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

Kaniṣka Dam† Prabal Roy Chowdhury‡

Abstract

We use a contracting framework to analyze the effect of strategic interaction among bank-loan officer-borrower hierarchies when monitors can possibly collude with the borrower by shading information in favor of the borrower. We identify two countervailing effects which can lead to inefficiencies in the bank-loan officer relationships. First, the 'free-riding' effect that arises from the strategic interaction among banks, and induces lower monitoring by the loan officers and weaker incentives relative to monopoly banking. The second one is the 'race-to-collusion' effect which emerges due to competition among the loan officers for graft, and leads to higher monitoring and stronger incentives. As a consequence, the effect of multiple-bank lending on the optimal incentive contracts is ambiguous relative to those under single-bank lending — when 'bank efficiency' is high, multiple bank lending implies lower monitoring effort; whereas if banks are not efficient, multiple bank lending enhances monitoring effort.

1 Introduction

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

The main objective of our paper is to develop a framework for financial intermediation that can reconcile these apparently conflicting pieces of evidence. To that end, we analyze a credit market in which many banks lend to a firm that can under-report its income, so that non-monitored lending is not feasible, and the task of monitoring is delegated to loan officers as banks are assumed not to have access

*We are grateful to Karna Basu, Fahad Khalil, Alejandro Manelli, Dilip Mookherjee, Saltuk Ozerturk, Ashok Rai, Jaime Sempere and Tridib Sharma for insightful comments.

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

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

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

to any monitoring technology. Berger and Udell (2002) recognize the importance of loan officers in producing soft information in the context of small business (SME) lending. In fact they argue that "the relevant relationship in SME lending is the loan officer-borrower relationship, not the bank-entrepreneur relationship". However, entrusting loan officers with the responsibility of monitoring can be problematic in that these agents may collude with the borrowing firm, and shade reports on the true financial states in its favor (e.g. Stein, 2002; Liberti and Mian, 2009).[2] We establish that the possibility of such *vertical collusion* between loan officers and the borrower generates some hitherto unexplored effects.[3]

We look at a one-period model with one firm and multiple lenders in which firm output can only be verified by costly monitoring. The novel feature of our model is that (unlike e.g. Winton, 1995), monitoring activity is delegated to loan officers. The firm, that owns a constant-returns-to-scale technology project, borrows startup capital from one or two banks at a given loan rate. It can hide the realized cash flow of the project at no cost in order to reduce interest payments. Consequently, any bank that lends has to hire a loan officer to monitor the borrower. While the monitoring level is contractible, its outcome, i.e. whether monitoring has been successful, or not, is not verifiable. In the event that a loan officer succeeds in monitoring, she can either report to her employer the true cash flow, or shade reports in favor of the firm in exchange for a bribe. Thus, the optimal contract involves each bank designing an incentive scheme (a share of repayment) for its loan officer such that she has no incentive to engage in bribery. We assume that each bank comes to know about collusion with an exogenously given probability, which we refer to as *bank efficiency*. In case of multiple-bank lending the optimal contract of one bank-monitor relationship depends on that of the other, i.e. contracts are subject to externality. Using a parsimonious model, we answer the following relevant questions pertaining to the internal organization of a bank under multiple-bank lending:

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

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

We solve for the optimal contract under both monopoly, as well as two-bank lending. Under either scenario, we find that the contractual structure, which consists of a stipulated monitoring effort and a share of repayment, is critically affected by the presence of collusion possibilities. Under both lending structures, each bank essentially maximizes the surplus arising out of the bank-loan officer relationship subject to an incentive constraint, called the *no-collusion constraint*. When bank efficiency, i.e. the exogenous probability of detecting bribery is high, there is little to gain from collusion for both the loan officer(s) and the borrower. Hence, loan officer incentives become perfectly aligned with those of the banks as the no-collusion constraint does not bind at the optimum. As a result, the outcome coincides with the non-delegation outcome, i.e. the case when each bank would have directly monitored the borrower. Collusion however becomes a more serious issue when bank efficiency is low when the

[2]Loan officers may also shirk in providing costly monitoring effort (moral hazard), the issue which we abstract from until Section 8.3 where we relax this assumption, and discuss its implications.

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

no-collusion constraint of each loan officer binds, which implies conflict of interests between a bank and its loan officer (non-aligned incentives). This in turn incentivizes the loan officer to monitor more intensely as bribery is feasible only if there is successful monitoring. Consequently, each bank must provide stronger incentives to deter collusion, and hence, we observe inefficient *over-monitoring* relative to the non-delegation case under both lending modes.

We next turn to comparing the outcomes under one- and two-bank lending. Our central result is that, for lower levels of bank efficiency, two-bank lending induces higher monitoring efforts, as well as stronger incentives, relative to one-bank lending. By contrast, for higher levels of bank efficiency, the results are reversed — two-bank lending yields lower monitoring efforts and weaker incentives. This result can be traced to two countervailing effects that emerge under two bank lending — namely, a *free-riding* effect and a *race-to-collusion* effect. Under two-bank lending, strategic interaction among banks leads to free-riding in monitoring as banks fail to coordinate their actions when they choose monitoring independently. Individual monitoring efforts thus 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). In our model, for high levels of bank efficiency, no bank faces any incentive problem on behalf of the loan officers, and hence, two-bank lending induces lower monitoring effort relative to single-bank lending because of the negative externality arising from the free-riding effect.

By contrast, for low levels of bank efficiency, the no-collusion constraint of each loan officer binds and consequently the individual monitoring efforts becomes strategic complements. Strategic complementarity of monitoring efforts, which we call the race-to-collusion effect, arises due to the following reason. Each bank faces a trade-off between extracting additional rent from its loan officer and providing incentives to deter collusion. With an increase in monitoring by any loan officer, the participation constraint of the other bank's loan officer gets relaxed. But the concerned bank cannot extract the additional rent by lowering incentive pay because this would violate the no-collusion constraint. The only way to exploit the relaxation of the participation constraint is to increase the loan officer's monitoring effort. An increase in the monitoring effort of any loan officer therefore induces an increase in the effort of the other. As the inefficiency at the bank level increases, the strategic complementarity becomes stronger (steeper best reply functions), and hence generating a stronger race-to-collusion effect due to which the incentives for over-monitoring get exacerbated under two-bank lending. As a result, two-bank lending leads to over-provision of monitoring efforts relative to single-bank lending.

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

We finally extend the analysis in several directions. First, we extend the two-bank lending mode to a situation where many (more than two) banks invest in the firm. We find that the earlier results go through qualitatively, in particular the trade-off between the free-riding and the race-to-collusion effects persist. Additionally, we find that with more than two banks the aggregate monitoring intensity unambiguously increases in the number of investors for all levels of bank efficiency as the benefits from public information due to monitoring success outweigh the costs arising from free-riding. Second, we endogenize the loan rates, using the zero-profit condition for each bank (competitive banking market). We find that two-bank lending implies a higher loan rate if and only if bank efficiency is low. Finally, we allow the monitoring level to be non-contractible (moral hazard), so that incentive contracts must not only satisfy the no-collusion constraint, but also an additional effort incentive constraint for each loan officer. Under both these extensions we find that our central results — namely, over-monitoring resulting from binding no-collusion constraint if bank efficiency is low, and second, that monitoring is higher (lower) under two-bank lending when bank efficiency is low (high), remain valid.

2 Related literature and our contribution

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

The paper closest in spirit to ours is by Carletti et al. (2007) who analyze a model with multiple lenders investing and monitoring multiple projects, and monitoring is subject to moral hazard. 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 than the second one, multiple-bank lending induces over-provision of per-project monitoring compared with single-bank lending. By contrast, our paper analyzes the importance of a bank's organizational structure in producing soft information (as in Berger and Udell, 2002; Stein, 2002), and explores a different channel — namely, race-to-collusion through which multiple-bank lending boosts up monitoring.

Prior to Carletti et al. (2007), both Carletti (2004) and Khalil et al. (2007) analyze the effect of multiple-bank lending on monitoring incentives.[5] In Carletti (2004) the central issue is of *ex ante* moral hazard, with monitoring aimed at preventing borrower misbehavior, whereas in Khalil et al. (2007) the central problem is of costly state verification. Further, while Carletti (2004) aims at explaining the

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

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

endogenous choice of alternative lending modes by a borrower, Khalil et al. (2007) focus on whether, under multiple-bank lending, equilibrium contracts are debt-like contracts (as in Winton, 1995). Both these papers though make the same fundamental point that, in the presence of multiple investors, free-riding due to strategic interaction among banks leads to under-provision of monitoring.[6] The present paper however differs from both the aforementioned works in that we explicitly take into account agency problems, in particular collusion possibilities, in bank-monitor-borrower hierarchies. This induces an interplay between the free-riding effect unearthed in Carletti (2004) and Khalil et al. (2007), and collusion incentives for the loan officers and the borrower. Consequently, as discussed above, our results are more nuanced.

The analysis of collusion in hierarchies goes back to Tirole (1986), who argued that while collusion possibilities among supervisors and agents creates additional inefficiencies, employing a supervisor can still be beneficial. Kofman and Lawarée (1993) extend this analysis to allow for external auditors who may have less information relative to internal auditors, but are more honest and can keep a check on collusive behavior. The present paper, while related, differs in that the monitors are symmetric, and both can prevent collusion by the other monitor if she is honest herself. Another related paper is Mookherjee and Png (1995) who study the issue of how to compensate corruptible law enforcers, finding that the collusive possibilities may generate non-obvious results. While Mookherjee and Png (1995) do not allow for multiple law enforcers, they also find that collusive possibilities can lead to over-monitoring.

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

3 The Model

The economy, which spans five dates $t = 0, 1, 2, 3, 4$, consists of three classes of risk neutral agents — a firm (borrower), two banks (lenders) i and j, and two loan officers (monitors) i and j. The indigent firm owns a constant-returns-to-scale project that yields a cash flow of $y \geq 0$ for every dollar invested in it. The project must be financed by bank lending in which each bank lends \$1 at a per unit opportunity cost of capital which we normalize to 1. The value of the per unit cash flow y is observable, but not publicly verifiable, unless a lending bank manages to gather verifiable information on the cash flow. In the absence of any such verification, repayment can therefore only be contingent on the announced per unit cash flow $\hat{y}$ by the borrower. In particular, given the reported cash flow $\hat{y}$, each bank receives a repayment $r\hat{y}$, where $r \leq 1$ is exogenously given.[7] Clearly, the firm has incentives to report a zero per unit cash flow in order to save on repayment, which gives rise to a borrower moral hazard problem. Consequently, non-monitored lending is not feasible because in this case no lender can recoup her initial

[6] While Diamond (1984) makes a fundamental point that non-cooperative contracting can lead to over-monitoring, neither Diamond (1984), nor the literature following from it explicitly allow for strategic interactions among lender, and the consequent contractual externalities.

[7] We endogenize r in Section 8.2.

investment of \$1.

