in collusion. Our model has a similar structure where the incentive pay is the collusion-deterring instrument. It is worth mentioning that Scheepens (1997) applies Tirole's (1986) model to a bank-loan officer-borrower hierarchy under adverse selection [about the borrower], and shows that collusion [between the loan officer and the borrower] can be deterred only if the bank can commit to use a committee who scrutinizes the loan proposals.

Another closely related paper is by Mookherjee and Png (1995), which studies the issue of how to compensate corruptible law enforcers in a single organizational hierarchy, also shows that collusive threats exacerbate the incentives for over-monitoring. This result finds resonance in one of our main results that collusive threats lead to over-monitoring irrespective of the lending structures whenever the borrower is of low-quality. Burlando and Motta (2015) on the other hand consider a single principal-auditor-agent hierarchy under adverse selection in which organizational structure can be contractually determined. Under the possibility of vertical collusion between the auditor and the agent, an efficient agent chooses to stay outside the firm without monitoring (an outsourcing contract), whereas an inefficient agent works within the firm and is monitored more intensely (an insourcing contract). This resembles our finding that a low-quality agent (borrower) is monitored more intensely, whereas a high-quality agent, for whom collusive threats are not significant, is monitored at a lower intensity. However, our framework assumes moral hazard instead of adverse selection.

Our principal contribution is to extend the single hierarchy framework (e.g. Tirole, 1986) to multiple hierarchies. There are two novel implications. First, when principals (banks) can commit to collusion-free contracts, the presence of another competing hierarchy can help mitigate the free-riding problem in monitoring that emerges because of the strategic interaction between the principals. However, this comes at a cost—multiple hierarchies exacerbate incentives for over-monitoring.

Second, when principals (banks) cannot commit to collusion-free contracts, the collusion-proofness principle may fail to hold under competing hierarchies. This feature bears resemblance to single hierarchies where a principal can rely on an external signal that is informative about collusion (e.g. Kofman and Lawarée, 1993; Strausz, 1997; Khalil and Lawarrée, 2006). When collusion can be detected with high probability, it is not in principal's interest to commit to collusion-free contracts as the expected gain of deterring collusion is small. In the present context, the collusion-proofness principle may fail to hold because one principal finds it optimal to free-ride on the information gathered by the other, and save on the cost of incentive provision by not hiring a monitor. Another paper that analyzes the failure of collusion-proofness principle under multiple bank lending is Tressel and Verdier (2011). In their model, each bank lends to one borrower, and banks, having an advantage in monitoring borrowers, may collude with the latter. However, Tressel and Verdier (2011) do not analyze the situation in which many banks lend to a single firm, and moreover, their mechanism is different from ours—following the opening up of the economy, foreign capital flows via FDI exacerbates collusion incentive problem of banks by making domestic banks' capital scarcer relative to uninformed capital, thereby increasing the relative cost of bank monitoring. As a result, a fraction of domestic banks collude with domestic firms.

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 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 borrower 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.[6] 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, any investment must be externally funded. Each bank invests \$1 in loan at an exogenously given loan rate $r \geq 1$.[7] Banks incur a 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. Bank monitoring helps control 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, 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.[8,9]

Banks however do not possess any monitoring technology, and hence, must delegate monitoring to loan officers.[10] The [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 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

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

[6] 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 $p_H = p > 0$ and no private benefits, and a "bad project" that yields yI with probability $p_L = 0$ and a strictly positive private benefit v.

[7] In Dam and Roy Chowdhury (2020), we endogenize loan rates allowing for the possibility that the banks charge different loan rates.

[8] 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 as 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 $\pi = 1$.

[9] Empirical measures of monitoring such as intervals of collateral reviews (Cerqueiro et al., 2016), number of times of loan invigilation (Gustafson et al., 2020) also suggest that bank monitoring is often an activity that is performed in the interim stage of a lending relationship. However, all our results qualitatively remain the same in the context of costly state verification in which monitoring is of ex-post nature.

[10] The assumption that banks do not own a monitoring technology holds for example if 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.

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., $C(1)$.[11] We normalize the outside option of the loan officers to 0.

We assume that monitoring *levels* or efforts m_i and m_j can be verified by the banks, and hence, can be contracted upon, whereas 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. Non-observability of the monitoring outcome opens up the possibility of collusion between the concerned loan officer and the firm. When a loan officer colludes with the borrowing firm, he lets the borrower shirk (and enjoy the private benefit) in exchange for a bribe, and reports to his employer that he has not learnt anything about the actions of the firm. This is a 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 by the borrower. 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.[12] 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 focus of our paper is to study the strategic effects of incentive contracts on bank monitoring under multiple-bank lending, and hence, we first analyze collusion-free contracts. In Section 6, we examine whether there is any loss of generality if we restrict attention only to collusion-free contracts.[13]

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 loan officer(s) decide whether to collude, which of course is feasible only if the concerned loan officer has been successful in monitoring. We assume that the outcome of the monitoring effort is private information, i.e., one loan officer does not observe whether the other one has succeeded in detecting firm behavior. In case at least one loan officer is successful and decides not to collude, the borrower is diligent. Otherwise, the borrower misbehaves. There are the following three potential scenarios:

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

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

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

[12] 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).

[13] In a longer version of the paper, Dam and Roy Chowdhury (2020), we analyze several extensions of our model.

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.

At date 5, the cash flow from the project is realized and the agreed upon payments are made.

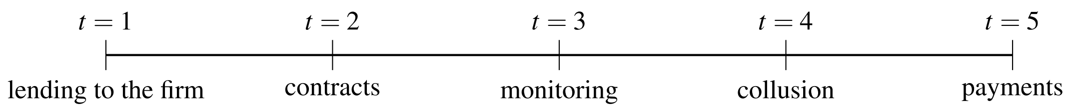


Figure 1: *The timing of events.*

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 optimal monitoring if the bank had access to the same monitoring technology $C(\cdot)$, and had chosen to directly monitor the borrowing firm. The optimal monitoring level solves

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

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 monitors 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 delegation, the bank does not face any incentive problem arising from collusion 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. 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. Tirole, 1986) which abstracts from the minutiae of any particular bargaining protocol involved in the collusion process, and instead focuses on deriving a no-collusion constraint which 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 low enough in case there is collusion, so that one cannot sustain a positive amount of bribe, whatever may be the bargaining protocol.

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:

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

Thus, collusion is not feasible if and only if

$$F(1) + M(1) \geq F(0) + M(0) \quad \Longleftrightarrow \quad prs \geq pr - z. \tag{2}$$

The *no-collusion constraint* (2) takes a simple form—the (expected) incentive pay of the loan officer, prs must exceed the stake in 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 contract*.

