Monitoring in Originate-to-Distribute Lending: Reputation versus Skin in the Game*

Andrew Winton
University of Minnesota

Vijay Yerramilli
University of Houston

Abstract

Banks face liquidity and capital pressures that favor selling off loans they originate, but loan sales undermine their monitoring incentives. A bank's loan default history is a noisy measure of its past monitoring choices, which can serve as a reputation mechanism to incentivize current monitoring. In equilibrium, higher reputation banks monitor (weakly) more intensively; if retention is credible, they generally retain less of the loans they originate. Monitoring is harder to sustain in periods with uncommon large spikes in loan demand ("booms"), especially for low-reputation banks, which are more likely to accommodate boom demand and forgo monitoring.

*A previous version of the paper was titled "Lender Moral Hazard and Reputation in Originate-to-Distribute Markets." We thank two anonymous referees, Itay Goldstein (Editor), Viral Acharya (discussant), Andres Almazan, Thomas Chemmanur, Thierry Foucault, Paolo Fulghieri, Thomas Gehrig (discussant), Bruno Gerard, Dirk Hackbarth (discussant), Charles Kahn, Rich Mathews (discussant), Elvira Sojli, Ajay Subramanyam, Zaime Zender (discussant), and seminar participants at the 2012 SFS Finance Cavalcade (University of Virginia), 2012 Paris Spring Corporate Finance Conference, 2012 Western Finance Association meetings (Las Vegas), 2012 European Finance Association meetings (Copenhagen), 2012 Summer Conference at the Indian School of Business, 2012 University of Oxford Reputation Symposium, 1st Oxford Financial Intermediation Theory Conference, Carlos III University, Erasmus University, Federal Reserve Banks of New York and Philadelphia, Georgia State University, London School of Economics, Norwegian Business School in Oslo, Oxford University, Pompeu Fabra University, Rotterdam Business School, Tilburg University, and the University of Minnesota, for their helpful comments or discussions on issues examined in the paper. All remaining errors are our responsibility. Send correspondence to Vijay Yerramilli, 334 Melcher Hall, University of Houston, Houston, TX 77204; telephone: (713) 743-2516. E-mail: vyerramilli@bauer.uh.edu.

In traditional theories of financial intermediation, banks must hold the loans they make so as to maintain their incentives to screen and monitor them, but present-day lenders often sell off the loans that they originate.[1] Although this "originate-to-distribute" (OTD) strategy can improve risk-sharing and the lender's capital or liquidity position, it begs the question of why a lender would continue to monitor loans it sells. In the two decades leading up to the financial crisis of 2007-2009, the typical response of practitioners to such concerns was that the lender's concern for its reputation would give it incentive to monitor even after it had laid off its exposure to credit risk. Subsequent revelations of poor credit underwriting even by highly-reputable institutions cast doubt on this view.[2] Nevertheless, as we discuss in our review of the literature, there has been little theoretical work on the question of when and to what extent reputation concerns will in fact cause lenders to monitor the loans they sell.

We examine this question in a model of repeated lending where all participants are rational and banks face capital or liquidity costs that favor selling off loans that they make, all else equal. Because the history of defaults on a bank's loans is a noisy indication of whether it has monitored its loans in the past, this history can serve as a reputation that may allow the bank to commit to monitor even if it sells off all or part of the loans it originates. Note that, although banks with fewer past defaults have higher reputations and banks with more past defaults have lower reputations, reputation in our model does not reflect any *learning* about the bank's hidden characteristics or "type"; indeed, the bank has no innate type in our model. Instead, reputation is a sanctioning mechanism that operates

[1] The idea of delegated monitoring is that banks monitor and enforce loan terms on borrowers on behalf of the bank's own depositors and shareholders. See Leland and Pyle (1977), Diamond (1984) and Holmström and Tirole (1997) for traditional theories of delegated monitoring, and Boyd and Prescott (1986) and Ramakrishnan and Thakor (1984) for models of delegated screening. Pennacchi (1998) and Gorton and Pennacchi (1995) are among the first to highlight the trend towards selling off originated loans.

[2] Keys, Mukherjee, Seru, and Vig (2010) and Purnanandam (2011) provide empirical evidence that securitization led to lax screening in the mortgage market. Piskorski, Seru, and Vig (2010) and Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff (2011) show that securitization affects servicing of loans, in particular, renegotiation of delinquent loans.

purely through the threat that poor *current* performance will be followed with lower *future* profits.

Exactly how this works depends on whether the bank can credibly commit to retain any part of loans it originates. In our baseline setting, we assume this is impossible: that is, even if the bank says it *will* retain some of the loan, it can turn around and anonymously sell off the remaining piece at the prevailing price. This implies that investors will expect that the bank will *not* retain any part of its loans in equilibrium. Nevertheless, equilibrium monitoring may still be possible: if the bank is expected to monitor more, its loans can be sold for a higher price; if it then fails to monitor, there is a greater chance the loans will default, hurting its reputation and the price it can get for its loans next period. If the increase in future expected profits from more intensive monitoring is sufficiently high, the bank prefers to monitor intensively now so as to maximize the chance that its reputation will stay high in the future.

However, in some cases it may be possible for the bank to commit to retain a stake in loans it originates. In the case of corporate loans, the loan may have a no-sale provision; if the loan is syndicated, the lead arranger (which is responsible for structuring and monitoring the borrower) holds a fraction of the loan initially, with other banks holding the rest. Alternatively, market makers in the secondary loan market may be able to observe the bank's loan sales, preventing it from subsequently unloading its share of the loan at a price that incorrectly assumes the bank is not doing this.

Such credible commitments have a big impact on the nature of reputation equilibria. Intuitively, the bank can now trade off retention (which gives it a direct share of the benefits from monitoring a given borrower, but uses up valuable risk-bearing capital or liquidity reserves) against reputation incentives. Moreover, because the bank's retention choice is observable, the bank can offset the negative reputation effect of past defaults by retaining more in the present. Under simple and plausible beliefs, there are only two possible equilibria.

In the first ("high monitoring") equilibrium, banks always monitor with full intensity, but low-reputation banks retain strictly more of the loan than high-reputation banks do. In the second ("low monitoring") equilibrium, high-reputation banks monitor fully and retain a positive fraction of their loans, whereas low-reputation banks do not monitor at all and sell off all of their loans.

When it exists, the high monitoring equilibrium dominates the low monitoring equilibrium, but it may fail to exist if bank capital constraints are relatively loose, monitored loans have a high chance of default, or loan losses given default are high. Even if the high monitoring equilibrium does not exist, the low monitoring equilibrium may not be feasible; its existence requires that a number of constraints be met, and these are often mutually exclusive. In this last case, no equilibrium with reputation-based monitoring is feasible.

Next, we consider what happens when loan demand varies stochastically over time. Our primary focus is on the case where periods of high demand ("booms") are relatively infrequent compared with periods of low ("normal") demand. Once again, equilibrium depends critically on whether or not the bank can commit to retaining all or part of the loan.

If the bank cannot commit to retaining part of its loans, we show that whenever the bank fully accommodates high levels of loan demand, it will not monitor. This happens because, when the bank handles normal levels of loan demand, its incentive compatibility constraint for monitoring holds with equality; that is, the bank's incentive to shirk and save on monitoring costs is exactly offset by the resulting damage to the bank's reputation and future profits. Accommodating loan demand in a boom means a larger loan volume now, increasing immediate effort savings from shirking (the bank can shirk on more loans); however, the reputation benefit is not commensurately higher because average loan volumes are likely to be lower in the future. Similarly, in booms, low-reputation banks always lend weakly more and monitor weakly less than high-reputation banks do: for any level of possible boom-time loan demand, the loss of reputation from shirking matters less for low-reputation banks than for high-reputation banks, and so low-reputation banks have less to lose from accommodating loan demand and shirking. Nevertheless, if loan demand in booms is high enough relative to demand in normal times, then even high-reputation banks will fully accommodate loan demand in booms.

By contrast, if the bank can commit to retain all or part of its loans, it may be possible to support an equilibrium like the "high monitoring" one described for the case of constant loan demand. In this "high monitoring, high lending" equilibrium, high-reputation banks retain less of their loans than low-reputation banks, and both types of banks retain a higher fraction of their loans in booms than in normal times. The conditions that guarantee such an outcome are almost identical to those that guarantee the existence of the high monitoring equilibrium in the case of constant loan demand. Intuitively, these conditions guarantee that the cost of retaining enough of the loan to fully commit to monitoring is low relative to the costs of shirking and getting lower future rents. However, the feasibility conditions fail if bank capital or liquidity needs are relatively low, monitored loans have a high chance of default, or loan losses given default are high.

Our results have a rich set of empirical implications. As we have noted, when originators can commit to retain part of their loans, the dominant equilibrium is one in which banks that suffer more defaults subsequently retain more of the loans they make. This is consistent with the empirical evidence in Gopalan, Nanda, and Yerramilli (2011), who find that a lead arranger of syndicated loans that experiences large defaults is likely to retain a larger fraction of the loans that it underwrites in the subsequent year. However, when loans are sufficiently risky (in terms of the likelihood of default or losses given default), banks with relatively many past defaults retain nothing and do not monitor at all. In settings where banks cannot credibly commit to retain part of their loans, an uncommon boom in loan demand can lead to more lending and less monitoring by banks, especially those with low reputations. As we discuss in our conclusion, this is broadly consistent with some of the facts leading up to the Financial Crisis of 2008-2009. However, there is one important proviso: in our model, investors are not fooled and *expect* low-reputation banks to shirk on high-risk loans in booms, whereas many market participants in the boom preceding the crisis seem to have been more credulous. As a result, our work suggests that a fully rational model of OTD lending with reputation concerns cannot fully explain the pattern of booms and busts that the crisis exemplifies.