Banks however do not possess any monitoring technology, and hence, must delegate monitoring to loan officers. The monitoring effort $m_i \in [0, 1]$ of loan officer i represents the probability that the loan officer secures verifiable information that the realized cash flow is y. The monitoring *effort* is verifiable, and hence, contractible, however the *outcome* of the monitoring process, i.e., whether it has been successful or not, is not perfectly observable, giving rise to collusive possibilities between the concerned loan officer and the firm. For loan officer i, m_i comes at a cost of

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

Each bank employs exactly one loan officer, with bank i offering its loan officer a contingent contract which specifies the monitoring effort m_i to be exerted. Further, as part of the incentive contract, loan officer i receives a share $s_i \in [0, 1]$ of repayment collected. Suppose loan officer i is successful in monitoring, i.e., she obtains information that can verifiably confirm that the true cash flow is y, while officer j fails in her monitoring efforts. Loan officer i can choose between two actions — report the true cash flow y to her employer, providing verifiable information regarding the same, or collude with the firm. We model collusion in the same spirit as Mookherjee and Png (1995). In the event of collusion, loan officer i has discretion over the cash flow $\hat{y} \in [0, y]$ that she reports to bank i, providing verifiable evidence regarding the same. The borrower may pay a bribe b_i to loan officer i to hide evidence and report $\hat{y}$ rather than y, thus saving on repayment by $r(y - \hat{y})$. In this case, loan officer i receives $s_i r \hat{y}$, an amount smaller than what she would have received by reporting truthfully. We assume away over-reporting, i.e., $\hat{y} > y$ and any possibility of extortion on behalf of the concerned loan officer.[8] We further assume that information about any such collusion is divulged to each bank with an exogenous probability λ, where $0 < \lambda < 1$. The source of such leaks may be an audit by an outsider, gossip among neighbors, or press. We will refer to λ as *bank efficiency* which may be correlated with bank size, proximity to the borrower, etc. as argued by Berger, Miller, Petersen, Rajan, and Stein (2005). In case the graft has been discovered, the borrower must cover the repayment evaded, $r(y - \hat{y})$ to each bank, and the loan officer must pay a penalty $\tau(y - \hat{y})$. We assume, without loss of generality, that $\tau = 0$.[9]

Because monitoring effort is contractible, the sole objective of offering the incentive contract is to deter collusion (ensuring a *collusion-free equilibrium*) so that bank lending is profitable in expectation. Therefore in equilibrium, $ry \geq 1$ can be interpreted as the gross loan rate. We assume that

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

The upper bound on the loan rate guarantees that the optimal monitoring effort is less than 1, whereas the lower bound implies that the repaid amount must cover the cost of capital plus the maximum monitoring cost, i.e., $C(1)$. Note that a necessary condition for the above inequalities to hold simultaneously is that $c > 2$.

In this framework monitoring is a public good, in that if one loan officer, say i, is successful in detecting the true cash flow of the firm (and decides not to collude), then there is verifiable information

[8] As in Kofman and Lawarée (1993), we assume that verifiably mis-reporting the cash flow to banks requires the cooperation of the firm. Thus extortion is not an option since the firm is not going to cooperate in over-reporting the cash flow.

[9] All our results remain qualitatively the same if we assume $\tau > 0$.

that the true cash flow is y, and the firm has to repay ry to each bank. Thus given individual monitoring efforts m_i and m_j and no collusion, the firm repays with probability

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

The aggregate monitoring intensity is increasing in individual monitoring effort since $\pi_i(m_i, m_j) = 1 - m_j \geq 0$. 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.

The timing of events is as follows. At date $t = 0$, each bank lends \$1 to the firm. At date 1, 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 $t = 2$, the cash flow from the project is realized. At date 3, the loan officers monitor at the level specified in the contract. At $t = 4$, the firm and the successful loan officers, if any, decide whether to collude.

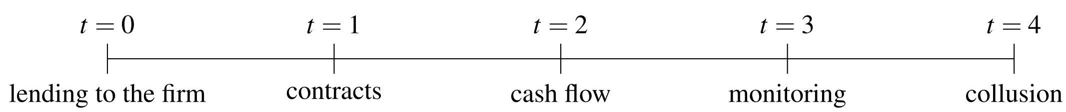

Figure 1: *The timing of events.*

4 Single-bank lending

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

4.1 Direct bank monitoring without delegation

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

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

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

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

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

Clearly, $m_1^*(r)$ is less than 1 by Assumption 1. Interestingly, optimal monitoring is independent of bank efficiency. This is because in the absence of any delegation, the bank does not face any collusion possibility. As we shall later find, collusion incentives do depend on bank efficiency.

4.2 Delegated monitoring

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

4.2.1 The bribery subgame: incentives to deter collusion

At date 4, collusion between the firm and the loan officer occurs only if the loan officer can successfully gather verifiable information about the cash flow of the firm. At this stage, the loan officer can either choose to report the true cash flow y, or he can collude with the borrower in exchange for a bribe b and under-report $0 \leq \hat{y} \leq y$ to the bank. Let $F(\cdot)$ and $M(\cdot)$ respectively be the payoffs of the *f*irm and the *m*onitor. When they collude in exchange of a bribe b, the payoffs are given by:

$$F(\hat{y}) = y - b - \lambda ry - (1-\lambda)r\hat{y}, \quad M(\hat{y}) = sr\hat{y} + b - C(m).$$

By contrast, when the loan officer report truthfully, their payoffs are given by:

$$F(y) = y - ry, \quad M(y) = sry - C(m).$$

Thus, collusion is not feasible if and only if

$$F(y) + M(y) \geq F(\hat{y}) + M(\hat{y}) \quad \Longleftrightarrow \quad s \geq 1 - \lambda. \tag{NC}$$

The preceding inequality, called the *no-collusion constraint* (NC), asserts that if the aggregate surplus from collusion is less than that under truthful reporting, then there are no feasible bribes that make collusion profitable. By contrast, if (NC) does not hold, then the firm and the loan officer have incentives to collude. Clearly the optimal contract must satisfy the non-collusion constraint (NC). Otherwise, the firm would default and the bank would not break even.[10]

4.2.2 Optimal incentives and monitoring

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

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

[10] We follow much of the literature (e.g. Kofman and Lawarée, 1993) in focussing on the no-collusion constraint, and abstracting from the minutiae of any particular bargaining protocol involved in the collusion process. But it is worth noting that (NC) can also be obtained from a generalized Nash bargaining solution which is given by:

$$b^*(\lambda) = \underset{b}{\operatorname{argmax}} \left\{ [F(\hat{y}) - F(y)]^{1-\beta} \, [F(\hat{y}) - F(y)]^{\beta} = [(1-\lambda)r(y-\hat{y}) - b]^{1-\beta} \, [b - sr(y-\hat{y})]^{\beta} \right\} = [\beta(1-\lambda) + (1-\beta)s]ry,$$

given any distribution of bargaining power $(1-\beta, \beta)$, and that the optimal announcement is given by $\hat{y} = 0$. Thus, to deter collusion, the bank must set s such that

$$sry \geq b^*(\lambda) = \{\beta(1-\lambda) + (1-\beta)s\}ry \quad \Longleftrightarrow \quad s \geq 1 - \lambda,$$

which is the same as (NC).

where note that the outside option of the loan officer is normalized to 0, and (ii) the feasibility constraint

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

Note that the contract $(m,\, s)$ is also subject to *limited liability* in the sense that all parties are paid nothing in case of no repayment. We denote by $\mathscr{F}$ the set of *feasible contracts*, i.e., the set of contracts that satisfies (PC), (NC) and (F). The bank solves the following maximization problem:

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

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

Proposition 2 *There are threshold values of bank efficiency, λ_1^{min}, λ_1^0 and $\bar{\lambda}_1$, with $0 < \lambda_1^{min} < \lambda_1^0 < \bar{\lambda}_1 < 1$, where λ_1^{min} is such that bank-lending is not feasible for any $\lambda \leq \lambda_1^{min}$.*

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

$$m_1(\lambda,\, r) = \begin{cases} 1 & \text{for } \lambda \in [\lambda_1^{min},\, \lambda_1^0], \\ \frac{2ry(1-\lambda)}{c} & \text{for } \lambda \in (\lambda_1^0,\, \bar{\lambda}_1], \\ \frac{ry}{c} & \text{for } \lambda \in (\bar{\lambda}_1,\, 1]. \end{cases}$$

Thus whenever bank efficiency is at an intermediate level, i.e. $\lambda_1^{min} \leq \lambda \leq \bar{\lambda}_1$, the equilibrium outcome involves over-monitoring relative to the non-delegated level of monitoring $m_1^*(r)$. Whereas the monitoring level is at the non-delegated level of monitoring $m_1^*(r)$ whenever bank efficiency is high, i.e. $\lambda > \bar{\lambda}_1$.

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

$$s_1(\lambda,\, r) = \begin{cases} 1-\lambda & \text{for } \lambda \in [\lambda_1^{min},\, \bar{\lambda}_1], \\ \frac{1}{2} & \text{for } \lambda \in (\bar{\lambda}_1,\, 1]. \end{cases}$$

The preceding results are depicted in Figure 2.

Proposition 2 is intuitive. In an optimal contracting model under limited liability such as ours, the typical trade-off the bank faces is between incentive provision (setting a high share of repayment in order to satisfy the no-collusion constraint) and rent extraction from the loan officer (setting a low share as long as the participation constraint holds). Thus, (PC) as well as (NC) together play a crucial role in determining the optimal contract. Note that $F(\hat{y}) - F(y) = (1-\lambda)r(y-\hat{y}) - b$ is the net gain to the borrower if he colludes with the loan officer. On the other hand, $M(y) - M(\hat{y}) = sr(y-\hat{y}) - b$ is the net gain to the loan officer from not colluding. Thus, the no-collusion constraint (NC) can be re-written as $M(\hat{y}) - M(y) \leq F(y) - F(\hat{y})$, i.e. the firm's surplus in case of collusion $M(\hat{y}) - M(y)$ is not large enough to lure away the loan officer from truthful reporting when he has a surplus of $F(y) - F(\hat{y})$. First consider a scenario where banks are very efficient, i.e., $\bar{\lambda}_1 < \lambda \leq 1$. In this case the borrower has little to gain from collusion, and hence, too little to offer to the loan officer in the way of bribe. Hence bribery becomes unprofitable, and the incentives of the bank and the loan officer become perfectly aligned, so that the non-delegated level of monitoring, $m_1^*(r)$ can be implemented via a low incentive pay. Moreover, given

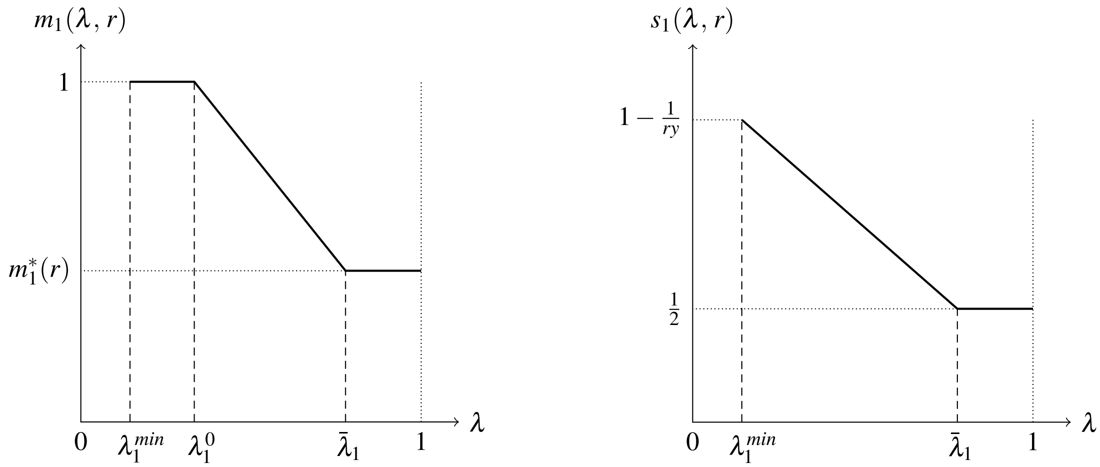

Figure 2: *Equilibrium monitoring and share of the loan officer as functions of bank efficiency. For $\lambda < \lambda_1^{min}$, bank-lending is not feasible. For $\lambda \in [\lambda_1^{min}, \bar{\lambda}_1]$, the no-collusion constraint binds, and there is over-monitoring relative to the non-delegated level $m_1^*(r)$. For For $\lambda > \bar{\lambda}_1$, the no-collusion constraint does not binds and monitoring is at the non-delegated level.*

that the no-collusion constraint (NC) does not bind, the bank can reduce the share of the loan officer until her entire rent is extracted and the participation constraint binds.