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

$$mprs - \frac{1}{2}cm^2 \geq 0, \tag{3}$$

and (ii) the *feasibility constraint*

$$(m, s) \in [0, 1] \times [0, 1]. \tag{4}$$

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 (2), (3) and (4). The bank thus solves the following maximization problem:

$$\max_{(m,s)\in\mathscr{F}} B(m, s) \equiv mpr(1-s) - 1. \tag{5}$$

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 $z \leq z_1^{min}$.*

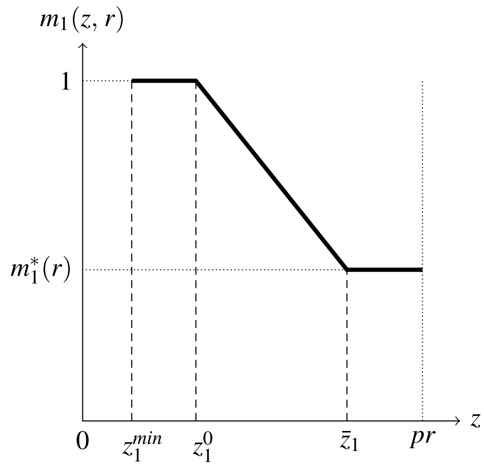
(a) *The optimal monitoring intensity is given by:*

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

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 $m_1^*(r)$. On the other hand, the monitoring level is at the non-delegated level whenever firm quality is high, i.e., $z > \bar{z}_1$.

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

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



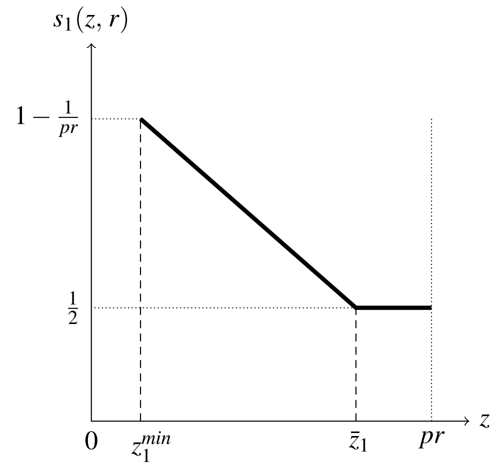


Figure 2: *Equilibrium monitoring and share of the loan officer as functions of firm quality. For $z < z_1^{min}$, bank-lending is not feasible. For $z \in [z_1^{min}, \bar{z}_1]$, the no-collusion constraint binds, and there is over-monitoring relative to the non-delegated level $m_1^*(r)$. For $z >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

The results in Proposition 2 are depicted in Figure 2. 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, (2) as well as (3) 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 in 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 in collusion is high. In this case, the no-collusion constraint binds. 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 a 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 over-monitoring. When z is lowered further, due to very high stake in collusion, the agency problem is so severe that the loan officer must be given incentives in the form of additional rent, i.e., (3) does not bind anymore (the loan officer earns *efficiency wage*), and the monitoring level is optimally set at its highest possible level, 1.

It is also worth noting that, under single-bank lending, the *collusion-proofness principle* (e.g. Tirole, 1986) 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.

Welfare. We compare the equilibrium monitoring with the socially efficient level under single-bank lending. At any monitoring effort m, [expected] welfare is defined as the sum of the expected borrower utility and the expected surplus of the bank-loan officer relationship, i.e.,

$$W_1(m) \equiv mpy + (1-m)v - \frac{1}{2}cm^2 - 1.$$

The socially optimal level of monitoring is the one that maximizes the above expression.

Proposition 3 *The socially optimal level of monitoring under single-bank lending is given by:*

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

Thus the equilibrium under single-bank lending with delegation involves over-monitoring relative to the socially efficient level, i.e., $m_1^{**}(z) \leq m_1(z, r)$ for all $z \in [z_1^{min}, pr]$.

The proof of the above proposition is trivial as $m_1^{**}(z)$ is a strictly increasing function of z which is maximized at $z = pr$, i.e., $m_1^{**}(pr) = pr/c = \min m_1(z, r)$. Therefore, the equilibrium under the single-bank lending mode entails over-monitoring relative to the socially efficient level of monitoring.

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 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

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

The first-order conditions of the maximization problems of banks i and j (similar to the above) yield the following best reply functions:

$$cm_i = pr(1-m_j), \tag{6}$$

$$cm_j = pr(1-m_i). \tag{7}$$

These reaction functions have a unique and symmetric solution $m_2^*(r)$.

Proposition 4 *When two banks lend to the borrower, and they can monitor her directly, the optimal monitoring effort is given by:*

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

and aggregate monitoring intensity is given by $\pi^*(r) = \pi(m_2^*(r), m_2^*(r)) = 1 - [1 - m_2^*(r)]^2$.

As in Proposition 1, the 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:

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

Thus, collusion is not feasible if and only if

$$F(1) + M_i(1) \geq F(0) + M_i(0) \quad \Longleftrightarrow \quad prs_i \geq 2(pr - z). \tag{8}$$

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

$$prs_j \geq 2(pr - z). \tag{9}$$

We then 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 different from (8) and (9). Let 0 represent collusion 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. In case one loan officer decides not to collude, there is no misbehavior on behalf of the borrower and the loan is paid. The payoffs in this case are given by:

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

Thus, collusion is not feasible if and only if

$$F(1) + M_i(1) + M_j(1) \geq F(0) + M_i(0) + M_j(0) \quad \Longleftrightarrow \quad pr(s_i + s_i) \geq 2(pr - z). \tag{10}$$

The no-collusion constraint (10) asserts that the aggregate (expected) incentive pay of the two loan officers must exceed their stake in collusion under two-bank lending, $2(pr - z)$. We impose only (8) and (9) in order to ensure no-collusion because they together imply (10).

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

$$\pi(m_i, m_j) prs_i - \frac{1}{2} c m_i^2 \geq 0, \tag{11}$$

and the *feasibility constraint*

$$(m_i, s_i) \in [0, 1] \times [0, 1]. \tag{12}$$

We denote by $\mathscr{F}_i$ the set of *feasible contracts* for bank i, i.e., the set of contracts that satisfies (8), (11) and (12). We define $\mathscr{F}_j$ for bank j in an analogous manner. It is worth noting that the feasible set of bank i, $\mathscr{F}_i$ 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:

$$\max_{(m_i, s_i) \in \mathscr{F}_i} B_i(m_i, m_j, s_i) \equiv \pi(m_i, m_j) pr(1 - s_i) - 1. \tag{13}$$

We can define the optimization problem of bank j in a similar fashion.