Our work also has implications for "skin-in-the-game" rules that force banks to retain a minimum stake in their loans. If such regulations can (albeit at a compliance cost) make loan retention credible, they will facilitate a shift from mixed-strategy equilibria to pure-strategy equilibria. Given that mixed-strategies may be less stable (since, by definition, the bank is indifferent between the equilibrium monitoring choice and any other), making retention credible could improve stability. In addition, if monitoring has positive externalities for other stakeholders in the firm, increasing low-reputation banks' monitoring level by shifting from a mixed-strategy equilibrium to a high-monitoring equilibrium would help these stakeholders without hurting the banks. On the other hand, uniform skin-in-the-game rules might be redundant or worse if banks can already credibly retain their loans, forcing higher-than-needed retention in some cases and not binding in others.

The rest of the paper is organized as follows. We review related literature in Section 1. We describe our baseline model in Section 2, and characterize the equilibrium in Section 3. In Section 4 we allow the bank to retain a portion of the loan on its books, and examine how the retention decision varies with the bank's reputation. We introduce stochastic loan demand in Section 5 and examine how this affects monitoring and lending volume decisions. We conclude in Section 6 with a discussion of the empirical and policy implications of our analysis.

1. Literature Review

While there have been many theoretical papers on reputation in finance, very few have focused on the reputation concerns of a bank that sells off all or part of the loans it originates. Early papers on loan sales assumed that the bank would retain a share of its loan sufficient to ensure monitoring incentives, either through explicit contractual commitments (Pennacchi 1998) or implicit commitments linked to reputation concerns (Gorton and Pennacchi 1995). More recently, Hartman-Glaser (2017) explicitly examines the impact of reputation on a bank's decision on retaining part of the loans it originates. We discuss Hartman-Glaser (2017) in more detail below, but one of the biggest differences between his work and ours is the way in which we model reputation.

Broadly speaking, there are two ways of capturing reputation in economic models. As we have already explained, our model is one of pure moral hazard where reputation is a sanctioning mechanism, encouraging good efforts in the present by threatening future punishment for poor current performance. This falls into the economics literature that begins with papers such as Klein and Leffler (1981), Green and Porter (1984), and Abreu (1986), in which agents are assumed to be completely strategic in their behavior. The other stream of reputation models builds on the work of Kreps and Wilson (1982) and Milgrom and Roberts (1982), in which some agents' behavior depends only on strategic considerations while other agents' behavior is completely determined by their (unobservable) innate type. Hartman-Glaser (2017) is an example of this second stream of models.

Within the group of finance papers that use the first (pure moral hazard) type of reputation model, none have examined how reputation and loan sales interact with monitoring incentives per se. A number of recent papers focus on the reputation of credit rating agencies rather than lenders; since these agencies do not invest in the securities they rate, these models do not speak to the interplay between retention and capital or liquidity concerns that bank lenders face. At a more technical level, these other papers either assume that shirking is detected with certainty and that the cost of damaging one's reputation when caught shirking is either exogenous (e.g., Bolton, Freixas, and Shapiro 2007, 2012) or takes the form of grim-trigger strategies from customers (e.g., Bar-Isaac and Shapiro 2013). We not only endogenize the value of reputation but also assume more realistically that shirking cannot be detected with certainty; moreover, we allow for less onerous responses than the grim-trigger punishment.[3] It follows that our model does a better job of capturing real-world outcomes, where defaults (which are at best a noisy measure of shirking) rarely lead to lenders being permanently excluded from adding value.[4]

As noted above, among reputation models where some agents are strategic and others behave in an exogenously fixed fashion, only Hartman-Glaser (2017) addresses moral hazard in an originate-to-distribute lending setting.[5] In his model, a lender originates loans and then learns whether the loan will be good or bad. The lender can then choose how much of the loan to retain and how much to sell. In the absence of reputation concerns, the strategic lender would perfectly signal loan quality by retaining more when the loan is better, which is the standard signaling result. However, because there is a chance that the lender may be an exogenously honest type that always truthfully reveals loan quality, a strategic lender may benefit by pursuing an honest (or sometimes honest) revelation strategy in the short

[3] In a grim-trigger response, customers faced with shirking respond by assuming that the shirker will always shirk in the future. This means that an apparent misstep (which may not even be actual shirking in a model where shirking is only visible with noise) leads to permanent exclusion, which is not realistic. Moreover, it is often optimal to impose this grim-trigger response with a probability less than one (cf. Green and Porter 1984; Abreu 1986), but this requires that many diffuse market participants coordinate on a randomized punishment, which again strains credibility.

[4] Another banking paper that employs a pure moral hazard model of reputation with grim-trigger strategies is Dinc (2000), although his focus is very different from ours. Dinc examines whether a bank can commit to rescue distressed entrepreneurs in order to preserve its reputation for being a relationship lender, and how this commitment varies with the level of competition from other arm's length and relationship lenders.

[5] A recent reputation paper that features both hidden types and moral hazard without assuming the existence of an "honest" type is Chen, Morrison, and Wilhelm, Jr. (2015). In their model of investment banking, the threat of punishment by counterparties mitigates the bank's moral hazard only when it has a high type reputation, which makes it worthwhile for banks to build their type reputation.

run. If discount rates are sufficiently low, pure truth-telling may be feasible; otherwise, strategic lenders mix between honestly revealing bad loans and misrepresenting them as good (and retaining more of the loan to aid in the misrepresentation). Generally, the higher the likelihood that the lender is of the innately honest type, the less the lender has to retain when it claims a loan is good rather than bad; however, the strategic lender becomes more and more likely to misreport bad loans as good.

Hartman-Glaser's model differs from ours in several respects. In addition to his assumption that some lenders are innately honest and others are strategic, he assumes that lying is always detected (bad loans are perfectly revealed as such after they have been sold). This means that strategic types are slowly revealed over time, after which they must always resort to the pure signaling strategy that holds in the absence of reputation. By contrast, in our model reputations can deteriorate but then recover. Moreover, in his model, the lender always retains less of the asset when she has a higher reputation; in our model, high-reputation lenders may either retain less or more, depending on whether the high-monitoring equilibrium is feasible. In his model, strategic lenders lie more often as reputation improves, whereas in our model, higher reputation leads to weakly higher monitoring. Finally, unlike us, Hartman-Glaser does not examine the impact of stochastically varying levels of loan demand.

2. Baseline Model: Assumptions and Framework

Consider a monopolist long-run lender ("bank") that exists for an infinite number of discrete periods (denoted $t = 0, 1, \ldots$). In each period, it faces a new borrower and a new set of secondary loan market investors who only exist for one period. All agents are risk neutral. Let δ denote the per-period discount factor. The bank's objective is to maximize the discounted value of its expected future payoffs.

At the beginning of each period, a borrower obtains a loan of one unit from the bank to fund its project. By the end of the period, the project either succeeds, yielding X, or fails, in which case the borrower defaults on the loan and the bank seizes the collateral value C, where $C < 1 < X$. The project's success or failure is verifiable, so default occurs if and only if the project fails. Let $R \leq X$ denote the *endogenous* loan repayment amount if the project succeeds. Thus, $R - C$ is the risky component of the loan that the bank obtains only if the project succeeds. We describe below how R is determined in equilibrium. Figure 1 illustrates the stage game in each period.

The bank can improve project outcomes by monitoring borrowers at a cost of $m > 0$. Monitoring allows the bank to better assess the "state" of the project so that it can liquidate bad projects early and continue good projects. The project succeeds with probability p if the bank does not monitor, and with probability $p + \Delta$ if it does, where $\Delta > 0$ denotes the impact of monitoring.[6] The bank's monitoring effort is unobservable, and cannot be contracted upon. We refer to $1 - p$ as the "baseline default probability" because it denotes the probability of default in the absence of any monitoring.

The borrower will undertake the project only if its expected payoff from the project exceeds the value of its outside option, $u \geq 0$. Let $\bar{q}$ denote the borrower's conjecture regarding the probability with which the bank monitors ("monitoring intensity"). Therefore, the borrower's expected payoff from undertaking the project is $(p+\bar{q}\Delta)(X-R)$. As the bank is a monopolist, it will set the loan repayment at the lowest value at which the borrower is indifferent between undertaking the project and pursuing the outside option. Letting $R(\bar{q})$

[6] In Section IA.1 of the Internet Appendix, we show that our reduced-form model of monitoring is consistent with a more detailed structure in which project cash flows depend on the underlying state of the project, "good" or "bad", such that it is optimal to liquidate the project in the bad state and to allow it to continue operating in the good state. Monitoring allows the bank to observe the state of the project perfectly at the intermediate date, thus allowing it to make an informed continuation vs. liquidation decision. On the other hand, if it doesn't monitor, it only observes a noisy signal on the state of the project, which causes it to wrongly liquidate a good project with positive probability.

denote this indifference value, it must satisfy

$$(p + \bar{q}\Delta)(X - R(\bar{q})) = u. \tag{1}$$

We assume that the bank has an incentive to raise immediate cash by selling the loan in the secondary loan market. Formally, we assume that the bank values immediate cash at $1 + \beta$ per dollar for some $\beta > 0$, whereas if it waits to collect loan payments, it only values those payments at 1 per dollar. The benefit of immediate cash could reflect explicit liquidity needs, but it could also be the shadow cost of binding minimum capital requirements: if the bank retains a (risky) loan, it must raise costly equity capital, whereas it will not face this cost if it sells the loan (see Dewatripont and Tirole 1995; DeMarzo and Duffie 1999; Parlour and Winton 2013).