By contrast, if banks are not sufficiently efficient, i.e. $\lambda \leq \bar{\lambda}_1$, bribery is very attractive for the borrower, and hence, the no-collusion constraint binds at the optimum. In other words, the incentives of the loan officer and those of her employer point in opposite directions. Therefore, the loan officer monitors more intensely relative to $m_1^*(r)$ in order to increase the possibility of receiving a bribe, implying a loss of efficiency. Consequently, the bank must strengthen incentives in order to deter collusion as bank efficiency diminishes. Finally, when bank efficiency is even lower, i.e., $\lambda_1^{min} \leq \lambda \leq \lambda_1^0$, incentive problem of the loan officer due to collusive threats is so severe that in order to incentivize her not to collude with the borrower, it is not in the interest of the bank extract the entire surplus (slack participation constraint). In other words, it is necessary for the bank to offer *efficiency wage* to the loan officer in order deter collusion. For these levels of bank efficiency, the incentives are set the highest possible level, and the optimal monitoring is at its maximum, i.e., equal to 1. Finally, for the lowest levels of bank efficiency ($\lambda < \lambda_1^{min}$), the loss of efficiency is the maximum, and hence, it is not profitable for the bank to lend anymore. The necessity of offering very strong incentive pay implies that the bank cannot breaks even if it continues to lend. As in Mookherjee and Png (1995), collusive threats (i.e., for $\lambda \leq \bar{\lambda}_1$) turn out to be an inefficient instrument [due to the binding no-collusion constraint] to enhance monitoring.

5 Two-bank lending

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

5.1 Direct bank monitoring without delegation

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

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

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

$$c m_i = ry(1 - m_j), \tag{BR$_i$}$$

$$c m_j = ry(1 - m_i). \tag{BR$_j$}$$

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

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

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

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

Both the individual monitoring effort and aggregate monitoring intensity are independent of the level of bank efficiency because banks do not collude with the firm.

5.2 Delegated monitoring

5.2.1 The bribery subgame

Collusion between the firm and loan officer i occurs only if monitoring by the loan officer is successful. At this stage, the loan officer can choose to either collude with the firm and under-report $\hat{y} \leq y$, or report truthfully. We shall examine subgame perfect Nash equilibria that involves truthful reporting by both the loan officers. Therefore, while examining the collusion decision of loan officer i, we assume that loan officer j does not collude even if he monitors successfully.[11] Given a bribe of b_i, the payoffs from collusion to the firm and loan officer i are respectively given by:

$$F(\hat{y}) = 2y - b_i - 2\lambda ry - 2(1-\lambda) r\hat{y}, \quad M_i(\hat{y}) = s_i r\hat{y} + b_i - C(m_i).$$

The payoffs from truthful reporting, on the other hand are given by:

$$F(y) = 2y - 2ry, \quad M_i(y) = s_i ry - C(m_i).$$

[11]None of our results changes qualitatively if one considers equilibria where truthful reporting is a dominant strategy for both the loan officers.

Thus, collusion between the firm and loan officer i is not feasible if and only if

$$F(y) + M_i(y) \geq F(\hat{y}) + M_i(\hat{y}) \quad \Longleftrightarrow \quad s_i \geq 2(1-\lambda). \tag{NC$_i$}$$

Similarly, the no-collusion constraint for loan officer j is given by:

$$s_j \geq 2(1-\lambda). \tag{NC$_j$}$$

Note that the loan officers do not bargain collectively with the borrower if at least one of them succeeds in monitoring, rather they compete with each other for receiving bribe.

5.2.2 Optimal incentives and monitoring

At date $t = 2$, each bank offer a contract to its loan officer, taking the contract offer of the rival bank as given. Apart from satisfying the no-collusion constraints, the contracts (m_i, s_i) and (m_j, s_j) must satisfy the following *participation constraints* of loan officers i and j

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

$$\pi(m_i, m_j) r y s_j - \frac{1}{2} c m_j^2 \geq 0, \tag{PC$_j$}$$

respectively, and the feasibility constraints

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

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

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

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

As in Proposition 2, the participation constraint of each loan officer will be non-binding for very low levels of bank efficiency. In this case, $m_i = m_j = 1$, and hence individual monitoring efforts are independent of each other.[12] Let $m_i(m_j)$ and $m_j(m_i)$ denote the best reply functions derived by solving ($\mathscr{M}_i$) when the participation constraints of the loan officers bind at the optimum. Lemma 1 derives properties of the best reply functions $m_i(m_j)$ and $m_j(m_i)$ that will be critical to whether individual monitoring level under single-bank lending is lower relative to that under two-bank lending or not (see the discussion following Proposition 5).

Lemma 1 *Consider the best reply functions in individual monitoring efforts, $m_i(m_j)$ and $m_j(m_i)$, when the participation constraints of the loan officers bind.*

[12] Constraints (PC$_i$) and (PC$_j$) both bind whenever at the optimum we have $m_i = m_j = 1$, i.e., in case of non-binding participation constraints, the individual monitoring efforts are independent of each other (flat best replies).

(a) *When neither of the two no-collusion constraints binds, the best reply functions are negatively sloped.*

(b) *When both the no-collusion constraints bind, the best reply functions are positively sloped.*

Suppose the participation constraints of the loan officers bind at the optimum. In this case, when neither of the two no-collusion constraints binds, loan officer incentives are perfectly aligned with those of the banks, so that the best reply functions are given by (BR_i) and (BR_j). Thus, the fact that $\pi_{ij}(m_i, m_j) < 0$ ensures that individual monitoring efforts are strategic substitutes, so that the best reply functions are negatively sloped. By contrast, when both the no-collusion constraints bind, the best replies $m_i(m_j)$ and $m_j(m_i)$ solve the following two non-collusion constraints respectively:

$$cm_i^2 = 4ry(1-\lambda)\pi(m_i, m_j), \tag{BR_i'}$$

$$cm_j^2 = 4ry(1-\lambda)\pi(m_i, m_j). \tag{BR_j'}$$

The best reply functions are upward-sloping for $m_i, m_j \in [0, 1]$. Intuitively, the trade-off each bank faces is either to provide incentives to deter collusion or to extract rent from its loan officer. 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, the bank cannot lower the loan officer's share as such a decrease would violate 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.[13] Moreover, given that bank i is constrained by (NC_i), it has already been operating at a sub-optimal level of m_i. Thus, an increase in m_i in this case increases the bank's profits.[14]

The optimal monitoring and share of repayment for each loan officer are characterized in Proposition 4 below. Under this lending mode, as in the case of single-bank lending, collusive threats exacerbate incentives for over-monitoring, and hence, banks are required to offer stronger incentives to deter collusion. Moreover, for very low values of bank efficiency banks do not break-even, and hence, stay out of business.

Proposition 4 *There are threshold values of bank efficiency, λ_2^{min}, λ_2^0 and $\bar{\lambda}_2$ with $0 < \lambda_2^{min} < \lambda_2^0 < \bar{\lambda}_2 < 1$, such that bank-lending is not feasible for any $\lambda \le \lambda_2^{min}$.*

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

$$m_2(\lambda, r) = \begin{cases} 1 & \text{for } \lambda \in [\lambda_2^{min}, \lambda_2^0], \\ \frac{8ry(1-\lambda)}{c+4ry(1-\lambda)} & \text{for } \lambda \in (\lambda_2^0, \bar{\lambda}_2], \\ \frac{ry}{c+ry} & \text{for } \lambda \in (\bar{\lambda}_2, 1]. \end{cases}$$

[13] Note that the right-hand-side of (12) is stricty increasing in m_i.

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

The aggregate monitoring intensity is given by:

$$\pi(\lambda, r) = \begin{cases} 1 & \textit{for } \lambda \in [\lambda_2^{min}, \lambda_2^0], \\ 1 - \left(\frac{c-4ry(1-\lambda)}{c+4ry(1-\lambda)}\right)^2 & \textit{for } \lambda \in (\lambda_2^0, \bar{\lambda}_2], \\ 1 - \left(\frac{c}{c+ry}\right)^2 & \textit{for } \lambda \in (\bar{\lambda}_2, 1]. \end{cases}$$

Thus whenever bank efficiency is at an intermediate level, i.e. $\lambda_2^{min} \le \lambda \le \bar{\lambda}_2$*, the outcome involves over-monitoring relative to the non-delegated level of monitoring, i.e.* $\pi^*(r)$*. Whereas for all* $\lambda > \bar{\lambda}_2$*, so that firm quality is high, the monitoring level is at the non-delegated level of monitoring.*

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

$$s_2(\lambda, r) = \begin{cases} 2(1-\lambda) & \textit{for } \lambda \in [\lambda_2^{min}, \bar{\lambda}_2], \\ \frac{c}{2(2c+ry)} & \textit{for } \lambda \in (\bar{\lambda}_2, 1]. \end{cases}$$

Proposition 4 is depicted in Figure 3. The intuition for this proposition is very similar to that for Proposition 2 and hence we refrain from discussing it.[15]

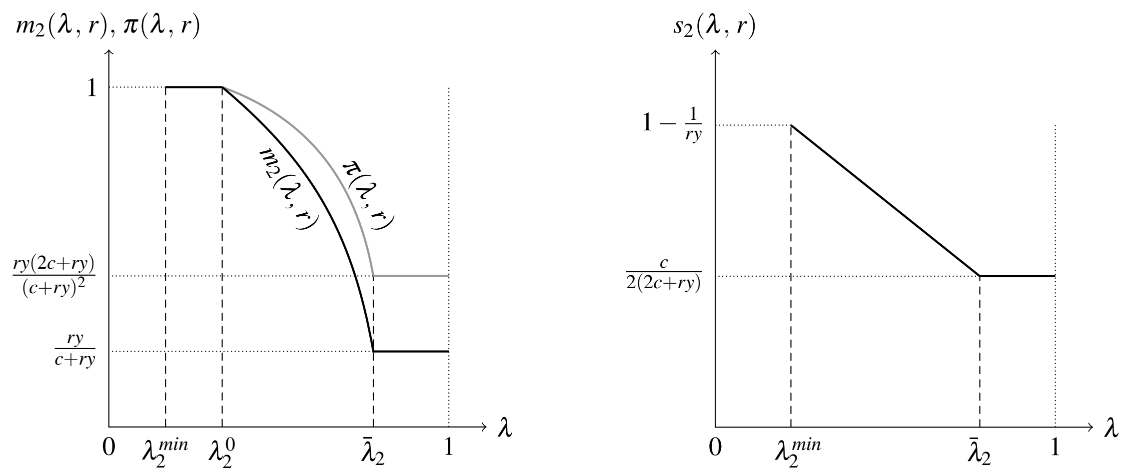

Figure 3: *The equilibrium individual monitoring effort, aggregate monitoring intensity and share of the loan officer as functions of bank efficiency. For* $\lambda < \lambda_2^{min}$*, bank-lending is not feasible. For* $\lambda \in [\lambda_2^{mint}, \bar{\lambda}_2]$*, the no-collusion constraints bind, and there is over-monitoring relative to the non-delegated level.*

6 Comparison of the two lending modes

In this section, we turn to the central question of this paper, namely 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.