We analyze an equilibrium in which the two no-collusion constraints either bind simultaneously or neither of them binds.[14] Let $m_i(m_j)$ and $m_j(m_i)$ denote the best reply functions of banks i and j, respectively. Lemma 1 derives properties of the best reply functions which will be critical to whether monitoring level under single-bank lending is lower relative to that under two-bank lending or not (see the discussion following Proposition 7).

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 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 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 (6) and (7). Thus, the fact that $\pi_{ij}(m_i, m_j) < 0$ ensures that monitoring efforts are strategic substitutes.

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 (8) (resp. (9)) to obtain the best replies $m_i(m_j)$ and $m_j(m_i)$:

$$cm_i^2 = 4(pr - z)\pi(m_i, m_j), \tag{14}$$

$$cm_j^2 = 4(pr - z)\pi(m_i, m_j). \tag{15}$$

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. 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 (8), 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.[15]

The optimal monitoring and share of repayment for each loan officer are characterized in Proposition 5 below. As in the case of single-bank lending, collusive threats exacerbate incentives for over-monitoring under two-bank lending as well, and hence, banks are required to offer stronger incentives to deter collusion.

Proposition 5 *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*

[14] There is an asymmetric equilibrium where only one of the two no-collusion constraints binds. Suppose (8) binds, but (9) does not. It is easy to show that the equilibrium payoff of bank i is lower than that under the symmetric equilibrium, whereas the the payoff of bank j can be lower or higher than that under the symmetric equilibrium depending on the parameter values. The detailed analysis of the asymmetric equilibrium is available upon request.

[15] 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, monitoring efforts remain to be strategic complements because $\delta > 0$ reinforces the effect of the trade-off between rent extraction and incentive provision.

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

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

The aggregate monitoring intensity is given by $\pi(z, r) = 1 - [1 - m_2(z, r)]^2$. *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.,* $m_2^*(r)$. *Whereas for all* $z > \bar{z}_2$, *so that firm quality is high, the monitoring efforts are at the non-delegated level.*

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

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

Proposition 5 is depicted in Figure 3.

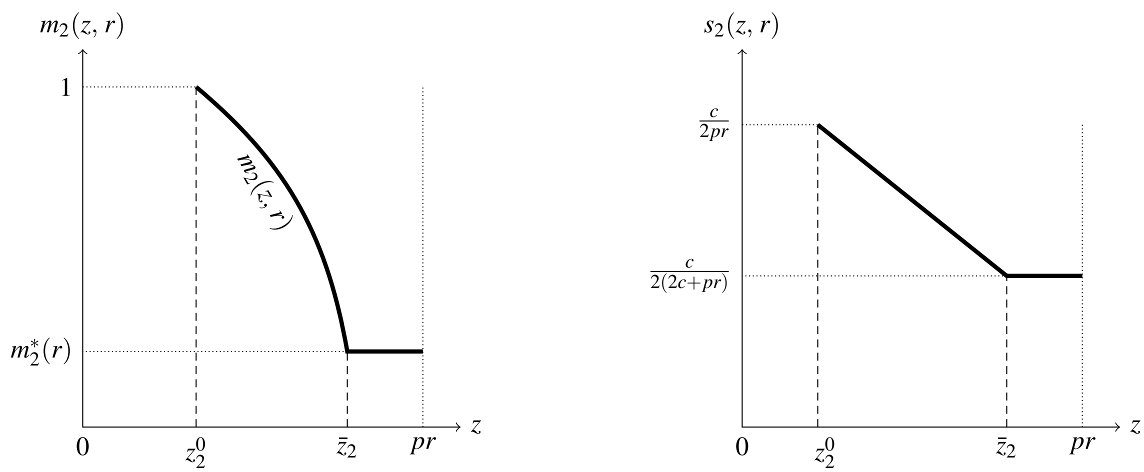


Figure 3: *The equilibrium monitoring effort and share of the loan officers as functions of firm quality. For* $z \leq \bar{z}_2$, *the no-collusion constraints bind, and there is over-monitoring relative to the non-delegated level.*

When the no-collusion constraints bind, there is a positive externality between the banks. With (8) and (9) binding, both s_i and s_j are fixed at a minimum, so that any strategic interaction between the lenders must be through changes in m_i and m_j. This externality is more severe for lower z—if z is very small (very high stake in collusion), s_j is very high due to binding (9). Therefore, the incentive to increase m_i is also high because m_j can be increased by a significant amount.[16]

The equilibrium association between monitoring and borrower quality described in Propositions 2 and 5 are similar because they share a common intuition. Under both lending structures, low values of z is equivalent

[16] Consider $m_i(m_j)$ defined by (14). For a given $\bar{m}_j$, differentiating (14) 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 (15).

to having a high stake in 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 in collusion given that the loan rate is exogenously given. Therefore, we obtain a positive association between the equilibrium share and the stake in collusion. Similar intuition is present in earlier works on vertical collusion (e.g. Tirole, 1986). 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 must have $m_2(z, r) < 1$ (which is equivalent to $z > z_2^0$). If either 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 $m_i = 1$.

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.[17] For completeness however, we take up this issue in Section 6, 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.

Welfare. At any symmetric monitoring effort, $m_i = m_j = m$, the [expected] social welfare, which is the sum of the expected firm utility and the aggregate expected surplus of the two bank-loan officer pairs, is given by:

$$W_2(m) \equiv 2[\pi(m, m)py + (1 - \pi(m, m))v] - cm^2 - 2.$$

The following proposition compares the symmetric equilibrium monitoring with the socially efficient level under two-bank lending.

Proposition 6 *The socially optimal level of monitoring under two-bank lending is given by:*

$$m_2^{**}(z) = \frac{2z}{c + 2z}.$$

Moreover, there is a unique threshold of firm quality $z^{**} \in (z_2^0, \bar{z}_2)$ such that the equilibrium under two-bank lending involves over-monitoring relative to the socially efficient level, i.e. $m_2^{**}(z) \leq m_2(z, r)$ if and only if $z \leq z^{**}$.

Unlike the single-bank lending structure, the equilibrium under two-bank lending may imply under-monitoring relative to the socially optimal level if firm quality is high, i.e., $z > z^{**}$.[18] This happens since $m_2(z, r)$ is too low (equal to the non-delegation level) because of the inefficiency arising due to strategic interaction between two banks even though the collusion incentive problem does not have any bite for high values of z. However, for $z \leq z^{**}$ there is still over-monitoring relative to the socially optimal level. In Section 5.2, we further elaborate on the issue of strategic interaction between the banks.