Given the belief $\bar{q}$ regarding the bank's monitoring choice, the price of the loan in the secondary market is

$$\begin{aligned} P(\bar{q}) &= (p + \bar{q}\Delta)(R(\bar{q}) - C) + C \\ &= (p + \bar{q}\Delta)(X - C) + C - u, \end{aligned} \tag{2}$$

where the second equation follows from equation (1). For simplicity, we begin by assuming that the bank cannot credibly commit to hold a fraction of the loan, either because borrowers and investors cannot observe whether the bank has sold the loan or not, or else because they can observe this only after a significant delay. We relax this assumption in Section 4.

We impose some parametric restrictions to focus on situations of economic interest, and to make the model tractable.

Assumption 1: $0 < \Delta < 1 - p$; *monitoring lowers but does not eliminate the probability of default.*

Because monitoring does not completely eliminate the possibility of default, defaults are a noisy signal of whether the bank has monitored or not. As a result, no equilibrium can support full monitoring indefinitely: defaults eventually occur, damaging bank reputation. A decrease in the baseline probability of default $1 - p$ lowers the probability that the loan defaults by bad luck even when the bank monitors.

Assumption 2: $\Delta(X - C) > m$*; monitoring is socially optimal.*

Monitoring improves the probability of success by Δ, so the incremental expected cash flow from monitoring is $\Delta (X - C)$. If this increase exceeds the cost of monitoring m, then monitoring is socially optimal.

Assumption 3: $p(X - C) + C \geq 1 + u$*; the borrower and the bank more than break even on the project even if the bank does not monitor.*

Assumption 3 implies that the secondary loan price $P(\bar{q}) \geq 1$ even if market participants do not believe that the bank will not monitor the loan (i.e., $\bar{q} = 0$). This assumption is not necessary, but it simplifies analysis by ensuring that the bank always prefers to participate in the loan market even if the borrower and investors believe that it will not monitor at all.

3. Equilibrium in the Baseline Model

Because monitoring cannot be contracted upon, it is clear that if the bank existed for only one period, it would not have any incentive to monitor the borrower once it had sold the loan. Anticipating this, the investors would then price the loan at $p(X - C) + C$. However, this need not to be true for a long-lived bank. As long as borrowers and investors can observe the performance history—default versus no default—of the loans made by the bank in previous periods, these market participants can condition their beliefs about how intensively the bank will monitor in the current period based on its past performance. In other words, the bank's past performance may affect its current reputation, and consequently the prices it can charge in the secondary loan market. In this section, we examine whether and to what extent such a reputation mechanism can incentivize the bank to monitor the borrower. Again, this reputation mechanism operates purely through the threat of future punishment for poor performance; there is no learning about the bank per se because the bank has no innate type in our model.

3.1. Definition of Equilibrium

As in any infinitely repeated game, there are many potential equilibria. We examine *perfect public equilibria* (PPE); that is, equilibria in which players choose their current strategies based only on past public signals ("public strategies") and not on their past (unobserved) actions (see Mailath and Samuelson 2006, p. 231).

DEFINITION: Let h denote the past history of defaults on the bank's loans. A perfect public equilibrium consists of the bank's monitoring strategy $q\left(h\right)$ and investor beliefs $\overline{q}(h)$, such that the following conditions hold:

1. Given investor beliefs $\overline{q}(h)$, the bank's monitoring choice $q(h)$ maximizes

$$-mq + \delta(p + \Delta q)(V(h|0) - V(h|1)) + \delta V(h|1),$$

   where $h|0$ and $h|1$ refer to states that are obtained if history h is followed by no default and default, respectively, in the current period, and V denotes the expected discounted value of the bank's profits.

2. The investor beliefs agree with the bank's equilibrium monitoring choice: $\overline{q}(h) = q(h)$.

For tractability, we restrict attention to PPE in which the only element of the bank's public history h that matters for bank actions and investor beliefs is whether the bank's most recent loan defaulted or not, which we denote using the indicator variable for default,

$d \in \{0, 1\}$. It is important to emphasize that this restriction entails almost no loss of generality, because in a pure moral hazard model such as ours, the history h does not contain any information regarding the bank's innate type or motives. We refer to $d \in \{0, 1\}$ as the bank's reputation (see Dellarocas 2005). Thus, the bank may be in one of two possible reputation states: "high" which corresponds to $d = 0$ or "low" which corresponds to $d = 1$. Moreover, the bank's reputation in the next period only depends on whether its current-period loan defaults or not (i.e., $h|0$ is equivalent to $d = 0$ and $h|1$ is equivalent to $d = 1$), which greatly simplifies the analysis.[7]

For expositional convenience, we use the term "high-reputation bank" ("low-reputation bank") to denote that the bank is in the high (low) reputation state. Throughout our analysis, we focus on equilibria in which borrowers and investors hold the highest beliefs about the bank's monitoring intensity that are consistent with such monitoring being incentive compatible. Also, although we allow the bank to use randomized strategies, we assume that the many investors in the market cannot coordinate on such strategies. This rules out equilibria which require investors to coordinate on "grim-trigger" punishments of the sort found in Green and Porter (1984) and Abreu (1986). We discuss the features and drawbacks of such equilibria in Section IA.3 of the Internet Appendix.

3.2. Monitoring Incentives and Bank Value

Ignoring the current period surplus from selling the loan (which is sunk by the time the bank chooses whether to monitor), the bank's expected payoff if it monitors is $V_{mon} \equiv -m + \delta \left[(p + \Delta) V(0) + (1 - p - \Delta) V(1) \right]$, and its expected payoff if it shirks on

[7] More generally, borrowers and investors could condition their beliefs about the bank's monitoring intensity based on the number of defaults $d(N)$ that the bank has sustained in the previous N periods (see Dellarocas 2005). Then, the bank's reputation $d(N) \in \{0, 1, ..., N\}$ could be in $N + 1$ possible states. For simplicity, we restrict the analysis in the paper to the case of $N = 1$. We discuss equilibria with $N > 1$ in Section IA.2 of the Internet Appendix, where we show that restricting attention to equilibria with $N = 1$ entails no loss of generality.

monitoring is $V_{shirk} \equiv \delta \left[pV(0) + (1-p)V(1) \right]$. Monitoring is incentive compatible if and only if $V_{mon} \geq V_{shirk}$. Upon inspection, it is clear that the bank faces the following trade-off: monitoring costs m, but it increases the probability of the bank being in the high reputation state next period by Δ, which is worth $\delta\Delta(V(0) - V(1))$ in present value terms. It follows that the incentive compatibility condition $V_{mon} \geq V_{shirk}$ is equivalent to

$$\Lambda \equiv V(0) - V(1) \geq \frac{m}{\delta\Delta}, \tag{3}$$

where Λ denotes the incremental discounted value of a high reputation over that of a low reputation.

The bank's current period surplus from selling the loan is $S(d) = (1+\beta) \cdot (P(\overline{q}(d)) - 1)$, where $\overline{q}(d)$ denotes the market's conjecture of the bank's monitoring. After substituting for $P(\overline{q})$ using equation (2), the bank's current period surplus can be written as

$$S(d) = A\overline{q}(d) + B, \tag{4}$$

where

$$A \equiv \Delta(1+\beta)(X - C), \tag{5}$$

and

$$B \equiv (1+\beta)(p(X-C) + C - 1 - u). \tag{6}$$

Here, B is the bank's "base level" surplus created by an unmonitored loan, and A is the additional surplus (gross of costs) created by monitoring the loan with probability 1. As $A > 0$, it follows that $S(d)$ is increasing in $\overline{q}(d)$. Moreover, since $B \geq 0$ by Assumption 3, it follows that $S(d) \geq 0$ for all d.

We have the following expression for the bank's value, $V(d)$, in equilibrium:

$$V(d) = A\overline{q}(d) + B - mq(d) + \delta[(p + \Delta q(d)) \cdot \Lambda + V(1)] \tag{7}$$

The bank's value has two components: its gross current period surplus, $A\overline{q}(d) + B$, which depends on the market's belief about the bank's monitoring, $\overline{q}(d)$, and the present value of its expected value next period, $\delta[(p + \Delta q(d))\Lambda + V(1)]$, less the cost of monitoring, $mq(d)$, which depend on the bank's actual monitoring in the current period, $q(d)$. The expression for expected value next period is obtained by noting that the bank will be in the high reputation state next period with probability $p + \Delta q(d)$.

Of course, in equilibrium, the market's conjecture is correct (i.e., $q(d) = \overline{q}(d)$), and the bank's monitoring effort, $q(d)$, satisfies the incentive compatibility condition (3) for all d. Formally, $q(d) = 0$ if condition (3) is violated, $q(d) \geq 0$ if condition (3) holds weakly, and $q(d) = 1$ if condition (3) holds strictly.

We have the following result. (Detailed proofs of all results are in the Appendix.)

Lemma 1 *In any equilibrium in which the bank monitors the loan with positive probability because of concern for its reputation ("reputational monitoring equilibrium"), the bank's incentive compatibility constraint (3) binds:* $\Lambda = \frac{m}{\delta\Delta}$*. The probability of monitoring is strictly higher if there was no default last period than if there was a default last period:* $q(0) > q(1)$*.*

Suppose the bank's incentive compatibility constraint holds strictly; i.e., $\Lambda > \frac{m}{\delta\Delta}$. Then the bank will strictly prefer to monitor in both the high- and low-reputation states, so that $q(0) = q(1) = 1$. But then it follows from the Bellman equation (7) that $\Lambda = V(0) - V(1) = 0$, violating the incentive compatibility condition. Therefore, we must have $\Lambda = \frac{m}{\delta\Delta} > 0$ in any reputational monitoring equilibrium. Making this substitution in the Bellman equation and simplifying, it follows that $\Lambda = [q(0) - q(1)]\,A$, and so $q(0) > q(1)$ in such an equilibrium. Combined with equation (1), an immediate implication of Lemma

1 is that $R(0) > R(1)$; the promised loan repayment is higher when the bank is in the high reputation state.