[15] The concavity of $m_2(\lambda, r)$ and $\pi(\lambda, r)$ with respect to λ arises because the individual monitoring efforts are substitutes in determining the aggregate monitoring intensity. In the case when $\pi(m_i, m_j) = m_i + m_j + \delta\, m_i m_j$, the functions $m_2(\lambda, r)$ and $\pi(\lambda, r)$ would have been strictly decreasing and convex on $[\lambda_2^0, \bar{\lambda}_2]$.

Proposition 5 *Two bank lending implies larger collusive threats, i.e., the no-collusion constraint of each loan officer binds over a larger range of values of bank efficiency ($\bar{\lambda}_2 > \bar{\lambda}_1$), and hence, exacerbates incentives for inefficient over-monitoring. Moreover, for a given loan rate r, there are unique threshold values of bank efficiency, $\lambda_m, \lambda_s \in (\lambda_2^{min}, \bar{\lambda}_2)$ such that:*

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

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

The results in Propositions 5 are depicted in Figure 4.

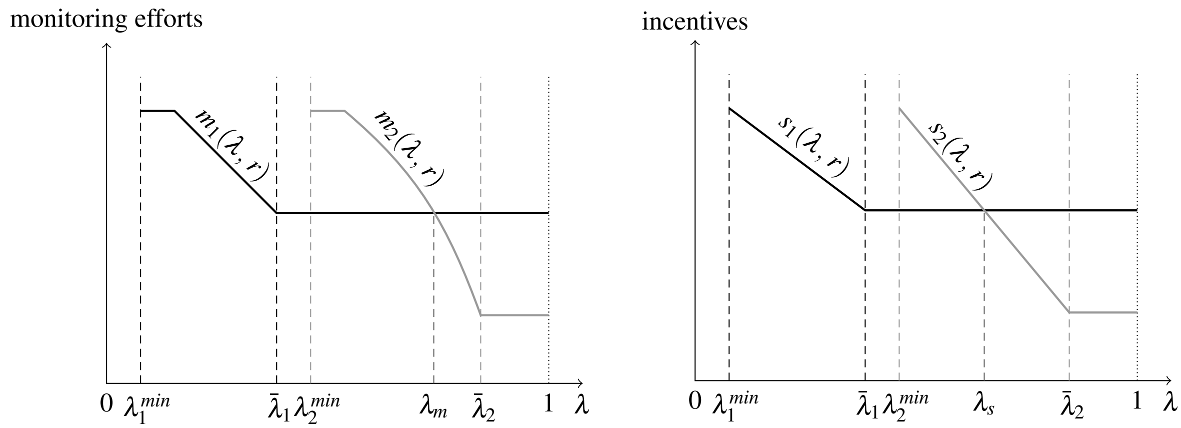

Figure 4: *The equilibrium individual monitoring efforts and incentives under the single- and two-bank lending structures. Under two-bank lending, each bank elicits higher monitoring effort and provides stronger incentives if and only if bank efficiency is low.*

Proposition 5 conveys the central message of our paper – when bank efficiency 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 bank inefficiency is at an intermediate level, two-bank lending implies more intense monitoring and stronger incentives to each loan officer. These results are driven by the two countervailing channels unearthed in Lemma 1 earlier. The first one is the strategic substitutability of individual monitoring efforts which gives rise to the *free-riding effect*, and the second one is the strategic complementarity that leads to the *race-to-collusion effect*. Clearly, the two effects point in opposite directions, and hence, the effect of two-bank lending on individual monitoring effort and loan officer incentives is ambiguous.

Which one of these two effects dominates depends on the level of bank efficiency, i.e. λ. When bank efficiency is high, there is no feasible bribe that makes collusion profitable, and hence, loan officer incentives are perfectly aligned with those of the banks, and the race-to-collusion effect plays no role. Only the free-riding effect is at play here, leading to lower monitoring effort for each monitor under two-bank lending. One would have obtained the same result under a common loan rate *r* had the banks been able to monitor the borrower directly, which has been shown by Carletti (2004) and Khalil et al. (2007). On the other hand, when bank efficiency is at an intermediate level, the no-collusion constraints bind. As we have argued earlier, this leads to over-monitoring under both single-bank, as well as two-bank lending. Under two-bank lending though, an additional effect comes into play — namely, the race-to-collusion effect, that amplifies the effect on m_i arising from a larger m_j. Thus for intermediate levels of bank efficiency, *individual* monitoring effort under two-bank lending exceeds that under single-bank lending. Consequently, the share of repayment must also be higher in an effort to deter collusion. Moreover, note that a lower λ shifts both the best reply functions given by (BR_i') and (BR_j') outward causing a higher equilibrium monitoring effort, so that the race-to-collusion effect is stronger for intermediate values of bank efficiency.[16]

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

Proposition 6 *For a given loan rate r, if $ry > 0.62c$, then there is a unique threshold value of bank efficiency, $\lambda_\pi \in (\lambda_2^{min}, \bar{\lambda}_2)$ such that the aggregate monitoring intensity under two-bank lending is higher than that under single-bank lending, i.e. $\pi(\lambda, r) > m_1(\lambda, r)$ if and only if $\lambda < \lambda_\pi$. By contrast, if $ry \leq 0.62c$, we have $\pi(\lambda, r) \geq m_1(\lambda, r)$ for all $\lambda \in [\lambda_2^{min}, 1]$.*

While comparing the aggregate monitoring levels under single- and multi-bank lending, a comparison of the individual monitoring levels alone provides a partial picture. There is an additional effect in play here in that monitoring is a public good, so that the presence of an additional monitor ensures that $\pi(m_i, m_j)$ is higher than both m_i and m_j. Further the public good effect, measured by $\frac{\partial \pi(m_i,m_j)}{\partial m_i} = 1 - m^*$ (from (1)), is larger when m^* is small, i.e. $\frac{ry}{c}$ is small. First consider the case when banks are highly efficient, i.e. $\lambda > \lambda_m$. While the free-riding effect ensures that individual monitoring effort is lower under two-bank lending, the aggregate monitoring intensity is not necessarily lower because of the public good effect. If $ry/c \leq 0.62$, then the public good effect is large as argued, ensuring that under two-bank lending, aggregate monitoring is higher, even though individual monitoring levels are lower (see Proposition 9 for a more general statement of the above results).[17] However, when the marginal benefit from individual monitoring effort is high relative to its marginal cost, i.e. $ry/c > 0.62$, the negative effect dominates, and hence, even the aggregate monitoring intensity is lower under two-bank lending. Next consider the case when bank efficiency is low. In this case the race-to-collusion effect ensures that individual monitoring levels are higher under two-bank lending, and hence, given the public good effect, so is the aggregate monitoring intensity.

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

$$\frac{dm_i}{d\lambda} = -\frac{2ry\pi(m_i, \bar{m}_j)}{cm_i - 2ry(1-\lambda)(1-\bar{m}_j)}.$$

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

[17] The fact that $ry > 0.62c$ is a necessary and sufficient condition for the results described in Part (b). Recall that we have assumed $ry \geq 0.5c + 1$. When $c < 8.33$, we have $0.5c + 1 > 0.62c$. Thus, $c < 8.33$ is a sufficient condition for $ry > 0.62c$.

7 Disentangling the two countervailing effects

In the previous section we demonstrate that a move from one-bank to two-bank lending has ambiguous consequences on monitoring, as well as loan officer incentives. We have argued that this is because of the interaction between two countervailing channels — namely, the free-riding and the race-to-collusion effects. In a bid to disentangle the two, in what follows we shut down one channel at a time, and examine how the shutting down of a particular channel affects monitoring and incentives under two-bank lending.

7.1 Strategic versus merged banks: the free-riding effect

We first compare the baseline two-bank framework analyzed in Section 5 with one where the free-riding channel is shutdown. To that end we examine a scenario where the two banks merge, with the merged bank maximizing joint-profits, but still employing two independent loan officers. The fact that the merged bank maximizes joint profits ensures that there is no free-riding, whereas the fact that the bank still employs two loan officers implies that the race-to-collusion effect is still present. 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:

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

We compare the optimal solution of the above maximization problem with that described in Proposition 4. The following proposition analyzes the role of the free-riding effect in determining the optimal monitoring and incentives.

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

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

Proposition 7 finds resonance in the literature on strategic managerial incentives in oligopolistic markets (e.g. Martin, 1993; Golan, Parlour, and Rajan, 2015; Dam and Robinson-Cortés, 2018), where strategic interaction among firms results in weaker incentives for managers, generating efficiency losses. In these models initially inefficient firms hire a manager apiece whose principal task is exert R&D effort so as to improve productive efficiency. The cost realization of each firm is independent of those of its rivals, and hence, there is no contract externality. When there are more firms in the market, output, and hence, profits per capita declines. This results in a relatively lower marginal value of cost reduction,

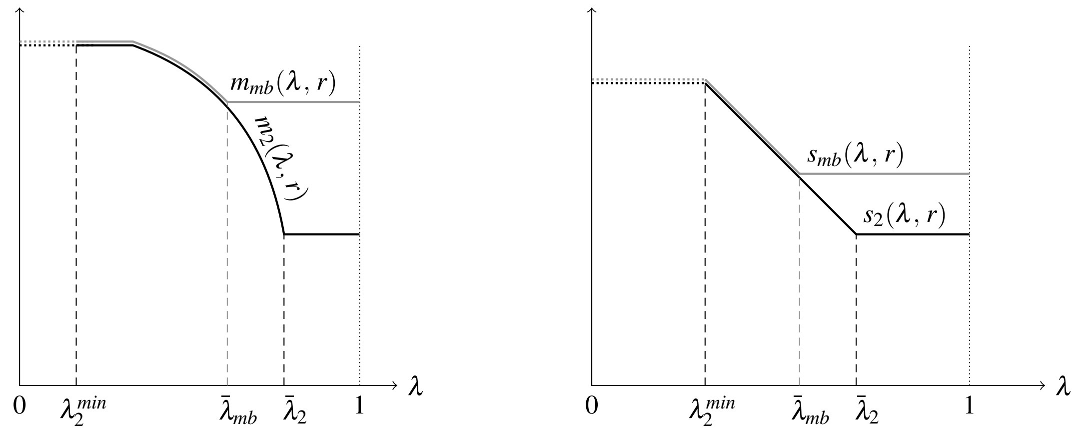

Figure 5: *The equilibrium monitoring effort and share under strategic and merged banks. Subscript 'mb' denotes the variables under merged banks.*

and thus lower marginal benefit of managerial effort. Consequently, as the level of strategic interaction intensifies with a growing number of firms, each firm finds it optimal weaken managerial incentives that elicit lower managerial efforts in equilibrium.