[17] Hertzberg et al. (2010) analyze the mandated policy of banks to rotate loan officers frequently, and show that such policy incentivizes loan officers to report more accurately. It may be debatable whether rotation of loan officers is an optimal policy, but clearly banks have interests in deterring collusion. In many micro-finance institutions there are clear guidelines for loan officers such as they must not accept gifts from clients, they should not deal with relatives, etc. (e.g. loan officer manual of the Ecuadorian MFI Banco Solidario which is available at `https://www.smartcampaign.org/storage/documents/Tools_and_Resources/Banco_Solidario-Loan_Office_Manual.pdf`). In the U.S., the Sarbanes Oxley Act was introduced in 2002 in order to protect investors by improving the accuracy and reliability of corporate disclosures. Sections 801 to 807 of this act, which are known collectively as the *Corporate and Criminal Fraud Accountability Act*, describe penalties for manipulation, destruction or alteration of financial records, while providing protections for whistle-blowers.

[18] Note that the efficient levels of monitoring under the two lending structures are in general different, i.e., $m_1^{**}(z) \neq m_2^{**}(z)$ because the welfare levels are different.

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 7 *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 $z \leq z_s$.*

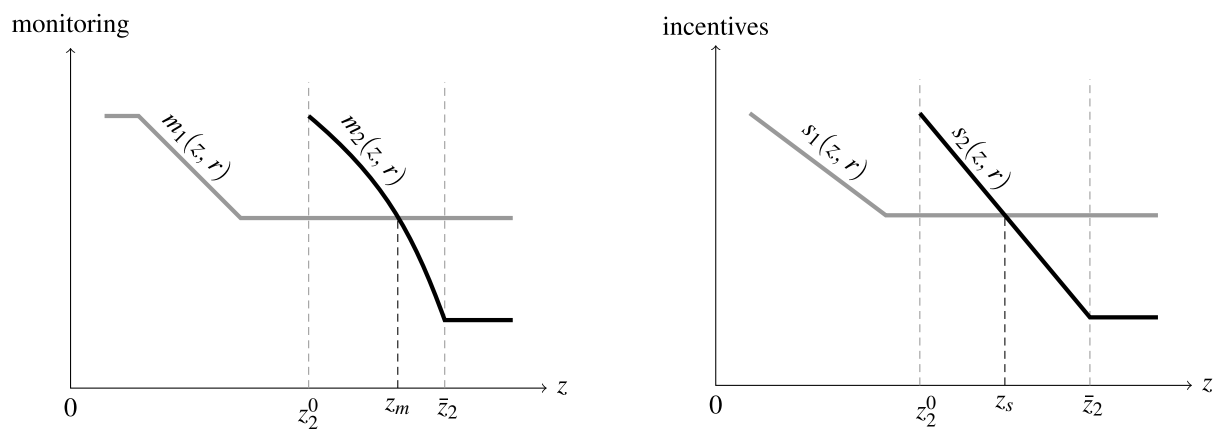


Figure 4: *The equilibrium monitoring efforts and incentives under the single- and two-bank lending structures. Under two-bank lending, each bank elicits higher monitoring effort and provides stronger incentives if and only if firm quality is low.*

The results in Propositions 7 are depicted in Figure 4. Proposition 7 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 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 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, 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 able to monitor the borrower directly (e.g. Carletti, 2004; 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 which amplifies the effect on m_i arising from a larger m_j. Thus for low levels of z, 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 (14) and (15) outward, the race-to-collusion effect is stronger for low values of firm quality causing even a higher equilibrium monitoring effort.[19]

Remark 1 We may have assumed that under two-bank lending, each bank lends \$0.50 to the borrower so that the aggregate loan size is constant across the two lending modes. In the present context this implies the same stake in collusion, $pr - z$ under both lending structures; however, each bank [under two-bank lending] has $r/2$ out of which each loan officer receives a fraction as incentive pay, which is only half of the revenue of the monopoly bank. By contrast, in our model, two-bank lending doubles the size of investment, and hence, the stake in collusion is also doubled, but each bank receives the same repayment as the monopoly bank does. Because our main focus is on the effect of two-bank lending on loan officer incentives, the second specification makes both the lending structures comparable. When each bank lends \$0.50 to the firm, a result similar to Proposition 7 obtains, however over a smaller parameter space (see Dam and Roy Chowdhury, 2020, Proposition 12).

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 this 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:

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

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 5. The following proposition analyzes the role of the free-riding effect in determining the optimal monitoring and incentives.

---

[19]One can also compare the aggregate monitoring intensities under the two lending structures, i.e., $m_1(z, r)$ and $\pi(z, r)$. However, this comparison is not very insightful in the sense that because of public good effect of the monitoring technology lower per bank monitoring under two-bank does not necessarily mean that aggregate monitoring is also lower. It can be shown that if c is not very high, then $\pi(z, r)$ is higher than $m_1(z, r)$ if and only if z is low (see Dam and Roy Chowdhury, 2020).

Proposition 8 *The 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 (8) and (9). 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, per bank as well as aggregate 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 $m_{mb}(z,r)$.

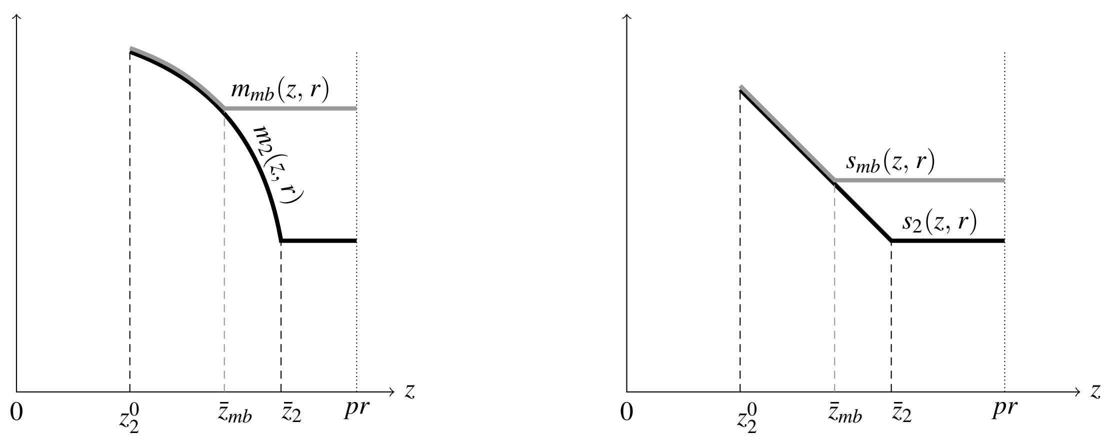


Figure 5: *The equilibrium monitoring effort and share under strategic and merged banks. Subscript 'mb' denotes the variables under merged banks.*

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 the 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 (10), i.e.,

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

i.e., the aggregate surplus from truth telling exceeds that under collusion. Bank i would thus choose (m_i, s_i) to maximize the objective function in (13), subject to (11), (12), and the (joint) no-collusion constraint (10). Comparing the optimal contract in this situation with that described in Proposition 5, we obtain the impact of the race-to-collusion effect on the optimal aggregate monitoring and loan officer incentives.