3.3. The "Full Monitoring" Equilibrium

Define

$$V^* \equiv (1-\delta)^{-1} \cdot \left(A + B - \frac{m(1-p)}{\Delta} \right) \tag{8}$$

We show in Section IA.6 of the Internet Appendix that V^* is the maximum bank value attainable under any PPE in the baseline model.

Proposition 1 *Any reputational monitoring equilibrium is feasible if, and only if,*

$$m \leq \delta \Delta A. \tag{9}$$

If Condition (9) is satisfied, then there exists a "full monitoring" equilibrium in which the bank always monitors the loan in the high-reputation state ($q(0) = 1$), but monitors with a strictly lower probability $q(1) = \hat{q} = 1 - \frac{m}{\delta \Delta A}$ in the low-reputation state. Under this equilibrium, the bank's value function is given by $V_{FM}(0) = V^*$ and $V_{FM}(1) = V^* - \frac{m}{\delta \Delta}$.

Lemma 1 implies that in any reputational monitoring equilibrium, the monitoring intensities $q(0)$ and $q(1)$ must satisfy the condition $q(1) = q(0) - \frac{m}{\delta \Delta A}$. For the equilibrium to be well defined, it is necessary that $q(1) \geq 0$. Substituting $q(0) \leq 1$ yields the necessity of the feasibility condition (9). Sufficiency follows by noting that if condition (9) is satisfied, then it is feasible to support an equilibrium in which $q(0) = 1$ and $q(1) = \hat{q} < 1$. We refer to this as the "full monitoring" equilibrium because it supports full monitoring by the bank in the high reputation state. It is easily verified that the feasibility condition (9) is more likely to hold as the monitoring cost m is lower, the discount factor δ is higher, the value of liquidity β is higher, the impact of monitoring Δ is higher, or project risk $X - C$ is higher.

Substituting $q(0) = 1$, $\Lambda = \frac{m}{\delta\Delta}$, and $V(1) = V(0) - \frac{m}{\delta\Delta}$ in equation (7) and then solving the resultant expression for $V(0)$ yields $V_{FM}(0) = V^*$; the expression for $V_{FM}(1)$ follows by incentive compatibility. It is easily verified that both $V_{FM}(0)$ and $V_{FM}(1)$ increase as the impact of monitoring, Δ, increases, and decrease as the base default probability, $1 - p$, increases.

A key feature of our model is that default is a noisy signal of bank monitoring, because monitoring does not completely eliminate the possibility of default. If instead monitoring completely eliminated the possibility of default (i.e., $p + \Delta = 1$), the value function would be $(1-\delta)^{-1}(A+B-m)$. In our setting, it is lower because even if the bank monitors, there may be a default due to bad luck; thus, defaults eventually occur, damaging bank reputation.

4. Reputation and Loan Retention

In this section, we depart from the base model and consider an alternative scenario in which the bank can credibly commit to retain a fraction $\alpha \in [0, 1]$ of the loan on its books. Consistent with how securitization operates in the real world, we assume that the bank chooses α *after* it sets the loan rate R with the borrower. We begin with a general definition of a perfect public equilibrium in this setting, where the bank's public history h now includes information on both its past loan outcomes and past retention decisions.

Definition: In this setting, a perfect public equilibrium is a set of bank strategies $\{\alpha(h), q(h, \alpha)\}$ and investor beliefs $\overline{q}(h, \alpha)$ such that the following conditions hold:

1. Given a loan rate R, retention choice α, and investor beliefs $\overline{q}(h, \alpha)$, the bank's monitoring choice $q(h, \alpha)$ maximizes

$$-mq + \alpha[(p + \Delta q)(R - C) + C] + \delta(p + \Delta q)(V(h|\alpha, 0) - V(h|\alpha, 1)) + \delta V(h|\alpha, 1),$$

where $h|\alpha, 0$ refers to the history consisting of h followed in the current period by retention α and no default, and $h|\alpha, 1$ refers to the history consisting of h followed in the current period by retention α and default.

2. Given investors' belief $\overline{q}(h, \alpha)$, and loan rate R, the bank chooses its retention $\alpha(h)$ to maximize

$$\begin{aligned} V(h,\alpha) &= (1+\beta)(1-\alpha)\cdot\overline{P}(\overline{q}(h,\alpha),R) + \alpha\cdot P(q(h,\alpha),R) - (1+\beta) - mq(h,\alpha) \\ &\quad + \delta(p+\Delta q(h,\alpha))(V(h|\alpha,0) - V(h|\alpha,1)) + \delta V(h|\alpha,1), \end{aligned}$$

where $\overline{P}(\overline{q}(h,\alpha),R) = (p+\Delta\overline{q}(h,\alpha))(R-C)+C$ denotes the price of the loan based on investors' belief $\overline{q}(h,\alpha)$, and $P(q(h,\alpha),R) = (p+\Delta q(h,\alpha))(R-C)+C$ denotes the loan's realized value based on the bank's actual monitoring $q(h,\alpha)$.

3. For any current retention choice α, investor beliefs agree with the bank's equilibrium monitoring choice: $\overline{q}(h,\alpha) = q(h,\alpha)$. Moreover, the equilibrium loan rate is given by

$$R(q(h,\alpha(h))) = X - \frac{u}{p+\Delta q(h,\alpha(h))}.$$

Note that monitoring may be sustained without any reputation considerations if α is sufficiently high so that $\alpha\Delta(R-C) \geq m$. Define

$$\alpha_{pr} \equiv \frac{m}{\Delta}\left(X - C - \frac{u}{p+\Delta}\right)^{-1} \tag{10}$$

to denote the level of retention at which monitoring is incentive compatible for the bank in the absence of reputation considerations if the loan rate R is set under the belief that the bank will monitor the loan; i.e., $R = R(1) = X - \frac{u}{p+\Delta}$ (the subscript "pr" stands for "pure retention"). However, given that the bank values immediate cash, it may not want to hold

α_{pr} fraction of the loan even if that supports full monitoring, relying instead on reputation to maintain its monitoring incentives.

Define

$$\bar{\alpha} \equiv \frac{m}{\Delta}\left(X - C - \frac{u}{p}\right)^{-1} \tag{11}$$

to denote the level of retention at which monitoring is incentive compatible for the bank even if the loan rate R is set under the belief that the bank will not monitor the loan; i.e., $R = R(0) = X - \frac{u}{p}$. Hence, it follows that $\overline{q}(h, \alpha) = 1$ if $\alpha \geq \bar{\alpha}$. Note that $\bar{\alpha} > \alpha_{pr}$.

Because α is a continuous variable, there are a potentially infinite number of reputation states that the bank may be in even if we restrict history dependence to the most recent period's default outcome and retention α_{t-1}. For tractability, we restrict attention to PPE in which bank actions and investor beliefs depend only on the previous period's default outcome unless the bank is perceived to have "cheated" on its previous period's retention. Since we are interested in equilibria that support bank monitoring, it is sensible to define cheating as when the bank deviates and retains less than its equilibrium retention amount in a state when it is expected to monitor with positive probability; i.e., if the bank deviates to an $\alpha < \alpha(d)$ in a state where $q(d) > 0$.

Hence, at any t, the bank may be in one of four possible reputations states, denoted $d_t \in \{0, 1, 0_{low}, 1_{low}\}$, which are defined as follows: if the bank did not cheat on its loan retention in the previous period, the state is $d = 0$ if that loan did not default and $d = 1$ if it did default; if the bank did cheat on its loan retention in the previous period, the state is $d = 0_{low}$ if that loan did not default and $d = 1_{low}$ if it did default. As before, let $V(d)$ denote the expected discounted value of the bank's profits in equilibrium in reputation state d, and let $\Lambda \equiv V(0) - V(1)$ denote the incremental value of high reputation.

We note that $d = 0_{low}$ and $d = 1_{low}$ are "off-equilibrium" states, and refer to these collectively as d_{low}. It is without loss of generality to set $\overline{q}(0_{low}, \alpha) = \overline{q}(1_{low}, \alpha) = 0$ if

$\alpha < \bar{\alpha}$, and $\overline{q}(0_{low}, \alpha) = \overline{q}(1_{low}, \alpha) = 1$ if $\alpha \geq \bar{\alpha}$; that is, investors believe that a bank that cheated on its previous period's retention will not monitor the loan in the current period unless it retains $\bar{\alpha}$ fraction of the loan or more. Because we restrict history dependence to the most recent period, this is the maximum "punishment" that investors can impose on the bank for having cheated in the previous period. Given these beliefs, a bank in the reputation state d_{low} may either retain $\alpha = 0$ and not monitor at all, or retain $\alpha = \bar{\alpha}$ and monitor with certainty. Regardless of the bank's choice of retention, $R(0_{low}) = R(1_{low}) = R(0)$ because the loan rate is set before the bank chooses its retention. It follows that

$$V(d_{low}) = \bar{A} + B + \delta p\Lambda + \delta V(1), \tag{12}$$

where

$$\bar{A} \equiv \max\left(0, A - \beta\bar{\alpha}\overline{P}(1, R(0)) + \delta\Delta\Lambda - m\right) \tag{13}$$

If the bank does not cheat on its retention, then it is clear from condition 1 in the equilibrium definition that the bank will monitor if, and only if, the following IC condition is satisfied:

$$\delta\Delta\Lambda + \alpha\Delta(R - C) \geq m. \tag{14}$$

On the other hand, if the bank cheats (i.e., if it deviates to $\alpha < \alpha(d)$ when $q(d) > 0$), then it is easy to show that $q(\alpha, d) = 0$. This is because if the bank cheats, then it will be in reputation state $d_{low} \in \{0_{low}, 1_{low}\}$ with certainty in the next period. Hence, monitoring will be incentive compatible if and only if $\alpha\Delta(R - C) \geq m$, which will never be satisfied because $\alpha < \alpha(d) \leq \alpha_{pr}$ and $R \leq R(1)$ imply that $\alpha\Delta(R - C) < m$.