7.2 Competitive versus cooperative loan officers: the race-to-collusion effect

Next we analyze the effect of shutting down the race-to-collusion effect, comparing the baseline framework in Section 5 with one where the loan officers behave cooperatively in the sense that if at least one of them succeeds in detecting borrower misbehavior, then they share this information among themselves. Further, they jointly collect the bribe in case they decide to collude with the firm.[18]

Consider a scenario where at least one of the loan officers has succeeded, and the loan officers decide to collude jointly with the borrower, taking a bribe in exchange for under-reporting the cash flow. Given the bribes of b_i and b_j, the payoffs from collusion to the firm and the loan officers are respectively given by:

$$F(\hat{y}) = 2y - b_i - b_j - 2\lambda ry - 2(1-\lambda)r\hat{y},\ M_i(\hat{y}) = s_i r\hat{y} + b_i - C(m_i),\ M_j(\hat{y}) = s_j r\hat{y} + b_j - C(m_j).$$

The payoffs from truthful reporting, on the other hand are given by:

$$F(y) = 2y - 2ry,\ M_i(y) = s_i ry - C(m_i),\ M_j(y) = s_j ry - C(m_j).$$

Thus, collusion between the firm and loan officers is not feasible if and only if

$$F(y) + M_i(y) + M_j(y) \geq F(\hat{y}) + M_i(\hat{y}) + M_j(\hat{y}) \quad \Longleftrightarrow \quad s_i + s_j \geq 2(1-\lambda). \qquad (\mathrm{NC}_{ij})$$

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

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

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

This proposition 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. To see this, consider first the no-collusion constraints (NC_i) and (NC_j). Note that (NC_i) can be written as

$$s_i r(y-\hat{y}) - b_i \geq 2(1-\lambda)r(y-\hat{y}) - b_i.$$

The right-hand-side of the above inequality is the net gain to the borrower from colluding with loan officer i, and the left-hand-sides is the net gain to loan officers i from truthful reporting. Thus, the binding no-collusion constraints dictate the minimum share of repayment to the concerned loan officer to deter collusion. A similar interpretation of (NC_j) goes through. When the loan officers cooperate, it follows from binding (NC_{ij}) that the minimum shares s_i and s_j are lower than that under competition between the loan officers. 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 $m_2(\lambda, r)$ to $m_{cp}(\lambda, r)$. Consequently, the required incentives are weaker, i.e., the optimal share reduces from $s_2(\lambda, r)$ to $s_{cp}(\lambda, r)$.

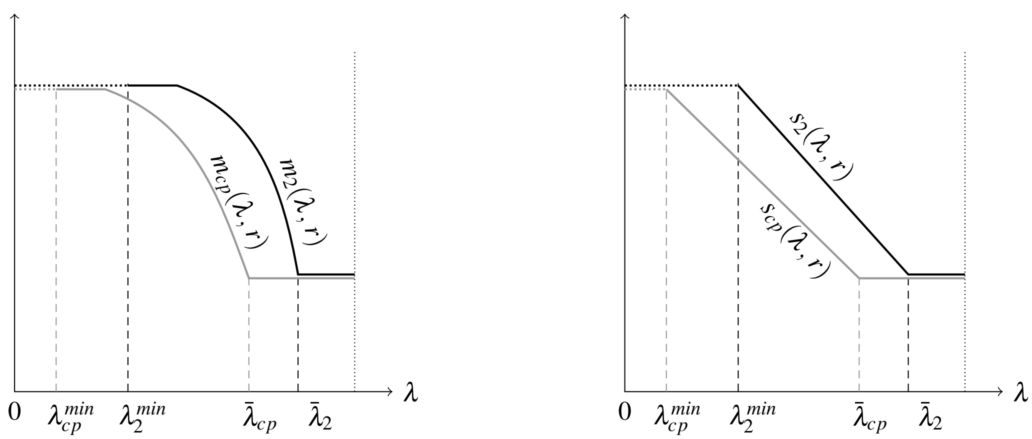

Figure 6: *The equilibrium monitoring effort and share under competing and cooperative loan officers. Subscript 'cp' denotes the variables under cooperative loan officers.*

Proposition 8 is related to the literature on optimal incentive contracts in a multi-agent situation. In analyzing an organization with a single principal and two agents, Ramakrishnan and Thakor (1991) assert that the principal may prefer that the agents cooperate, rather than compete. Similarly, Macho-Stadler and Pérez-Castrillo (1993) demonstrate that an increase in the capacity of the agents to cooperate leads to more efficient outcomes. The above proposition points in a similar direction. If banks could have enforced cooperation between the loan officers, the loss in efficiency due to incentives for over-monitoring would have been minimized. However, in our framework, banks cannot ensure cooperation between the loan officers because collusion is not always publicly verifiable (i.e. $\lambda < 1$).

Remark 1 *The intuitions developed in Propositions 7 and 8 are quite robust. Proposition 7, for example, goes through if, under both strategic and merged banks, loan officers behave cooperatively, rather than* *competitively. The results can be summarized in a figure that is very similar to Figure 5. Similarly, Proposition 8 goes through if, under both competitive and cooperative loan officers, we assume that the banks are merged, rather than competitive. Again the results can be summarized in a figure that is very similar to Figure 6.*[19]

8 Extensions

In the preceding analysis we have kept the framework parsimonious in an effort to focus on the essential trade-offs between the race-to-collusion and the free-riding effect. We now extend the analysis in various directions as robustness checks.

8.1 Lending with many banks

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

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

Mimicking our preceding argument, the no-collusion constraint of loan officer i in the bribery subgame is given by:

$$s_i \geq n(1 - \lambda). \tag{NC$_i^n$}$$

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

$$\pi(\mathbf{m}) r y s_i - \frac{1}{2} c m_i^2 \geq 0, \tag{PC$_i^n$}$$

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

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

We look at the symmetric equilibrium in the choice of monitoring efforts and incentives, i.e. $m_i = m$ and $s_i = s$ for all $i \in \{1, \ldots, n\}$.[20] As before, when none of the participation constraints bind at the optimum, we have $m_i = 1$ for all i. In this case, $\pi(\mathbf{m}) = 1$ and $s_i = n(1 - \lambda)$ for all i. By contrast, when the participation constraint (PC$_i^n$) of each loan officer i binds, there are two situations. When none of the no-collusion constraints bind, the symmetric equilibrium monitoring efforts are given by the following first-order condition:

$$(1 - m)^{n-1} = \gamma(r) m, \tag{$\star$}$$

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

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

where $\gamma(r) \equiv c/ry \geq 1$. The above equilibrium condition yields the monitoring effort which is same as the non-delegated level of monitoring effort. By contrast, when all the no-collusion constraints bind, the symmetric equilibrium share is given by $s = n(1-\lambda)$, and the equilibrium monitoring effort solves

$$2(1-\lambda)n[1-(1-m)^n] = \gamma(r)m^2. \qquad (\star\star)$$

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

Proposition 9 *Let $m(\lambda, n)$, $s(\lambda, n)$ and $\pi(\lambda, n)$ respectively denote the monitoring effort, share of repayment, and the aggregate monitoring intensity in the symmetric equilibrium.*

(a) *For each $n \geq 2$, there are threshold values $\lambda^{min}(n)$, $\bar{\lambda}(n) \in (0, 1)$ of bank efficiency with $\lambda^{min}(n) < \bar{\lambda}(n)$ such that bank lending is not feasible under n-bank lending for any $\lambda < \lambda^{min}(n)$, and the equilibrium outcome involves over-monitoring for all $\lambda \in [\lambda^{min}(n), \bar{\lambda}(n)]$. Moreover, the threshold $\bar{\lambda}(n)$ is strictly increasing in n with $\lim_{n\to\infty} \bar{\lambda}(n) = 1$;*

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

(c) *For any two n and n' with $n' > n \geq 2$, equilibrium aggregate monitoring intensity is always lower under n-bank lending, i.e. $\pi(\lambda, n) < \pi(\lambda, n')$ for all $\lambda \in [\lambda^{min}(n'), 1]$.*

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

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

8.2 Endogenous loan rate

In the baseline model we have not only assumed the loan rate to be exogenous, but also that the rates are the same under single- and two-bank lending modes in order to make the two lending structures comparable. However, in reality, rates are endogenous and vary according to market structure. We therefore endogenize the loan rates under both lending modes, and show that the results presented in Proposition 5 remain valid. We will assume that the banking market is competitive, so that loan rates are endogenously determined from the zero profit conditions.

8.2.1 Single-bank lending

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

$$R_1(\lambda) = \begin{cases} \frac{1}{\lambda} \quad \text{for } \lambda \in [\lambda_1^{min}, \lambda_1^0], \\ \sqrt{\frac{c}{2\lambda(1-\lambda)}} \quad \text{for } \lambda \in (\lambda_1^0, \bar{\lambda}_1], \\ \sqrt{2c}, \quad \text{for } \lambda \in (\bar{\lambda}_1, 1]. \end{cases}$$

The threshold values of bank efficiency, λ_1^{min}, λ_1^0 and $\bar{\lambda}_1$ are derived in the Appendix (see the proof of Proposition 10). Note that $R_1(\lambda)$ is strictly decreasing and convex on $[\lambda_1^{min}, \bar{\lambda}_1]$, and independent of bank efficiency thereafter. Substituting the above expression into that of optimal monitoring and share of repayment described in Proposition 2, we obtain the equilibrium monitoring and incentives under single-bank lending.

8.2.2 Two-bank lending

Under two-bank lending with delegated monitoring and the aggregate loan amount \$2, consider a time line where the banks first post r_i and r_j, with the subsequent game following the action sequence described in Figure 1. We first analyze the no-collusion constraint of each loan officer. Given r_i and r_j, and the bribe b_i to loan officer i, the payoffs from collusion to the firm and loan officer i are respectively given by:

$$F(\hat{y}) = 2y - b_i - \lambda(r_i + r_j)y - (1-\lambda)(r_i + r_j)\hat{y}, \quad M_i(\hat{y}) = s_i r_i \hat{y} + b_i - C(m_i).$$

The payoffs from truthful reporting, on the other hand are given by:

$$F(y) = 2y - (r_i + r_j)y, \quad M_i(y) = s_i r_i y - C(m_i).$$

Thus, collusion between the firm and loan officer i is not feasible if and only if

$$F(y) + M_i(y) \geq F(\hat{y}) + M_i(\hat{y}) \quad \Longleftrightarrow \quad s_i R_i \geq (1-\lambda)(R_i + R_j), \tag{NC_i^0}$$

where $R_i \equiv r_i y$ and $R_j \equiv r_j y$. Similarly, the no-collusion constraint for loan officer j is given by:

$$s_j R_j \geq (1-\lambda)(R_i + R_j). \tag{NC_j^0}$$

The participation constraints are given by:

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

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

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

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

subject to (PC_i^0), (NC_i^0) and (F_i).