[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 9 *The monitoring effort is 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.*

Proposition 9 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 (10) 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 $s_{cp}(z, r)$.

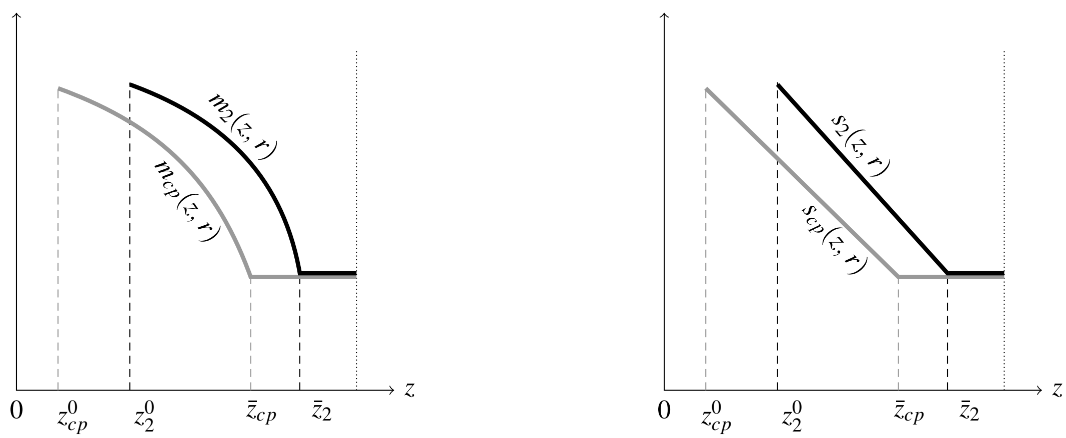


Figure 6: *The equilibrium monitoring effort and share under competing and cooperating loan officers. Subscript 'cp' denotes the variables under cooperating loan officers.*

Proposition 9 is related to the literature on optimal incentive contracts in a multi-agent situation in which an increase in the capacity of the agents to cooperate leads to more efficient outcomes (e.g. Ramakrishnan and Thakor, 1991; Macho-Stadler and Pérez-Castrillo, 1993). In the present context, 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 2 The intuitions developed in Propositions 8 and 9 are quite robust. Proposition 8, 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 9 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.

6 The collusion-proofness principle under two-bank lending

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 has incentives to free-ride on the information gathered by its loan officer, 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 no-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 the collusion-free contract, whereas the other bank does;

(3) An equilibrium strategy profile is said to be (C,C) if neither of them implements the 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 the 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 5], 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, whose effort is given by:*

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

(c) *The equilibrium monitoring effort under two-bank lending (with one bank not implementing the collusion-free contract) is higher than that under single-bank lending if $z \leq \hat{z}_2$. By contrast, the equilibrium 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 $z < z_2^C$, the unique equilibrium is of type (C,NC), so that the collusion-proofness principle fails to hold. The results in Proposition 10 are depicted in Figure 7.

We postpone the detailed technical analysis of the aforementioned equilibria under two-bank lending until the Appendix [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 stake in collusion is now doubled, so that the no-collusion constraint is more difficult to satisfy. 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

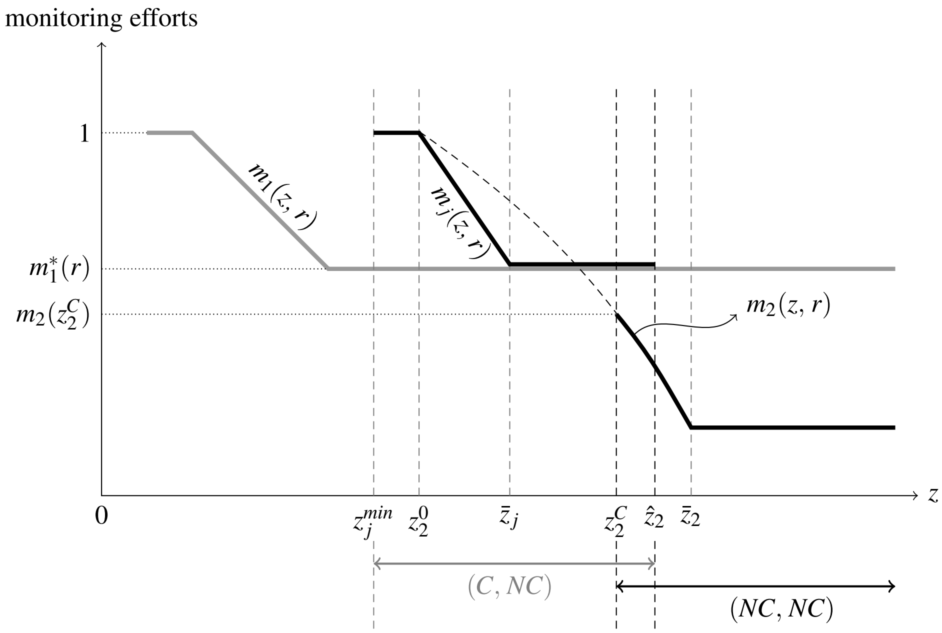


Figure 7: *The equilibrium monitoring efforts under two-bank lending has two segments—$m_j(z, r)$ corresponds to (C, NC) equilibrium, and $m_2(z, r)$ corresponds to (NC, NC) equilibrium. While the segment $m_j(z, r)$ lies above $m_1(z, r)$, the monitoring under single-bank lending, the segment $m_2(z,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

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 dashed 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.[21] From Figure 7, it is also clear 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 7 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 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

[21] In the Appendix we show that in any (C, NC) equilibrium, bank i can employ a loan officer by setting $s_i = 0$ and $m_i > 0$ as long as m_i not too high so that loan officer i's participation constraint is satisfied.

effort.

7 Concluding Remarks

In settings where borrower (mis)behavior can only be verified through costly monitoring by lenders, multiple-bank lending is in general viewed as detrimental to efficiency in the sense that it can lead 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 if monitoring activities must be delegated and there is a possibility of vertical collusion between a loan officer and the borrower, namely a race-to-collusion effect. Under multiple-bank lending, the race-to-collusion effect ensures that if borrower quality is low, then incentives offered to each monitor are stronger in order to deter collusion, leading to a higher level of monitoring.