For notational convenience, we use $q(d) \equiv q(d, \alpha(d))$ to denote the bank's monitoring under the equilibrium retention strategy $\alpha(d)$.

Lemma 2 *In any reputational monitoring equilibrium, either $q(d) = 0$ or $q(d) = 1$; i.e.,* *mixed monitoring strategies are not supported. Moreover, if $q(d) = 0$, then $\alpha(d) = 0$.*

To support $q(d) \in (0,1)$, the IC constraint (14) must hold with equality. But then, because the bank's monitoring incentives improve with α, it is easy to show that there exists a sufficiently small $\varepsilon > 0$ so that the bank will be strictly better off by increasing its retention to $\alpha' = \alpha(d) + \varepsilon$ and monitoring with certainty. Hence, we cannot have an equilibrium in which $q(d) \in (0,1)$; it must be that either $q(d) = 0$ or $q(d) = 1$. Moreover, we cannot have an equilibrium in which $q(d) = 0$ and $\alpha(d) > 0$, because the bank is strictly better off by deviating to $\alpha' = 0$ and lowering its liquidity costs. Therefore, if $q(d) = 0$, then it must be that $\alpha(d) = 0$.

An immediate implication of Lemma 2 is that the full monitoring equilibrium characterized in Proposition 1 is infeasible if the bank can credibly commit to retain a fraction of the loan.

It follows that there are at most only four types of monitoring choices in equilibrium: $q(d) \in \{0,1\}$ for $d \in \{0,1\}$. The next lemma shows that these conditions impose restrictions on retention, too.

Lemma 3 *In any reputational monitoring equilibrium, we cannot have both $\alpha(0) \geq \alpha(1)$ and $q(0) \leq q(1)$.*

We show in the proof of Lemma 3 that if $\alpha(0) \geq \alpha(1)$ and $q(0) \leq q(1)$, then the incremental value of high reputation, Λ, is negative, in which case, the bank will not have incentives to monitor the loan. Therefore, in any reputational monitoring equilibrium, we need either $\alpha(0) < \alpha(1)$ or $q(0) > q(1)$ or both.

Proposition 2 *There are only two types of reputational monitoring equilibria that are feasible:*

1. *A high monitoring ("HM") equilibrium, in which the bank fully monitors the loan regardless of reputation (i.e.,* $q(1) = q(0) = 1$*) and retains a bigger portion of the loan in the low reputation state (i.e.,* $\alpha(0) < \alpha(1) \leq \alpha_{pr}$*); and*

2. *A low monitoring ("LM") equilibrium, in which the bank retains a portion of the loan and monitors with certainty in the high reputation state (i.e.,* $q(0) = 1$ *and* $\alpha(0) > 0$*) but sells off the loan entirely and does not monitor in the low reputation state (i.e.,* $q(1) = 0$ *and* $\alpha(1) = 0$*).*

Given Lemma 2, there are only three possible combinations of $\{q(0), q(1)\}$ that can arise in any reputational monitoring equilibrium (the fourth possible combination, $q(0) = q(1) = 0$, has no monitoring at all). Of these, we show that the combination $q(1) = 1$ and $q(0) = 0$ cannot be an equilibrium, because if the bank has incentives to monitor in the low reputation state, then it will also have incentives to monitor in the high reputation state. This leaves only two types of reputational monitoring equilibria that may be feasible with retention: either (a) $q(0) = q(1) = 1$, or (b) $q(0) = 1$ and $q(1) = 0$. In the first case (the HM equilibrium), Lemma 3 requires that $\alpha(0) < \alpha(1)$; thus, the bank always monitors fully, but the high-reputation bank benefits by retaining less of its loan than the low-reputation bank does. In the second case (the LM equilibrium), the high-reputation bank monitors fully and the low-reputation bank does not monitor at all; by Lemma 2, the low-reputation bank retains nothing, but the high-reputation bank must retain some positive amount in order to maintain its monitoring incentives.

We characterize the LM equilibrium in Section IA.4 of the Internet Appendix, where we also show that the LM equilibrium is dominated by the HM equilibrium whenever the latter is feasible (Proposition IA.3). There is also no guarantee that the LM equilibrium is feasible when the HM equilibrium is infeasible. Indeed when $u = 0$, we show that the LM equilibrium is infeasible whenever the HM equilibrium is infeasible (Proposition IA.4).

Given the dominated and limited nature of the LM equilibrium, the remainder of this section focuses on the HM equilibrium.

Because the HM equilibrium has the bank monitor fully ($q = 1$) regardless of its reputation, the driving force for reputation is the difference in retention between high-reputation and low-reputation banks. Although in principle one could have the low-reputation bank retain $\alpha(1) > \alpha_{pr}$ and have market participants react to any deviation to lower retention by assigning the bank a "punishment" reputation d_{low} in the following period, we now show that this is implausible.

Recall that the loan rate is set before the retention decision; given that the low-reputation bank is supposed to monitor fully, the loan rate is $R(1)$. If a bank retains less than it is supposed to, its reputation the following period is d_{low}, which is effectively independent of the outcome d of the loan this period. That means there is no longer any reputation benefit to monitoring this period. Therefore, the bank will monitor if and only if its retained share makes it worthwhile to do so; that is, using Condition (14), if and only if $\alpha(R(1) - C) > m$. If the low-reputation bank deviates from $\alpha(1) < \alpha_{pr}$ to some lower α', it is immediate that $\alpha'(R(1) - C) < m$. However, if $\alpha(1) > \alpha_{pr}$ and the bank deviates to $\alpha' = \alpha(1) - \epsilon$, then for ϵ sufficiently small, we will have $\alpha'(R(1) - C) > m$ and the bank will monitor fully. As in the case of deviations to higher retention, it seems very unlikely that dispersed, independently-acting market participants would punish such a deviation the following period: they know the bank did monitor fully (and retained enough to guarantee that), so why would all agree this bank couldn't be trusted? For this reason, in our main analysis we require that $\alpha(1) \leq \alpha_{pr}$.[8]

Our next result fully characterizes the HM equilibrium.

[8] In Section IA.4.2 of the Internet Appendix, we briefly analyze how allowing $\alpha(1) > \alpha_{pr}$ in the HM equilibrium affects our results.

Proposition 3 *For any $\rho \in (0,1]$, the high monitoring equilibrium is feasible if and only if*

$$\alpha_{pr} \cdot \beta P(1) \geq \frac{m}{\delta \Delta} \iff \delta\beta \left[p + \Delta + \frac{C}{X - C - \frac{u}{p+\Delta}} \right] \geq 1 \tag{15}$$

and

$$\begin{aligned} A - m \quad & - (1-\rho)\,\alpha_{pr}\beta P(1) - \frac{\rho m \left[(1+\delta)(1-\delta\Delta) - \delta p\right]}{\delta\Delta(1+\delta)} \\ \geq \quad & \frac{u\Delta(1+\beta)}{(p+\Delta)(1+\delta)} + \frac{\delta \max\left[0, A - \beta\bar{\alpha}\overline{P}(1, R(0)) - m(1-\rho)\right]}{(1+\delta)} \end{aligned} \tag{16}$$

If these conditions are satisfied, the following high monitoring equilibrium is feasible: $\alpha(0) = (1-\rho)\,\alpha_{pr}$, $\alpha(1) = \alpha(0) + \frac{\rho m}{\delta\Delta\beta P(1)}$, and $q(0) = q(1) = 1$. Under such an equilibrium, bank value in the high reputation state is

$$V_{HM}(0) = (1-\delta)^{-1} \cdot \left[A + B - (1-\rho)\,\alpha_{pr} \cdot \beta P(1) - m(1-\rho) - \frac{\rho m (1-p)}{\Delta} \right], \tag{17}$$

and bank value in the low reputation state is $V_{HM}(1) = V_{HM}(0) - \frac{\rho m}{\delta\Delta}$.

As noted above, in the HM equilibrium, the bank always monitors fully, so the value of higher reputation comes from being able to retain less of the loan: $\Lambda = [\alpha(1) - \alpha(0)]\beta P(1)$. There are two sets of conditions that must be satisfied for the HM equilibrium to be feasible. First, the IC constraint (14) must hold in both reputation states so that the bank has incentives to monitor. Because $\alpha(1) > \alpha(0)$, Constraint (14) will hold for the low-reputation bank if it holds for the high-reputation bank. Furthermore, in the proof of this proposition, we show that any HM equilibrium in which the IC constraint holds strictly for the high-reputation bank delivers strictly lower expected bank profits in both reputation states than another feasible HM in which the IC constraint binds with equality for the high-reputation bank. This lets us focus on HM equilibria in which the IC constraint binds with equality for the high-reputation bank. Second, the bank must not prefer deviating to a zero-retention and no-monitoring strategy. Since the low-reputation bank has a lower value than the high-reputation bank, this non-deviation condition will always hold for the high-reputation bank if it holds for the low-reputation bank.