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

$$R_2(\lambda) = \begin{cases} \frac{1}{2\lambda - 1} & \text{for } \lambda \in [\lambda_2^{min}, \lambda_2^0], \\ \frac{c}{4\left(\sqrt{c(1-\lambda)(2\lambda-1)} + c(1-\lambda)\right)} & \text{for } \lambda \in (\lambda_2^0, \bar{\lambda}_2]. \end{cases}$$

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

$$R_i^0 = \frac{c\left(4(1+c) - \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2} \text{ and } R_j^0 = \frac{c\left(4(1+c) + \sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}$$

for $\lambda \in (\bar{\lambda}_2, 1]$. The above rates are independent of bank efficiency.[21] Because $R_i^0 \neq R_j^0$, the equilibrium monitoring efforts m_i and m_j, as well as the equilibrium shares s_i and s_j are also asymmetric. The threshold values of bank efficiency, λ_2^{min}, λ_2^0 and $\bar{\lambda}_2$ are derived in the Appendix.

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

8.2.3 Comparison of the two lending modes under endogenous loan rates

We compare the equilibrium loan rates and loan officer contracts under the two lending structures.

Proposition 10 (a) *There is a unique threshold value of bank efficiency, $\lambda_R \in (\lambda_2^{min}, \bar{\lambda}_2)$ such that the equilibrium loan rates under two-bank lending is higher than that under single-bank lending if and only if $\lambda < \lambda_R$;*

(b) *Moreover, the individual monitoring efforts and the aggregate monitoring intensity are higher (lower), and the incentives for loan officers are stronger (weaker) under two-bank lending for low (high) values of bank efficiency.*

The above proposition shows that the result of Proposition 6 is robust even if the loan rates are endogenous. Because each loan officer receives a fraction of the repayments by the firm, under endogenous loan rates, each bank has an additional instrument to incentivize its loan officer because the incentive pays are sR under single-bank lending, and s_iR_i and s_jR_j under two-bank lending. In the presence of collusion incentive problem, stronger incentive pays (both loan rate and share) are required to deter collusion under two-bank lending because of the race-to-collusion effect. By contrast, when such incentive problems are absent, under two-bank lending, a lower monitoring is elicited by a weaker incentive pay. The comparison between the equilibrium loan rates can be depicted graphically in a figure similar to Figure 4.

8.3 Non-contractible monitoring effort

The main objective of our paper is to highlight the role of the loan officers' incentives to collude with the borrower under relationship lending. Therefore, we have assumed that monitoring efforts are contractible, and focused on the aforementioned incentive problem in bank-monitor-borrower hierarchies. If monitoring efforts were non-contractible, then each bank would face an additional *effort incentive constraint*. Under single-bank lending, the maximization problem ($\mathscr{M}$) would be subject to an additional incentive constraint which is given by:

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

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

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

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

Because of the presence of the additional effort incentive constraints in each of the maximization programs, the equilibrium monitoring intensities are lower than those described in Propositions 2 and 4. However, all our results remain qualitatively the same even if monitoring efforts are non-contractible (see Dam and Roy Chowdhury, 2019, for a detailed analysis).

9 Concluding Remarks

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

Our results have important testable implications both with respect to production of soft information and multiple-bank lending. The critical role that loan officers play in producing soft information 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 regularly, the incumbent officer tends to report more accurately because if her successor produces different but verifiable information, then this may hurt her reputation. Our framework with multiple loan officers under the possibility of vertical collusion resembles that of Hertzberg et al. (2010), although we do not consider a dynamic long-term lender-borrower relationship. However, the success of incentive schemes in deterring collusion, either in Hertzberg et al. (2010) or in the present paper, is not based on a folk theorem like argument (e.g. Tirole, 1986). Berger et al. (2005) find that small banks tend to have greater proximity to their borrowers, interact more closely with them, and have more exclusive relationships; some or all of these factors may contribute to bank efficiency (as measured by the exogenous probability of detecting collusion). Thus, if we agree on the fact that that bank efficiency is negatively related to bank size, our result that more efficient banks are able to produce soft information at lower (incentive) costs, regardless of the lending structure, conforms to the findings of Uchida et al. (2012) that, in the context of Japanese SME lending, loan officers at smaller banks produce more soft information.

Appendix

Proof of Proposition 2

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

$$\mathscr{L} = \underbrace{mry(1-s)-1}_{B(m,s)} + \mu_P \underbrace{\left(mrys - \frac{1}{2}cm^2\right)}_{g^P(m,s)} + \mu_N \underbrace{(s-1+\lambda)}_{g^N(m,s)} + \mu_F \underbrace{(1-m)}_{g^F(m,s)},$$

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

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

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

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

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

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

First, we consider the case when (F) binds, i.e., $m = 1$. We argue that, in this case, (NC) must bind. Suppose not, i.e., $\mu_N = 0$. Substituting, $m = 1$ and $\mu_N = 0$ into (3), we obtain that $\mu_P = 1$ because $ry > 0$. Then, (2) reduces to $ry - c = \mu_F > 0$ which contradicts our assumption that $ry \leq c$. Given that $m = 1$ and $s = 1 - \lambda$, the multipliers μ_P, μ_N and μ_F are determined from the two first-order conditions (2) and (3), and the complementary slackness condition associated with the participation constraint, $\mu_P g^P(1, 1-\lambda) = 0$, i.e.,

$$ry\lambda + \mu_P(ry(1-\lambda) - c) - \mu_F = 0, \tag{7}$$

$$(\mu_P - 1)ry + \mu_N = 0, \tag{8}$$

$$\mu_P\left(ry(1-\lambda) - \frac{c}{2}\right) = 0. \tag{9}$$

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

$$ry(1-\lambda) \geq \frac{c}{2} \iff \lambda \leq 1 - \frac{c}{2ry} \equiv \lambda_1^0.$$

We now prove that $\mu_F = ry\lambda \geq 1$. Suppose not, i.e., $\lambda < 1/ry$ which implies $B(1, 1-\lambda) = ry\lambda - 1 < 0$. This is not possible, and hence, $\lambda \geq 1/ry \equiv \lambda_1^{min}$. This threshold is strictly positive because $ry > 0$.

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

$$mry(1-\lambda) = \frac{1}{2}cm^2 \iff m = \frac{2ry(1-\lambda)}{c}.$$

Because (F) is slack, we have

$$m = \frac{2ry(1-\lambda)}{c} < 1 \iff \lambda > \lambda_1^0.$$

Substituting $m = 2ry(1-\lambda)/c$, $s = 1-\lambda$ and $\mu_F = 0$ into (2) and (3), we obtain

$$\mu_P = \frac{\lambda}{1-\lambda} \quad \text{and} \quad \mu_N = \frac{2r^2y^2(1-2\lambda)}{c}.$$

Now, $\mu_N \geq 0$ implies $\lambda \leq \frac{1}{2} \equiv \bar{\lambda}_1$. Finally, consider the sub-case when (NC) does not bind at the optimum, i.e., $\mu_N = 0$. Thus, substituting $\mu_N = \mu_F = 0$ into (2) and (3), we obtain

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

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

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

Proof of Lemma 1

The binding participation constraint of each loan officer yields

$$rys_i = \frac{cm_i^2}{2\pi(m_i, m_j)}. \tag{12}$$

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

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

$$\text{subject to} \quad \frac{cm_i^2}{2\pi(m_i, m_j)} \geq 2ry(1-\lambda). \tag{NC$_i'$}$$

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

$$ry(1-m_j) - cm_i = 0, \tag{BR$_i$}$$

$$ry(1-m_i) - cm_j = 0. \tag{BR$_j$}$$

Note that $m_i'(m_j) = m_j'(m_i) = -ry/c < 0$, and hence, m_i and m_j are strategic substitutes. Now we prove that, when both the no-collusion constraints bind, m_i and m_j are strategic complements for $m_i, m_j \in [0, 1]$. In this case the best reply functions $m_i(m_j)$ and $m_j(m_i)$ solve the following two equations, respectively.

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

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

where $a \equiv 2ry(1-\lambda) \geq 0$. Analyzing the behavior of $m_i(m_j)$ suffices to prove the assertion as the behavior of $m_j(m_i)$ is symmetric. Note first that (BR$_i'$) can be written as

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

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

$$2a \leq c. \tag{13}$$

Solving for m_i from (BR_i') we get

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

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

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

The above expression is positive if and only if $2a \geq c$, which holds from (13). ■

Proof of Proposition 4

The proof is very similar to that of Proposition 2, and hence, we would omit the details. Instead, we will describe the threshold values λ_2^{min}, λ_2^0 and $\bar{\lambda}_2$, and the optimal values of (m_i, m_j), (s_i, s_j) and $\pi(m_i, m_j)$. The equilibrium is symmetric, and hence, $m_i = m_j = m_2$ and $s_i = s_j = s_2$. When, both the no-collusion constraints bind, and neither participation constraint binds, we have $m_2 = \pi = 1$ and $s_2 = 2(1-\lambda)$. In this case, the payoff of each bank is given by $B_2(1, 1, 2(1-\lambda)) = 1 \cdot ry(1 - 2(1-\lambda)) - 1 = ry(2\lambda - 1) - 1$, which must be non-negative yielding

$$\lambda \geq \frac{1}{2}\left(1 + \frac{1}{ry}\right) \equiv \lambda_2^{min}.$$

When $m_2 < 1$, the participation constraints (PC_i) and (PC_j) bind. We have two sub-cases — (a) (NC_i) and (NC_j) both bind, and (b) neither of them binds. First, consider the case when the no-collusion constraints bind. The optimal m_i and m_j are solutions to the system (BR_i') and (BR_j'), which has two symmetric solutions

$$(0, 0) \quad \text{and} \quad \left(\frac{8ry(1-\lambda)}{c + 4ry(1-\lambda)}, \frac{8ry(1-\lambda)}{c + 4ry(1-\lambda)}\right),$$

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 given by the other symmetric solutions, and hence,

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

Using (12), we get the optimal share which is given by:

$$s_i = s_j = 2(1-\lambda).$$

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

$$m_2(\lambda, r) = \frac{8ry(1-\lambda)}{c + 4ry(1-\lambda)} < 1 \quad \Longleftrightarrow \quad \lambda > 1 - \frac{c}{4ry} \equiv \lambda_2^0.$$

Finally, consider when neither (NC_i) nor (NC_j) binds. Because both $m_i(m_j)$ and $m_j(m_i)$ defined by (BR_i) and (BR_j) are linear and downward-sloping, there is a unique solution to the system of equations, which is also symmetric. This is given by:

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

Thus,

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

Using (12), we get the optimal share which is given by:

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

The non-binding no-collusion constraints imply

$$\frac{c}{2(2c+ry)} > 2(1-\lambda) \iff \lambda > \frac{7c+4ry}{4(2c+ry)} \equiv \bar{\lambda}_2.$$

This completes the proof of the proposition. ■

Proof of Proposition 5

We first show that $\bar{\lambda}_2 > \bar{\lambda}_1$ which is equivalent to $3c+2ry > 0$. The last inequality always holds because both ry and c are strictly positive. To be consistent with Figure 4, it is also necessary to show that $\bar{\lambda}_1 < \lambda_2^{min}$. This holds because the inequality is equivalent to $1/ry > 0$.