The main objective of our paper has been to focus on the strategic effects of incentive contracts under multiple-bank lending, and hence, our baseline framework analyzes situations in which banks are constrained to offer collusion-free contracts to the loan officers, finding that under such constraints all banks engage in monitoring. This is an example of *consortium lending* when a group of banks lends 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 a scenario resembles a *lending syndicate* in which one 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 argument in favor of syndicated lending even when the lenders are identical with respect to monitoring costs, and show that when borrower quality is low, syndicated lending dominates consortium lending.

In Dam and Roy Chowdhury (2020), we consider several extensions of our baseline model. We first show that our results remain valid even if we endogenize the loan rates, and the banks in principle can charge different rates. Second, our results also holds under fairly weak assumptions when monitoring effort is not contractible (moral hazard). Finally, Proposition 7 continues to hold when we allow the firm to borrow from $n > 2$ banks.

Our results have important testable implications both with respect to production of 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 resembles that of Hertzberg et al. (2010), although we do not consider a dynamic long-term lender-borrower relationship.

Our framework can find applications, either directly or with some modifications, in many other scenarios with multiple hierarchies. For one, consider a scenario where multiple international organizations provide various kinds of help to a country—one providing aid, another providing technical knowhow, etc., with officers in these organizations overseeing the proper use of these inputs. As a topical example, we can think of a Covid-affected country getting financial aid from the World Bank, loan from a developed nation, and health expertise from the WHO. For another, consider two firms cooperating on a technology-driven project, outsourcing at least a part of technology development to an independent laboratory, with the middle managers in the firms keeping tabs on the laboratory.[22]

[22] Majewski (2004) shows that when firms are direct competitors in the product market, outsourcing collaborative R&D is more likely to occur.

Finally, our results and intuitions generalize to vertically related markets that also involve competing hierarchies. The contractual outcome is similar to a situation where vertically related firms choose between *integration* through vertical merger and *separation* through a delegation contract (e.g. Bonanno and Vickers, 1988). In a vertically integrated firm, the conflict of interest between two production units ceases to exist, and the units take decisions as a single entity. By contrast, under vertical separation, incentive problem becomes non-trivial. Note that the stake in collusion mandates a minimum incentive pay via the no-collusion constraint of each loan officer. When the no-collusion constraint binds in a given bank-loan officer relationship, the incentive pay or equivalently the revenue share is tilted towards the loan officer, and an outcome similar to vertical separation results in. On the other hand, when the no-collusion constraint does not bind, loan officer share is low, and we obtain an outcome similar to vertical integration. Thus, endogenous determination of share implies endogenous choice of ownership structures.[23] Under two-bank lending, the stake in collusion increases, and so does the minimum incentive pay. As a result, vertical separation like outcome becomes more likely.

Appendix: Proofs

Proof of Proposition 2

Consider the maximization problem (5). 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 (2) for any $z \le pr$. Thus, the only relevant feasibility constraint we are left with is $m \le 1$. 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 conditions are given by:

$$\frac{\partial \mathscr{L}}{\partial m} = pr(1-s) + \mu_P(prs - cm) - \mu_F = 0, \tag{16}$$
$$\frac{\partial \mathscr{L}}{\partial s} = m(\mu_P - 1) + \mu_N = 0, \tag{17}$$
$$\mu_k g^k(m, s) = 0 \quad \text{for } k = P, N, F, \tag{18}$$
$$g^k(m, s) \ge 0 \quad \text{for } k = P, N, F, \tag{19}$$
$$\mu_k \ge 0 \quad \text{for } k = P, N, F. \tag{20}$$

First, we consider the case when $\mu_F > 0$ and $m = 1$. We argue that, in this case, the no-collusion constraint (2) must bind. Suppose not, i.e., $\mu_N = 0$. Substituting, $m = 1$ and $\mu_N = 0$ into (17), we obtain that $\mu_P = 1$. Substituting $\mu_P = 0$ into (16), we obtain $pr - c = \mu_F > 0$ which contradicts our assumption that $pr \le c$. 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 (16) and (17), and the complementary slackness condition $\mu_P g^P(m, s) = 0$ associated with the participation constraint, i.e.,

$$z + \mu_P(pr - z - c) = \mu_F, \quad \mu_P + \mu_N = 1, \quad \text{and} \quad \mu_P\left(pr - z - \frac{c}{2}\right) = 0.$$

[23]Similar results are shown by Grossman and Helpman (2002); Legros and Newman (2013); Dam and Serfes (2020) although the our mechanism is quite different, and works through the channel of collusion.

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, (3) 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 $m < 1$, and hence, $\mu_F = 0$. We argue that, in this case, the participation constraint (3) must be binding. If, on the contrary, (3) does not bind, then $\mu_P = 0$, and hence, (16) 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 the no-collusion constraint (2) binds, i.e., $prs = pr - z$. Substituting the value of s into the binding participation constraint (3), we obtain

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

Because (4) 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 (16) and (17), 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 the no-collusion constraint (2) is slack, i.e., $\mu_N = 0$. Thus, substituting $\mu_N = \mu_F = 0$ into (16) and (17), we obtain

$$pr(1 - s) + \mu_P(prs - cm) = 0, \tag{21}$$

$$m(\mu_P - 1) = 0. \tag{22}$$

From (22) 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 (21), we obtain $m = pr/c$. From the binding participation constraint we get $s = 1/2$. Because the no-collusion constraint is slack, we have $pr/2 > pr - z \iff z > \bar{z}_1$.

Proof of Lemma 1

Consider the maximization problem of Bank i. As in the proof of Proposition 2, it is easy to argue that the only relevant feasibility constraints are $m_i \leq 1$ and $m_j \leq 1$. Thus, 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 conditions (for bank i) are given by:

$$\frac{\partial \mathscr{L}_i}{\partial m_i} = (1-m_j)pr(1-s_i) + \mu_P^i\left[(1-m_j)prs_i - cm_i\right] - \mu_F^i = 0, \tag{23}$$

$$\frac{\partial \mathscr{L}_i}{\partial s_i} = \pi(m_i, m_j)(\mu_P^i - 1) + \mu_N^i = 0, \tag{24}$$

$$\mu_k^i g_i^k(m_i, m_j, s_i) = 0 \quad \text{for } k = P, N, F, \tag{25}$$

$$g_i^k(m_i, m_j, s_i) \geq 0 \quad \text{for } k = P, N, F, \tag{26}$$

$$\mu_k^i \geq 0 \quad \text{for } k = P, N, F. \tag{27}$$

Bank j has similar optimality conditions. We first show that, at the optimum, we must have both $m_i < 1$ and $m_j < 1$. Suppose to the contrary that $m_j = 1$. Then, (23) 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 (27). Therefore $m_j < 1$ and $\mu_F^j = 0$. Likewise, we have $m_i < 1$ and $\mu_F^i = 0$.