There are multiple combinations of $\alpha(0)$ and $\alpha(1)$ that can satisfy these two conditions; because $\alpha(0)$ must be between 0 and α_{pr}, we can index these equilibria with the variable $\rho \in (0,1]$, where $\alpha(0) = (1-\rho)\alpha_{pr}$. If the IC constraint binds with equality in the high reputation state, it follows that $\alpha(1) = \alpha(0) + \frac{\rho m}{\delta\Delta\beta P(1)}$.[9] The feasibility condition (15) in Proposition (3) is necessary to ensure that $\alpha(1) \leq \alpha_{pr}$. We note that this condition is more likely to be met when C is high and $X - C$ is low; i.e., when the loan is relatively safe. It is also more likely to hold if bank capital or liquidity needs β are high. Although the expression on the left-hand side is non-monotonic in $p + \Delta$, the condition is met if $p + \Delta$ is high enough that $\delta\beta(p+\Delta) \geq 1$, which means that the chance of default for a monitored loan cannot be too high. The second feasibility condition (16) is necessary to ensure that the non-deviation condition is satisfied in the low reputation state. If both the feasibility conditions (15) and (16) are satisfied, then the HM equilibrium described in the proposition becomes feasible.

Corollary 1 *Among the high monitoring equilibria described in Proposition 3, the one indexed by $\rho = 1$ – that is, $\alpha(0) = 0, \alpha(1) = \frac{m}{\delta\Delta\beta P(1)}$, and $q(0) = q(1) = 1$ – is most likely to be feasible and is optimal for the high-reputation bank. This equilibrium is feasible if and only if condition (15) holds and*

$$A - \frac{m(1+\delta-\delta p)}{\delta\Delta} + \delta \min\left[A, \beta\bar{\alpha}\overline{P}(1, R(0))\right] \geq \frac{u\Delta(1+\beta)}{(p+\Delta)} \tag{18}$$

Moreover, this equilibrium achieves the constrained first-best outcome: the value of the high-

[9] Note that the limiting case of $\rho = 0$ corresponds to a "pure retention" equilibrium in which the bank retains α_{pr} fraction of the loan regardless of its reputation (i.e., $\alpha(0) = \alpha(1) = \alpha_{pr}$) and fully monitors the loan.

reputation bank, $V_{HM}(0)$, is equal to V^*.

Corollary 1 follows by noting that when the feasibility condition (15) is satisfied, then the second feasibility condition (16) is more likely to be met when ρ is higher, and that $V(0)$ is increasing in ρ. Hence, the high monitoring equilibrium indexed by $\rho = 1$, with $\alpha(0) = 0$ and $\alpha(1) = \frac{m}{\delta \Delta \beta P(1)}$ is most likely to be feasible and is optimal for the high-reputation bank.

It is easily verified from equation (17) that $[V_{HM}(0)]_{\rho=1} = V^*$. In other words, both the full monitoring equilibrium (in the setting without loan retention) and the high monitoring equilibrium (in the setting with loan retention) achieve the maximum possible bank value, V^*. Therefore, making loan retention credible in our setting does not necessarily improve bank value.[10]

However, it is worth emphasizing that the equilibria with loan retention are more stable than the pure-reputation equilibria characterized in Section 3. In the case when retention is not credible, reputational equilibria require that the high- and low-reputation banks must pick specific monitoring intensities even though they are actually indifferent to changing their monitoring choices. This is a common problem with mixed strategy equilibria, of course. By contrast, once retention is credible, the bank never mixes its monitoring choice, avoiding this problem.

5. Reputation and Lending Booms

In the baseline model, we assumed that the bank faces a constant demand for loans each period, which we normalized to 1 unit. In this section, we allow for stochastic "lending booms" (i.e., periods during which there is a sharp increase in the demand for bank loans), and

[10] We must emphasize that the HM equilibrium supports strictly higher monitoring in the low reputation state compared to the full monitoring equilibrium without retention. In our model the borrower is indifferent to the level of monitoring because we have assumed that the bank captures all the gains from monitoring. However, if we assume that other stakeholders of the borrowing firm benefit from the project's success (Bebchuk and Goldstein 2011), then making loan retention credible could deliver a higher social surplus.

examine how the bank's monitoring incentives and reputation considerations vary between boom periods and periods of lower ("normal") loan demand.

5.1. Assumptions

We modify the baseline case by assuming that, in each period, the level of loan demand from the entrepreneur is now stochastic: the entrepreneur demands only 1 unit of loan ("normal" demand) with probability ϕ, but demands $\gamma > 1$ units of loans ("boom" demand) with probability $1 - \phi$. The project's output and the bank's monitoring cost are both assumed to be directly proportional to the scale of the loan. Let $s \in \{n, b\}$ denote the state of the economy in terms of loan demand, where n denotes normal and b denotes boom. We assume that the state s is independently and identically distributed over time.

At the beginning of each period, the bank observes the state of the economy s and then decides on its loan volume, denoted ℓ. For simplicity, in a boom, we restrict the bank to choose between either 1 or γ as its loan amount. If loan demand is normal, the bank simply lends 1 unit as in the baseline model. Although investors observe the bank's loan volume $\ell \in \{1, \gamma\}$, they never observe the loan demand s. This assumption simplifies the possible links between bank actions, states, and investor beliefs about bank monitoring; in particular, the market cannot differentiate between the bank lending 1 unit in a normal economy from the bank lending 1 unit in a boom.

To begin our analysis, we assume that the bank cannot commit to retain any part of the loans it makes; we relax this restriction in Section 5.5.

5.2. Definition of Equilibrium

Note that the bank's public history h now includes information on both its past loan outcomes (default or no default) and past lending volumes (1 or γ units). As in the baseline model, for tractability, we restrict attention to PPE in which the only element of h that matters for bank actions and investor beliefs is the performance of its most recent loan, $d \in \{0, 1\}$. Thus, the market's belief about bank monitoring depends only on d and current period lending volume ℓ.[11] Moreover, because the current state of loan demand $s \in \{n, b\}$ reveals nothing about future loan demand, we assume that the bank's actual monitoring also depends only on d and ℓ, and not on s. We impose these simplifying restrictions to maintain tractability while capturing the key elements of the stochastic loan demand setting.

Let $V(d)$ denote the bank's expected discounted value at the beginning of any given period, before the state $s \in \{n, b\}$ is realized. This in turn can be written as

$$V(d) = \phi V_n(d) + (1 - \phi) V_b(d), \tag{19}$$

where $V_s(d)$ represents the bank's discounted expected profits given that it has a reputation of d and the current state of the economy is s.

Definition: We examine perfect public equilibrium in which the set of the bank's lending and monitoring strategies $\{\ell(d, s), q(d, \ell)\}$ and investor beliefs $\overline{q}(d, \ell)$ satisfy the following conditions:

1. Given its lending strategy $\ell(d, s)$ and investor beliefs $\overline{q}(d, \ell)$, the bank's monitoring choice $q(d, \ell)$ maximizes

$$-mq \cdot \ell(d, s) + \delta(p + \Delta q)(V(0) - V(1) + \delta V(1).$$

2. Given investor beliefs $\overline{q}(d, \ell)$, the bank's lending volume in the boom state $\ell(d, b)$

[11] In Section IA.5 of the Internet Appendix, we analyze an alternative reputational equilibrium in which the bank would not lose or gain reputation if its loan volume is γ, and reputation depends on defaults only if loan volume is 1. The qualitative results under this alternative equilibrium are similar to what we find in this section.

maximizes

$$V_b(d,\ell) = \ell \cdot [A\overline{q}(d,\ell) + B - q(d,\ell)\cdot m] + \delta[(p + q(d,\ell)\cdot\Delta)\Lambda + V(1)].$$

In the normal state, $\ell\,(d,n) = 1$ regardless of d.

3. The investors' beliefs agree with the bank's equilibrium monitoring choice: $\overline{q}(d,\ell) = q(d,\ell)$.

5.3. Monitoring Incentives and Bank Value

Because investors do not observe the state of the loan market, the bank's reputation next period depends only on whether the current loan defaults or not. Given our assumption that bank monitoring only depends on d and ℓ, and not on the state of loan demand, there are only two cases to consider.

First, consider the case where the bank chooses to lend only one 1 unit, which can happen in either of the demand states. Then, by the same logic as in the base model, it follows that the bank monitors if, and only if, the following incentive compatibility constraint holds:

$$V(0) - V(1) \equiv \Lambda \geq m/\delta\Delta. \tag{IC(1)}$$

Although condition $IC(1)$ is the same as condition (3) in the base model, the discounted value functions $V(0)$ and $V(1)$ are more complex now.

Next, consider the case where the bank chooses to fulfill the boom demand and lend γ units. Because monitoring costs γm, the bank monitors if, and only if, the following condition holds:

$$\Lambda \geq \gamma m/\delta\Delta. \tag{IC(γ)}$$

Because Λ does not depend on the current state of loan demand, $IC(\gamma)$ is stricter than

$IC(1)$.