To show Part (a), note that because $\bar{\lambda}_1 < \lambda_2^{min}$, we have $m_1(\lambda_2^{min}, r) = ry/c$. Thus, $m_2(\lambda_2^{min}, r) = 1 \geq ry/c = m_1(\lambda_2^{min}, r)$. On the other hand, $m_2(\bar{\lambda}_2, r) = ry/(c+ry) < ry/c = m_1(\bar{\lambda}_2, r)$. Because $m_2(\lambda, r)$ is strictly decreasing on $[\lambda_2^0, \bar{\lambda}_2]$, there is a unique $\lambda_m \in (\lambda_2^{min}, \bar{\lambda}_2)$ such that $m_2(\lambda, r) > m_1(\lambda, r)$ if and only if $\lambda < \lambda_m$.

For Part (b), note that $s_2(\lambda_2^{min}, r) = 1 - 1/ry > 1/2 = s_1(\lambda_2^{min}, r)$ because $2 < c/2+1 \leq ry$, and $s_2(\bar{\lambda}_2, r) = (c/2)(2c+ry) < 1/2 = s_1(\bar{\lambda}_2, r) \iff c+yr > 0$. As $s_2(\lambda, r)$ is strictly decreasing on $[\lambda_2^0, \bar{\lambda}_2]$, there exists a unique $\lambda_s \in (\lambda_1^{min}, \bar{\lambda}_2)$ such that $s_2(\lambda, r) > s_1(\lambda, r)$ if and only if $\lambda < \lambda_m$. ■

Proof of Proposition 6

Note that $\pi(\lambda_2^{min}) = 1 > ry/c = m_1(\lambda_2^{min})$. On the other hand,

$$\pi(\bar{\lambda}_2, r) < m_1(\bar{\lambda}_2, r) \iff \frac{ry(2c+ry)}{(c+ry)^2} < \frac{ry}{c} \iff r^2y^2 + cry - c^2 > 0.$$

The above inequality holds for $2ry > (\sqrt{5}-1)c$, i.e. $ry > 0.62c$. ■

Proof of Proposition 7

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

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

Denote the threshold values of bank efficiency and the equilibrium variables with subscript 'mb'. In a symmetric equilibrium, as in Proposition 4, when both no-collusion constraints (NC_i) and (NC_j) bind, the optimal contracts coincide with those under two bank lending. Therefore, we have $\lambda_{mb}^{min} = \lambda_2^{min}$, $\lambda_{mb}^0 = \lambda_2^0$, $m_{mb}(\lambda, r) = m_2(\lambda, r)$, $\pi_{mb}(\lambda, r) = \pi(\lambda, r)$ and $s_{mb}(\lambda, r) = s_2(\lambda, r)$. 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

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

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

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

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

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

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

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

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

Proof of Proposition 8

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

$$\max_{\{m_i, s_i\}} \pi(m_i, m_j)ry(1-s_i)-1, \tag{$\mathscr{M}_i$}$$
$$\text{subject to (PC}_i\text{), (NC}_{ij}\text{), (F}_i\text{).}$$

We analyze a symmetric equilibrium, i.e., $m_i = m_j$ and $s_i = s_j$. As in the proof of Proposition 4, when the feasibility constraints bind, i.e., $m_i = m_j = \pi(m_i, m_j) = 1$, the participation constraints do not bind, and the no-collusion constraints bind implying $s_i = s_j = 1-\lambda$. At these values the equilibrium payoffs of the banks are given by $B_i = B_j = ry\lambda - 1$, the non-negativity of which yields $\lambda_{cp}^{min} = 1/ry = \lambda_1^{min}$. Next, consider the case when (F$_i$) and (F$_j$) are slack, i.e., $m_i = m_j < 1$. In this case, the participation constraints must bind. There are two sub-cases. First, the no-collusion constraints bind at the optimum, i.e., $s_i = s_j = s_{cp}(\lambda) = 1-\lambda$. The binding participation constraints, and the fact that $\pi(m,m) = 1-(1-m)^2$ imply:

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

Given that $c > 0$, we have $m_{cp}(\lambda, r) < m_2(\lambda, r)$ and $\pi_{cp}(\lambda, r) < \pi(\lambda, r)$. Clearly, $s_{cp}(\lambda, r) = 1-\lambda < 2(1-\lambda) = s_2(\lambda, r)$. Finally, in the sub-case when the no-collusion constraints are slack, the equilibrium contracts coincide with those under two-bank lending. The threshold value of bank efficiency $\bar{\lambda}_{cp}$ solves

$$\frac{c}{2(2c+ry)} = 1-\lambda,$$

whereas $\bar{\lambda}_2$ solves

$$\frac{c}{2(2c+ry)} = 2(1-\lambda).$$

Therefore, $\bar{\lambda}_{cp} < \bar{\lambda}_2$, 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. ■

Proof of Proposition 9

The proof is similar to those of Propositions 2 and 4, and hence, many details will be omitted. First, we derive the no-collusion constraint of loan officer i. Given a bribe of b_i, the payoffs from collusion to the firm and loan officer i are respectively given by:

$$F(\hat{y}) = ny - b_i - \lambda nry - (1-\lambda)nr\hat{y}, \quad M_i(\hat{y}) = s_i r\hat{y} + b_i - C(m_i).$$

The payoffs from truthful reporting, on the other hand are given by:

$$F(y) = ny - nry, \quad M_i(y) = s_i ry - C(m_i).$$

Thus, collusion between the firm and loan officer i is not feasible if and only if

$$F(y) + M_i(y) \geq F(\hat{y}) + M_i(\hat{y}) \quad \Longleftrightarrow \quad s_i \geq n(1-\lambda). \tag{NC$_i^n$}$$

To solve for the symmetric equilibrium contracts, i.e., $m_i = m$ and $s_i = s$ for all $i = 1, \ldots, n$, we first consider the case when (F_i) binds for all i. In this case the no-collusion constraints bind, and the participation constraints are slack. Therefore, $m_i = 1$ and $s_i = n(1-\lambda)$ for all i, and $\pi(\mathbf{m}) = 1$. Bank i's equilibrium payoff is given by:

$$B_i(1, n(1-\lambda)) = 1 \cdot ry(1 - n(1-\lambda)) - 1 \quad \text{for all } i = 1, \ldots, n.$$

The above expression is non-negative if

$$\lambda \geq \frac{1}{n}\left(n - 1 + \frac{1}{ry}\right) \equiv \lambda^{min}(n).$$

Next, consider the case when (F_i) does not bind for any i. In this case, (PC_i^n) must bind for all i. The binding participation constraint of loan officer i defines the optimal share as a function of the vector of efforts, i.e. $s_i(\mathbf{m})$. Thus, the maximization problem of bank i boils down to:

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

$$\text{subject to} \quad s_i(\mathbf{m}) \geq n(1-\lambda). \tag{NC_i^n}$$

Note that

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

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

$$ry\pi_i(\mathbf{m}) = cm_i.$$

Thus, using symmetry, the above first-order condition reduces to ($\star$). Denote by $m(n)$ the symmetric equilibrium monitoring effort. We first show that $m'(n) < 0$. Note that the left-hand-side of ($\star$), $h(m,n) \equiv (1-m)^{n-1}$ is strictly decreasing in m with $h(0,n) = 1$ and $h(1,m) = 0$ for all n. Also, $h(m,n)$ is strictly convex in m for all $n > 2$, and linear for $n = 2$ with the slope being -1. Moreover, $h(m,n)$ is strictly decreasing in n, and hence $h(m,n) \leq h(m,2)$ for all $n > 2$. On the other hand, the right-hand-side of ($\star$), $\gamma(r)m$ is a strictly increasing straight line with the slope being $\gamma(r) \geq 1$. Because $m(n)$ is determined by the intersection of $h(m,n)$ and $\gamma(r)m$, the intersection point moves to the left as n grows, which implies that $m'(n) < 0$. Moreover, because $\gamma(r) \geq 1$, the intersection between $h(m,2)$ and $\gamma(r)m$ is always to the left of a point where $m = 0.5$, i.e. $m(2) \leq 0.5$ which implies that $m(n) < 0.5$ for all $n > 2$. The aggregate monitoring intensity is given by $\pi(n) = 1 - (1-m(n))^n = 1 - \gamma(r)m(n)(1-m(n))$. The last equality follows from ($\star$). Thus,

$$\pi'(n) = -\gamma(r)(1 - 2m(n))m'(n) \geq 0$$

because $m(n) \leq 0.5$ for all $n \geq 2$, and $m'(n) < 0$. The equilibrium incentives are determined from the binding participation constraint of each loan officer, i.e.

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

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

$$s(n) > n(1-\lambda) \quad \iff \quad \lambda > 1 - \frac{s(n)}{n} \equiv \bar{\lambda}(n).$$

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

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

$$s^0(\lambda, n) = n(1-\lambda),$$

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

$$[1-(1-m)^n]rys^0(\lambda,n) = \frac{1}{2}cm^2 \iff \underbrace{2[1-(1-m)^n]n(1-\lambda)}_{H(m,n)} = \underbrace{\gamma(r)m^2}_{G(m)}.$$

The solution to the above equation gives the equilibrium monitoring effort. Note that $G(m)$ is strictly increasing and strictly convex in m with $G(0) = 0$ and $G(1) = \gamma(r)$. On the other hand, $H(m,n)$ is strictly increasing and strictly concave in m with $H(0,n) = 0$ for all $n > 0$ and $H(1,n) = 2n(1-\lambda)$. Note that, whenever $m = \pi = 1$, from the binding participation constraint we have $s = c/2ry$. At this optimal share, the no-collusion constraint does not hold, i.e. $c/2ry < n(1-\lambda)$ if and only if $\lambda < 1-(c/2nry) \equiv \lambda^0(n)$. Thus, bank lending is feasible only if $\lambda \geq \lambda^0(n)$ which is equivalent to $H(1,n) = 2n(1-\lambda) \leq \gamma(r) = G(1)$. Thus, the intersection of $H(m,n)$ and $G(m)$ is unique for each n, which gives the symmetric equilibrium monitoring effort $m(\lambda,n)$ which is unique for each n. Moreover, $H(m,n)$ is strictly increasing in n which implies that $m(\lambda,n)$ is strictly increasing for each $\lambda \in [\lambda^0(n), \bar{\lambda}(n)]$. The aggregate monitoring intensity is given by $\pi(\lambda,n) = 1-[1-m(\lambda,n)]^n$. Because $m(\lambda,n)$ is strictly increasing in n, we have that $\pi(\lambda,n)$ is strictly increasing in n for each $\lambda \in [\lambda^0(n), \bar{\lambda}(n)]$. Part (b) of the proposition follows from the fact that both $m(\lambda,n)$ and $s(\lambda,n)$ are strictly increasing in n, and both $m(n)$ and $s(n)$ are strictly decreasing in n. In the proof of Proposition 6 we have shown that $\bar{\lambda}_1 \equiv \bar{\lambda}(1) < \lambda_2^0 \equiv \lambda^0(2)$. But for any n and n' with $n' > n \geq 2$, it is not easy to show that $\bar{\lambda}(n) < \lambda^0(n')$. But the result holds irrespective of whether the last inequality is satisfied or not. Therefore, we have taken $\max\{\bar{\lambda}(n), \lambda^0(n')\}$ as the infimum of both λ_m and λ_s. On the other hand, the fact that both $\pi(\lambda,n)$ and $\pi(n)$ are strictly increasing in n implies part (c). ■

Proof of Proposition 10

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

$$B_1(\lambda,R) = 1\cdot R\cdot\lambda - 1 = R\lambda - 1.$$

Setting the above expression equal to zero, we obtain $R_1(\lambda) = 1/\lambda$. Because, $R_1(\lambda) \leq c$, we have $\lambda \geq 1/c \equiv \lambda_1^{min}$.