Next, we show that both the participation constraint must bind at the optimum. Suppose at least one of them, say (11), is slack, and hence, $\mu_P^i = 0$. Given that $\mu_F^i = 0$, (23) 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. The binding participation constraint of loan officer i yields

$$prs_i = \frac{cm_i^2}{2\pi(m_i, m_j)}. \tag{28}$$

Substituting for s_i into the objective function (13) and the no-collusion constraint (8), the above maximization problem of bank i reduces to:

$$\max_{m_i} \ \pi(m_i, m_j)pr - \frac{1}{2}\, cm_i^2 - 1,$$

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

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:

$$pr(1-m_j) - cm_i = 0, \tag{29}$$

$$pr(1-m_i) - cm_j = 0. \tag{30}$$

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.

$$cm_i^2 = 2a\pi(m_i, m_j), \tag{31}$$

$$cm_j^2 = 2a\pi(m_i, m_j), \tag{32}$$

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 (31) 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 $2a \leq c$. Solving for m_i from (31) 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 $2a \geq c$.

Proof of Proposition 5

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) and (s_i, s_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 both the participation constraints are binding. We have two cases—(a) both the no-collusion constraints (8) and (9) 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 (31) and (32), 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. Using (28), 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

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

It is worth noting that, at $z = z_2^0$, each bank's expected payoff is given by $pr - c/2 - 1$ which is positive under Assumption 1.

Next, consider the case when neither (8) nor (9) binds. Because both $m_i(m_j)$ and $m_j(m_i)$ defined by (29) and (30) are linear and downward-sloping, there is a unique solution to the system of equations, which is also symmetric. This is given by:

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

Using (28), 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

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

Proof of Proposition 6

Note that $m_2^{**}(z)$ is strictly increasing and concave on $[z_2^0, pr]$, whereas $m_2(z, r)$ is non-increasing in z. Because $m_2^{**}(z_2^0) = 1 - \frac{2c}{c+4pr} < 1 = m_2(z_2^0, r)$, and $m_2^{**}(\bar{z}_2) - m_2(\bar{z}_2, r) = \frac{pr(7c+4pr)}{4c^2+9cpr+4p^2r^2} - \frac{pr}{c+pr} = \frac{cpr(3c+2pr)}{(c+pr)(4c^2+9cpr+4p^2r^2)} > 0$, the result follows.

Proof of Proposition 7

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 $pr > c/2$, which 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 $z < z_m$. The proof of part (b) is similar because $s_2(z_2^0, r) = c/2pr \geq 1/2 = s_1(z_2^0, r)$ and $s_2(\bar{z}_2, r) = (c/2)(2c+pr) < 1/2 = s_1(\bar{z}_2, r)$.

Proof of Proposition 8

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. The optimal contract (m_i, s_i) between loan officer i and the merged entity solves

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)pr(2 - s_i - s_j) - 2,$$
$$\text{subject to (8), (11), (12).}$$

Denote the threshold values of firm quality and the equilibrium variables with subscript 'mb'. In a symmetric equilibrium, as in Proposition 5, when both no-collusion constraints (8) and (9) bind, the optimal contracts coincide with those under two bank lending. Therefore, we have $z_{mb}^0 = z_2^0$, $m_{mb}(z, r) = m_2(z, r)$ 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

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

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 9

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

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)pr(1 - s_i) - 1,$$
$$\text{subject to (10), (11), (12).}$$

We analyze a symmetric equilibrium, i.e., $m_i = m_j$ and $s_i = s_j$. When the no-collusion constraint (10) is slack, the equilibrium contracts coincide with those under two-bank lending. When (10) binds, we have $s_i = s_j = 1 - z/pr$. It follows from the binding participation constraints that

$$m_i = m_j = m_{cp}(z, r) = \frac{4(pr - z)}{c + 2(pr - z)}.$$

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 10

As discussed earlier, (C, C) cannot be an equilibrium because the banks do 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 5, neither (8) nor (9) binds, so that the banks do not worry about the incentive problems emerging from collusion possibilities. Therefore the equilibrium in Proposition 5 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 lemmata.

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 any bank, say i, wants to deviate. 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:

$$[m_j + (1-\beta)m_i(1-m_j)]prs_i + 2\beta m_i(1-m_j)(pr-z) - \frac{1}{2}cm_i^2 \geq 0. \tag{33}$$

Given that bank i does not implement a collusion-free contract it would optimally would set $s_i = 0$. 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, subject to (33), solves

$$\max_{m_i \in [0,1]} m_j pr - 1.$$

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 5]. 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:

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

Let $B_2(z, r)$ be the payoff of each bank in the (NC, NC) equilibrium at the monitoring effort $m_2(z, r) = 8(pr-z)/(c+4(pr-z))$, which is given by:

$$B_2(z, r) = \frac{16c(pr-z)(2z-pr)}{(c+4(pr-z))^2} - 1.$$

Bank i deviates from the (NC, NC) equilibrium if and only if

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

where note that we ignore the participation constraint (33).[24] To summarize, for all $z \geq z_2^C$, the equilibrium outcomes in Proposition 5 continue to constitute an equilibrium. Thus, the collusion-proofness principle holds in

[24] The participation constraint (33) 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.

the sense that there is one equilibrium where imposing the no-collusion constraints is 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 $B_2(z, r)$ is maximized. $\square$

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

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

Proof From Lemma 2 it follows that there cannot be an equilibrium of type (NC, NC) for any $z \le z_2^C$, and hence, (C, NC) is the only possible equilibrium for these values of z. Clearly, it suffices to analyze an equilibrium where 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

$$\max_{(m_j, s_j) \in [0,1]^2} m_j pr(1 - s_j) - 1,$$

$$\text{subject to } m_j prs_j - \frac{1}{2} cm_j^2 \ge 0,$$

$$prs_j \ge 2(pr - z). \tag{34}$$

The above program is similar to the one of single-bank lending with the only modification that the no-collusion constraint is (34) instead of (2). The optimal monitoring effort of loan officer j is given by:

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

where $z_j^{min} = (pr+1)/2$ such that $B_j(m_j^C(z, r)) \ge 0$ if and only if $z \ge z_j^{min}$, and $\bar{z}_j = 3pr/4 < \bar{z}_2$. As bank i does not deter collusion in this candidate equilibrium, it optimally sets $s_i = 0$, and hence, its expected payoff is given by:

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

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. We first calculate bank i's payoff from such a deviation when it solves the maximization problem (13) taking into account that $m_j = m_j^C(z, r)$.