By a similar logic as in the base model, the bank's current period surplus from selling ℓ units of the loan is $S(d,\ell) \equiv \ell \cdot [\bar{q}(d,\ell) \cdot A + B]$, where A and B are just as in Section 3. We can now write the bank's value function in each state of the economy. First, if the economy is in a normal state, we have

$$V_n(d) = A\bar{q}(d,1) + B - q(d,1) \cdot m + \delta[(p + q(d,1) \cdot \Delta)\Lambda + V(1)]. \tag{20}$$

If instead the economy is in a boom, we have

$$V_b(d,\ell) = \ell \cdot [A\bar{q}(d,\ell) + B - q(d,\ell) \cdot m] + \delta[(p + q(d,\ell) \cdot \Delta)\Lambda + V(1)]. \tag{21}$$

We can now establish the following result:

Lemma 4 *In any equilibrium in which monitoring occurs with positive probability, either $IC(1)$ binds and $IC(\gamma)$ fails to hold, or else $IC(\gamma)$ binds and $IC(1)$ holds strictly.*

Suppose the equilibrium parameters are such that $IC\,(1)$ holds strictly and either $IC\,(\gamma)$ also holds strictly, or $IC\,(\gamma)$ fails to hold. In the former case, the bank will always monitor the loan fully regardless of its lending volume, whereas in the latter case the bank will always monitor fully when it lends 1 unit but will not monitor if it lends γ units. In either of these cases, the bank earns the same expected future profits regardless of its reputation, so that $\Lambda = 0$ and there is no gain to monitoring, which is a contradiction. Hence, given that $IC(\gamma)$ is stricter than $IC(1)$, one of the two possibilities described in the lemma must hold if there is any monitoring in equilibrium.

It follows that there are two broad classes of monitoring equilibria in this setting: those in which the bank's monitoring incentives hold when it lends 1 unit but not more, and those in which its monitoring incentives hold when it lends either 1 unit or γ units but hold strictly when it lends 1 unit. However, our next result shows that the second class of equilibria only exists if ϕ is sufficiently low, that is, when booms are very common.

Lemma 5 *An equilibrium in which the bank' s monitoring incentives hold when it lends γ units and hold strictly when it lends 1 unit (i.e., $IC\,(1)$ constraint holds strictly and $IC\,(\gamma)$ constraint binds) cannot exist if $m > \delta\Delta\,(1-\phi)\,A$.*

If the bank monitors with certainty when it lends 1 unit, then its value function in the normal state does not vary with reputation. Therefore, the incremental value of high reputation, Λ, is entirely driven by differences in the value function between high- and low-reputation banks in the boom state; formally $\Lambda = (1-\phi)\,(V_b\,(0) - V_b\,(1))$. Hence, such equilibria are feasible only when the probability of a boom (i.e., $1-\phi$) and the additional surplus to the bank from monitoring the loan (i.e., A) are sufficiently high in comparison to the monitoring cost, m.

To focus on situations of most economic interest, we will assume that booms are *uncommon*, by imposing the following assumption:

Assumption 4: $m > \delta\Delta(1-\phi)A = \delta\Delta^2(1-\phi)(1+\beta)(X-C)$.

5.4. Equilibria when Booms are Uncommon

Suppose Assumption 4 holds, so that booms are relatively uncommon. Then, as per Lemma 5, the only feasible monitoring equilibria are those in which $IC(1)$ binds and $IC(\gamma)$ fails to hold, so that the bank monitors with positive probability if it lends 1 unit but will not monitor if it lends γ units in a boom (i.e., $q\,(d,\gamma) = \overline{q}(d,\gamma) = 0$ for $d \in \{0,1\}$). There are four possible equilibria to consider, based on how the bank chooses its lending volume $\ell \in \{1,\gamma\}$ in a boom, in the high and low reputation states. In each of these equilibria, three incentive compatibility conditions must hold: (a) $IC(1)$ must bind, so we must have $\Lambda = \frac{m}{\delta\Delta}$; (b) in a boom, a high-reputation bank must not prefer to switch its loan volume choice, $\ell_b(0)$; and (c) in a boom, a low-reputation bank must not prefer to switch its loan volume choice $\ell_b(1)$.

We have the following results:

Proposition 4 *A monitoring equilibrium in which a low-reputation bank lends 1 unit during booms whereas a high-reputation bank lends γ units is infeasible.*

A low-reputation bank has strictly higher incentive than a high-reputation bank to lend γ units in a boom and shirk on monitoring because it has less to lose from any resultant loss of reputation next period. Hence, equilibria in which a high-reputation bank lends γ in a boom whereas a low-reputation bank lends 1 unit are infeasible. This means that only three equilibria are possible: a "tight credit" equilibrium in which the bank never accommodates the higher loan demand in booms (Proposition 5); a "partially-tight credit" equilibrium in which the low-reputation bank accommodates higher loan demand in booms but the high-reputation bank does not (Proposition 6); and a "loose credit" equilibrium in which both high- and low-reputation banks accommodate higher loan demand during booms (Proposition 7).

Proposition 5 *A tight credit ("TC") monitoring equilibrium in which the bank lends 1 unit in both economic states regardless of its reputation is feasible if and only if $m \leq \delta\Delta \cdot (A - (\gamma - 1)B)$. In this equilibrium, the bank monitors with probability 1 in the high reputation state, and with probability $\widehat{q} = 1 - m/\delta\Delta A$ in the low reputation state.*

The TC equilibrium in Proposition 5 is identical to the full-monitoring equilibrium from Proposition 1 in the base model, but with more limited feasibility: the condition $m/\delta\Delta \leq A - (\gamma - 1)B$ is more restrictive than the corresponding condition in Proposition 1, which was $m/\delta\Delta \leq A$. The tighter limit on feasibility arises from the added constraint that a low-reputation bank must weakly prefer lending 1 unit in a boom and monitoring with intensity $\widehat{q}$ over lending γ units and not monitoring at all. (As noted above, this constraint is sufficient to guarantee that a high-reputation bank will also not deviate to lending more.) This requires that loan demand γ during booms cannot be too high, so that the additional gross value A created by monitoring is sufficiently large in comparison to the additional base (unmonitored) surplus $(\gamma-1)B$ that can be obtained by accommodating higher loan volume.

Proposition 6 *A partially tight credit ("PTC") monitoring equilibrium in which a high-reputation bank lends 1 unit during booms whereas a low-reputation bank lends γ units is feasible if and only if $A \geq (\gamma-1)B$ and $m/\delta\Delta \leq A-(1-\phi)(\gamma-1)B$. In this equilibrium, a high-reputation bank monitors with probability $\hat{q}_{0,ptc} \equiv \min\{1, [(\gamma-1)B/A + m/\delta\Delta A]\}$ in both economic states. By contrast, a low-reputation bank monitors with probability $\hat{q}_{1,ptc} = \phi^{-1}\left[\hat{q}_{0,ptc} - m/\delta\Delta A - (1-\phi)(\gamma-1)B/A\right]$ in a normal economy and does not monitor at all in a boom.*

The PTC equilibrium comes about because the low-reputation bank has strictly higher incentive to lend more in a boom than the high-reputation bank. Note that the feasibility condition in Proposition 5 is stricter than those in Proposition 6. It follows that the PTC equilibrium is feasible whenever the TC equilibrium is feasible and sometimes when the latter is not.

Proposition 7 *A loose credit ("LC") monitoring equilibrium in which the bank always lends γ units during booms regardless of reputation is feasible if and only if $m \leq \phi\delta\Delta \cdot \min\{A, (\gamma-1)B\}$. In this equilibrium, the bank never monitors in a boom. In a normal economy, the high-reputation bank monitors with probability $\hat{q}_{0,lc} \equiv \min\{1, (\gamma-1)B/A\}$ and the low-reputation bank monitors with probability $\hat{q}_{1,lc} = \hat{q}_{0,lc} - m/\phi\delta\Delta A$.*

In the LC equilibrium, the bank always accommodates the higher loan demand in booms, regardless of reputation. For this to be feasible, booms must be sufficiently unlikely (ϕA must exceed $m/\delta\Delta$); otherwise, because reputation doesn't affect profits in booms, it would be impossible to support incentive compatibility based on the difference between high- and low-reputation values in normal times alone. Also, loan demand γ in booms must be sufficiently high ($(\gamma-1)\phi B$ must exceed $m/\delta\Delta$) to ensure that the high-reputation bank does not deviate to lending 1 unit in a boom.

Recall that the bank is in the high reputation state ($d = 0$) at the initial date, $t = 0$. Our next result characterizes the "optimal" equilibrium from among the three equilibria described above, that is, the equilibrium that delivers the highest total surplus, $V(0)$, at the initial date. It is clear that none of the three equilibria described above are feasible if $m > \delta\Delta A$.

Proposition 8 *If $m \leq \delta\Delta A$, then there exist thresholds $\hat{\gamma}_{tc} \equiv \frac{(A-(m/\delta\Delta))}{B} + 1$ and $\hat{\gamma}_{ptc}(\phi) \equiv \frac{(A-(m/\delta\Delta))}{(1-\phi)B} + 1$ such that:*

1. *If $\gamma \leq \hat{\gamma}_{tc}$, the TC equilibrium is feasible, (weakly) dominates the PTC equilibrium, and strictly dominates the LC equilibrium.*

2. *If $\hat{\gamma}_{tc} < \gamma \leq \hat{\gamma}_{ptc}(\phi)$, the PTC equilibrium is feasible and (weakly) dominates the LC equilibrium whenever the latter is feasible, whereas the TC equilibrium is infeasible in this region.*

3. *The LC equilibrium is the only feasible equilibrium if $\gamma > \hat{\gamma}_{ptc}(\phi)$ and $m \leq \phi\delta\Delta \cdot \min\{A, (\gamma-1)B\}$.*

Suppose $\gamma \leq \hat{\gamma}_{tc}$ so that both the TC and PTC equilibria are feasible. It is clear that the bank's monitoring in the high reputation state under the TC equilibrium is weakly higher compared to the PTC equilibrium ($\hat{q}_{0,ptc} \leq 1$) and strictly higher compared to the LC equilibrium (because $\hat{q}_{0,lc} = (\gamma-1)B/A < 1$ when the TC equilibrium is feasible). Because the benefits of accommodating the higher loan demand are also small when γ is small, it follows that, when it is feasible, the TC equilibrium (weakly) dominates the PTC equilibrium (i.e., $V_{tc}(0) \geq V_{ptc}(0)$) and strictly dominates the LC equilibrium (i.e., $V_{tc}(0) > V_{lc}(0)$).