Next, consider the case when $m_1(\lambda,R) \leq 1$. We first analyze the sub-case when both the participation and no-collusion constraints bind. Thus, $s_1(\lambda,R) = 1-\lambda$ and $m_1(\lambda,R) = 2R(1-\lambda)/c$, given which, the bank's payoff is given by:

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

Setting the above equal to zero we obtain

$$R_1(\lambda) = \sqrt{\frac{c}{2\lambda(1-\lambda)}}, \quad \text{and} \quad m_1(\lambda) \equiv m_1(\lambda, R_1(\lambda)) = \sqrt{\frac{2(1-\lambda)}{c\lambda}}.$$

The threshold value of bank efficiency, λ_1^0 is determined by solving

$$\frac{1}{\lambda} = \sqrt{\frac{c}{2\lambda(1-\lambda)}} \quad \Longleftrightarrow \quad \lambda_1^0 = \frac{c}{2+c}.$$

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

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

Thus, zero-profit condition yield

$$R_1(\lambda) = \sqrt{2c}, \quad \text{and} \quad m_1(\lambda) \equiv m_1(\lambda, R_1(\lambda)) = \sqrt{\frac{2}{c}}.$$

The other threshold value of bank efficiency, $\bar{\lambda}_1$ is determined by solving

$$\sqrt{\frac{c}{2\lambda(1-\lambda)}} = \sqrt{2c} \quad \Longleftrightarrow \quad \bar{\lambda}_1 = \frac{1}{2}.$$

Note that $c > 2$ implies that $\lambda_1^{min} < \lambda_1^0 < \bar{\lambda}_1$.

Equilibrium loan rates under two-bank lending. Given the loan rates R_i and R_j, in the contracting stage, we consider three situations. First, let the feasibility constraints bind, i.e., $m_i = m_j = 1$, (PC_i^0) and (PC_j^0) are slack, and (NC_i^0) and (NC_j^0) bind. The above implies $\pi(1,1) = 1$ and $s_iR_i = s_jR_j = (1-\lambda)(R_i + R_j)$. Thus, the payoffs of the banks are given by:

$$B_i \equiv 1 \cdot R_i(1-s_i) - 1 = R_i - (1-\lambda)(R_i + R_j) - 1,$$
$$B_j \equiv 1 \cdot R_j(1-s_j) - 1 = R_j - (1-\lambda)(R_i + R_j) - 1.$$

Setting each of the above expressions equal to zero, we obtain $R_i = R_j = 1/(2\lambda - 1)$. Thus, $s_i = s_j = 2(1-\lambda)$. Because, $R_i, R_j \leq c$, we have

$$\frac{1}{2\lambda - 1} \leq c \quad \Longleftrightarrow \quad \lambda \geq \frac{1}{2}\left(1 + \frac{1}{c}\right) \equiv \lambda_2^{min}.$$

Next, consider the case when $m_i, m_j \leq 1$, and both the participation and no-collusion constraints of both loan officers bind, i.e.,

$$\pi(m_i, m_j)R_is_i = \frac{1}{2}cm_i^2 \quad \text{and} \quad \pi(m_i, m_j)R_js_j = \frac{1}{2}cm_j^2,$$
$$R_is_i(\lambda, R_i, R_j) = (1-\lambda)(R_i + R_j) = R_js_j(\lambda, R_i),$$

which together yield

$$cm_i^2 = 2\pi(m_i, m_j)(1-\lambda)(R_i+R_j) = cm_j^2.$$

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

$$m_i(\lambda, R_i, R_j) = m_j(\lambda, R_i, R_j) = \frac{4(1-\lambda)(R_i+R_j)}{c+2(1-\lambda)(R_i+R_j)},$$
$$\text{and } \ \pi(\lambda, R_i, R_j) = \frac{8c(1-\lambda)(R_i+R_j)}{(c+2(1-\lambda)(R_i+R_j))^2}.$$

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

$$B_i(\lambda, R_i, R_j) = \pi(\lambda, R_i, R_j)R_i(1-s_i(\lambda, R_i, R_j)) - 1 = 0,$$
$$B_j(\lambda, R_i, R_j) = \pi(\lambda, R_i, R_j)R_j(1-s_j(\lambda, R_{i,R_j})) - 1 = 0.$$

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

$$R_i(\lambda) = R_j(\lambda) = \frac{c}{4\left(\sqrt{c(1-\lambda)(2\lambda-1)} - c(1-\lambda)\right)},$$
$$R_i(\lambda) = R_j(\lambda) = \frac{c}{4\left(\sqrt{c(1-\lambda)(2\lambda-1)} + c(1-\lambda)\right)}.$$

We set $R_i(\lambda)$ and $R_j(\lambda)$ at the smaller root. For the above roots to be real, we require that $2\lambda - 1 \geq 0$. The threshold value of bank efficiency, λ_2^0 solves

$$\underbrace{\frac{1}{2\lambda-1}}_{g(\lambda)} = \underbrace{\frac{c}{4\left(\sqrt{c(1-\lambda)(2\lambda-1)} + c(1-\lambda)\right)}}_{f(\lambda;c)}.$$

Note that $g(\lambda)$ is strictly decreasing on $[1/2,1]$ with $\lim_{\lambda\to 1/2} g(\lambda) = \infty$ and $g(1) = 1$. On the other hand, $f(\lambda; c)$ is strictly increasing in λ for all $\lambda \in [1/2, 1]$ and for any $c > 2$ with $f(1/2, c) = 1/2$ and $\lim_{\lambda\to 1} f(\lambda; c) = \infty$. Thus, $\lambda_2^0 \in (1/2, 1)$ is unique. The equilibrium individual monitoring efforts and the aggregate monitoring intensity are given by:

$$m_i(\lambda) = m_j(\lambda) = m_2(\lambda) = \frac{2(1-\lambda)}{(1+c)(1-\lambda) + \sqrt{c(1-\lambda)(2\lambda-1)}},$$
$$\pi(\lambda) = 1 - [1-m_2(\lambda)]^2.$$

The equilibrium incentives s_i and s_j are given by $s_2(\lambda) = s_i(\lambda, R_i(\lambda), R_j(\lambda))$.

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

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

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

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

from which we obtain

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

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

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

The above system has (a) two pairs of asymmetric and real solutions, (b) one pair of symmetric, but complex solutions, and (c) two pairs of asymmetric, but complex solutions.[22] The pairs of real solutions are given by:

$$\left(R_i^0 = \frac{c\left(4(1+c)-\sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2},\ R_j^0 = \frac{c\left(4(1+c)+\sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}\right),$$
$$\left(R_i^0 = \frac{c\left(4(1+c)+\sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2},\ R_j^0 = \frac{c\left(4(1+c)-\sqrt{2c(4+2c-c^2)}\right)}{(2+c)^2}\right).$$

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

$$m_i^0 = \frac{2c-\sqrt{2c(4+2c-c^2)}}{c(2+c)} \quad \text{and} \quad m_j^0 = \frac{2c+\sqrt{2c(4+2c-c^2)}}{c(2+c)},$$
$$\pi^0 = \pi(R_i^0, R_j^0) = \frac{2}{c}.$$

The corresponding equilibrium shares of repayment are determined from binding (PC_i^0) and (PC_j^0), which are given by:

$$s_i(\lambda) = \frac{4-\sqrt{2c(4+2c-c^2)}}{4(2+c)}, \quad s_j(\lambda) = \frac{4+\sqrt{2c(4+2c-c^2)}}{4(2+c)}.$$

The threshold value of bank efficiency, $\bar{\lambda}_2$ is determined by solving $R_2(\lambda) = R_j^0$, where

$$R_2(\lambda) = \frac{c}{4\left(\sqrt{c(1-\lambda)(2\lambda-1)} - c(1-\lambda)\right)}.$$

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

It is easy to see that $\bar{\lambda}_2$ is well-defined and strictly less than 1. Also, we have $\lambda_2^{min} < \lambda_2^0 < \bar{\lambda}_2$.

Comparison of the two lending modes. For $2 < c \leq 1+\sqrt{5}$, it is easy to show that $\lambda_1^{min} < \bar{\lambda}_1 < \lambda_2^{min} < \bar{\lambda}_2$. Because $\bar{\lambda}_1 < \lambda_2^{min}$, $R_1(\lambda_2^{min}) = R_1(\bar{\lambda}_1) = \sqrt{2c} < c = R_2(\lambda_2^{min})$. On the other hand, $R_2(\bar{\lambda}_2) = R_j^0 < \sqrt{2c} = R_1(\bar{\lambda}_2)$. As $R_2(\lambda)$ is strictly decreasing on $[\lambda_2^{min}, \bar{\lambda}_2]$ there is a unique $\lambda_R \in (\lambda_2^{min}, \bar{\lambda}_2)$ such that $R_2(\lambda) > R_1(\lambda)$ if and only if $\lambda < \lambda_R$. This proves Part(a) of the proposition. The proof of Part (b) is very similar to that of Propositions 5 and 6 with the exception that, unlike Proposition 6-(b), the aggregate monitoring intensity under two-bank lending never lies above that under single-bank lending because $\sqrt{2/c} > 2/c$ for any $c \in (2, 1+\sqrt{5}]$. ■

References

Berger, Allen, Nathan Miller, Mitchell Petersen, Raghuram Rajan, and Jeremy Stein (2005), "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks." *Journal of Financial Economics*, 76, 237–269.

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.

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.

Dam, Kaniṣka and Alejandro Robinson-Cortés (2018), "Executive Compensation and Competitive Pressure in the Product Market: How Does Firm Entry Shape Managerial Incentives?" Mimeo, CIDE and Caltech.

Dam, Kaniṣka and Prabal Roy Chowdhury (2019), "Loan Officer Incentives under Multiple Bank Lending and Threat of Collusion." CIDE Working paper.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.