First, consider the case when the no-collusion constraint of loan officer i binds. In this case, the best reply $m_i(m_j^C(z, r))$ solves (14), i.e., $cm_i^2 = 4(pr - z)\pi(m_i, m_j)$, which is given by:

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

Next, consider the case when the above constraint does not bind. In this case, the best reply $m_i(m_j^C(z, r))$ solves

(6), i.e., $cm_i = pr(1-m_j)$, which is given by:

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

Note that $\tilde{m}_i(z,r)$ is strictly decreasing in z, whereas $\bar{m}_i(z,r)$ is non-decreasing in z (see the left panel of Figure 8), and they intersect each other only once at $z = \bar{z}_i$. In fact, $\bar{z}_i$ solves $m_i(pr/c) = (pr/c)(1-pr/c)$, and is given by:

$$\bar{z}_i = \frac{pr}{4}\left(3 + \frac{c^2}{2c^2 - 2cpr + p^2r^2}\right).$$

It is easy to show that $\bar{z}_i > \bar{z}_2$ under Assumption 1. Thus, for any $z \in [z_j^{min} < \bar{z}_2]$, $\tilde{m}_i(z,r)$ lies strictly above $\bar{m}_i(z,r)$, and 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).[25]

We next turn to comparing bank i's payoff from not implementing the collusion-free contract, 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 $\hat{z}_2 = z^{dev}$. $\square$

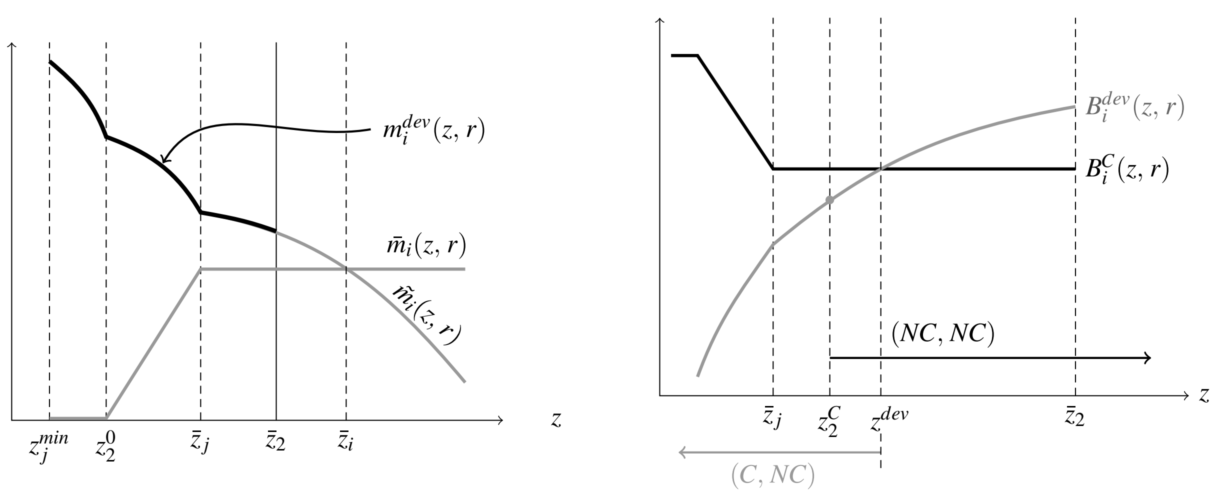


Figure 8: *In the left panel, the dark solid curve represents $m_i^{dev}(z,r)$. The right panel depicts $B_i^C(z,r)$ and $B_i^{dev}(z,r)$. For high values of firm quality, i.e., $z > z^{dev}$ both banks offer collusion-free contracts, whereas for all $z < z_2^C$ at least one bank does not offer collusion-free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

Combining the preceding two lemmata, we have Proposition 10.

[25]Note that if $\tilde{m}_i(z,r) > \bar{m}_i(z,r)$, then the optimal monitoring level is given by $\tilde{m}_i(z,r)$ because 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).

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.

Bonanno, Giacomo and John Vickers (1988), "Vertical Separation." *The Journal of Industrial Economics*, 3, 257–265.

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

Burlando, Alfredo and Alberto Motta (2015), "Collusion and the Organization of the Firm." *American Economic Journal: Microeconomics*, 7, 54–84.

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 Prabal Roy Chowdhury (2020), "Race to Collusion: Monitoring and Incentive Contracts for Loan Officers under Multiple-Bank Lending." Economics and Planning Unit Discussion Paper 20-05. Indian Statistical Institute, Delhi. Available at https://www.isid.ac.in/~epu/dispapers/dp20_05.pdf.

Dam, Kaniṣka and Konstantinos Serfes (2020), "A Price Theory of Vertical and Lateral Integration under Productivity Heterogeneity." Mimeo, CIDE and Drexel University.

Detragiache, Enrica, Paolo Garella, and Luigi Guiso (2000), "Multiple versus 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.

Goldmann, Peter (2010), *Financial Services Anti-Fraud Risk and Control Workbook*. John Wiley & Sons, Inc.

Grossman, Gene and Elhanan Helpman (2002), "Integration versus Outsourcing in Industry Equilibrium." *The Quarterly Journal of Economics*, 85–120.

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.

Khalil, Fahad and Jacques Lawarrée (2006), "Incentives for Corruptible Auditors in the Absence of Commitment." *The Journal of Industrial Economics*, 54, 269–291.

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

Legros, Patrick and Andrew Newman (2013), "A Price Theory of Vertical and Lateral Integration." *The Quarterly Journal of Economics*, 128, 727–770.

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.

Majewski, Suzanne (2004), "How Do Consortia Organize Collaborative R&D? Evidence from the National Cooperative Research Act." Harvard Law School John M. Olin Center for Law, Economics and Business Discussion Paper Series.

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.

Ono, Arito and Iichiro Uesugi (2009), "Role of Collateral and Personal Guarantees in Relationship Lending: Evidence from Japan's SME Loan Market." *Journal of Money, Credit and Banking*, 41, 935–960.

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

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.

Scheepens, Joris (1997), “Collusion and Hierarchy in Banking.” CentER for Economic Research Discussion Paper No. 9701.

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

Strausz, Roland (1997), “Collusion and Renegotiation in a Principal-Supervisor-Agent Relationship.” *Scandinavian Journal of Economics*, 99, 497–518.

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.