Extending this reasoning, it follows that the PTC equilibrium is the dominant equilibrium only when it is feasible and the TC equilibrium is infeasible, which happens for intermediate values of γ.

Finally, if the boom-time loan demand (γ) and the likelihood of booms $(1-\phi)$ are sufficiently high, then neither the TC equilibrium nor PTC equilibrium is feasible. Hence, in this parameter region characterized by $\gamma > \hat{\gamma}_{ptc}(\phi)$, the LC equilibrium is the dominant equilibrium whenever it is feasible.

5.5. Retention, Reputation, and Lending Booms

Thus far, our analysis of stochastic lending booms has been based on the assumption that the bank cannot commit to retain a portion of the loan. The main takeaway is that if the size of the boom γ is sufficiently high, then the only feasible monitoring equilibrium is one in which the bank always lends γ in a boom and does not monitor regardless of its reputation (Propositions 7 and 8). In this section, we examine if this inefficient no-monitoring outcome in booms can be eliminated if the bank can commit to retain a portion of the loan.

For simplicity, we will assume that the bank's retention decision also only depends on its reputation d and loan volume ℓ, and not the state of the economy s. Let $\alpha(d,\ell)$ denote the bank's retention strategy. Let $q(d,\ell,\alpha)$ the bank's monitoring choice, and let $\bar{q}(d,\ell,\alpha)$ denote the market's conjecture of the bank's monitoring. Based on the logic established in Section 4, the bank's IC constraint for monitoring when it lends 1 unit may be written as

$$\delta\Delta\Lambda + \alpha\Delta \cdot [R - C] \geq m. \tag{22}$$

Similarly, the IC constraint corresponding to $\ell = \gamma$ may be written as

$$\delta \Delta \Lambda + \gamma \alpha \Delta \cdot [R - C] \geq \gamma m. \tag{23}$$

To focus on situations of most economic interest, we assume that γ is large enough that, in the absence of retention, the bank would always prefer to meet the boom demand (i.e., $\ell(d) = \gamma$ in the boom state for $d \in \{0, 1\}$). We also focus attention on high monitoring equilibria with retention in which the bank always fully monitors the loan but retains a larger portion of the loan in the low reputation state; i.e., $q(d, \ell) = 1$ for all combinations of d and ℓ, $\alpha(0, 1) < \alpha(1, 1) \leq \alpha_{pr}$ and $\alpha(0, \gamma) < \alpha(1, \gamma) \leq \alpha_{pr}$. We refer to these equilibria as high monitoring and high lending (HMHL) equilibria.

We define γ_ϕ as follows to denote the expected loan demand:

$$\gamma_\phi \equiv \phi + (1 - \phi) \gamma \tag{24}$$

Proposition 9 *If $\alpha_{pr} \geq \frac{m}{\delta \Delta \beta P(1)}$ and $A - m - \alpha_{pr} \beta P(1) \geq \frac{u\Delta(1+\beta)}{p+\Delta}$, then there exists a high monitoring and high lending equilibrium (HMHL) for every $\rho \in (0, 1]$ in which the bank always meets the boom demand and fully monitors the loan regardless of its reputation and lending volume (i.e., $q(d, \ell) = 1$ for all d and ℓ), but its retention strategy $\alpha(d, \ell)$ varies with reputation and lending volume as follows: $\alpha(0, 1) = (1 - \rho) \alpha_{pr}$, $\alpha(1, 1) = \frac{\rho m}{\delta \Delta \beta P(1)} + (1 - \rho) \alpha_{pr}$, $\alpha(0, \gamma) = \frac{(\gamma - \rho)\alpha_{pr}}{\gamma}$, and $\alpha(1, \gamma) = \frac{1}{\gamma} \left(\frac{\rho m}{\delta \Delta \beta P(1)} + (\gamma - \rho) \alpha_{pr} \right)$. Under this equilibrium, the bank value in the high reputation state, $V(0)$, is given by*

$$V(0) = (1 - \delta)^{-1} \cdot \left(\gamma_\phi \cdot [A + B - m - \alpha_{pr} \beta P(1)] + \rho \alpha_{pr} \beta P(1) - \frac{\rho m}{\Delta} (1 - p - \Delta) \right). \tag{25}$$

The bank value in the low reputation state is $V(1) = V(0) - \frac{\rho m}{\delta \Delta}$.

Under the HMHL equilibria, the incremental value of high reputation Λ depends on the difference in retention between the low and high reputation states. There are several conditions that must be satisfied in equilibrium. First, the bank's IC constraint must hold in both reputation states ($d \in \{0,1\}$) and for both lending volumes ($\ell \in \{1,\gamma\}$). Because monitoring incentives depend partly on reputation, it follows that the bank will retain a larger portion of the loan in the low reputation state, regardless of the loan volume (i.e., $\alpha(1,\ell) > \alpha(0,\ell)$ for all ℓ). As in Section 4, we show that the efficient equilibria are those in which the IC constraint binds with equality in the high reputation state. Then, it is clear from the IC constraints (22) and (23) that, for monitoring to be incentive compatible, the high-reputation bank will have to retain a larger portion of the loan when it meets the boom demand (i.e., $\alpha(0,\gamma) > \alpha(0,1)$).

Second, we need to check that the bank does not deviate to a no-retention strategy with zero monitoring after the loan rate has been set at $R(1)$, both in the normal state and the boom state. Since $V(0) > V(1)$, we only need to check this for the low-reputation bank, which is more likely to prefer such a deviation. Third, we need to check that, in the boom state, the bank will not deviate to a strategy of lending 1 unit and retaining $\alpha(d,1)$ portion of the loan. The feasibility conditions listed in Proposition 9 are *sufficient*, but not necessary, to ensure that all these equilibrium conditions are satisfied.[12]

Corollary 2 *Among the equilibria characterized in the proposition above, the one indexed by $\rho = 1$ is most likely to be feasible and is optimal for the high-reputation bank.*

The corollary follows by noting that if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, then $V(0)$ is increasing in ρ. Moreover, we show in the proof of Proposition 9 that the more general feasibility conditions

[12] We show in the proof of Proposition 9 that the HMHL equilibria may be feasible even if $A - m - \alpha_{pr}\beta P(1) < \frac{u\Delta(1+\beta)}{p+\Delta}$, provided $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$ and $A + B - m - \alpha_{pr}\beta P(1) \geq 0$. In this case, the HMHL equilibria exist for every $\rho \in (0,1]$ that satisfies the following condition:

$$\left[\begin{array}{c} (\gamma + \delta\gamma_{\phi}) \cdot (A - m - \alpha_{pr}\beta P(1)) - \gamma\frac{u\Delta(1+\beta)}{p+\Delta} \\ -\delta\bar{A} + (1+\delta)\,\rho\alpha_{pr}\beta P(1) - \frac{\rho m[(1+\delta)(1-\delta\Delta)-\delta p]}{\delta\Delta} \end{array}\right] \geq 0.$$

for the existence of the HMHL equilibria are more likely to be satisfied for higher values of ρ. We also show in Section IA.6 of the Internet Appendix that the high-reputation bank's value under the HMHL equilibrium with $\rho = 1$ equals the maximum bank value that can be attained by any perfect public equilibrium under the setting with stochastic loan demand when loan retention is credible.

A final observation on this HMHL equilibrium concerns the two sufficient conditions in Proposition 9. We show in the proof of Corollary 2 that these two conditions are also sufficient to guarantee the existence of the optimal HM equilibrium characterized in Corollary 1 (i.e., the HM equilibrium with $\rho = 1$). This implies that the conditions for HMHL equilibria depend more on loan and bank characteristics than on properties of stochastic loan demand such as the frequency or size of booms.

6. Discussion and Conclusion

We derive the conditions under which reputation concerns will cause lenders to monitor loans they sell. Because the history of defaults on a bank's loans is a noisy indication of whether it has monitored its loans in the past, this history can serve as a reputation that may allow the bank to commit to monitor even if it sells off all or part of the loans it originates. That is, banks with fewer past defaults have higher reputation. In our model reputation is a sanctioning mechanism that operates purely through the threat of lower future profits following poor current performance.

As we have seen, the nature of equilibrium depends critically on whether or not the bank can commit to retain a specific fraction of the loan. If the bank is expected to sell off the loan entirely, then monitoring may still be feasible in equilibrium but the high-reputation bank monitors more intensively than a low-reputation bank. On the other hand, if the bank can credibly commit to retain a portion of the loan, then some additional conditions are required to support equilibria in which the bank always monitors the loan fully but retains a smaller portion in the high reputation state. When such "high monitoring" equilibria are feasible, they achieve the constrained first-best value for the high-reputation bank. We now discuss when these various cases are likely and their implications for actual OTD loan markets.

In situations where retention commitments are not credible, the bank's monitoring incentives must come from reputation alone. As we have seen, this can work, but the equilibria are fragile: the bank's incentive compatibility condition always holds with equality, and a low-reputation bank must follow a mixed strategy in monitoring that precisely matches investor expectations. Furthermore, large infrequent booms in loan demand can undermine monitoring incentives, especially for a bank that has a low reputation. In practice, maintaining such equilibria may be difficult, especially if the underlying constellation of parameters shifts over time in a way investors cannot quickly observe.

There are many reasons why banks may be unable to commit to retention. If trading in shares of loans (or claims on loan securitization) is decentralized or difficult for other participants to track, then banks may find it easier to unload their retained shares without alerting investors. Alternatively, the bank may unload its share even if this tips off the market maker (and later other investors): getting a price for the retained share that reflects shirking does not change the fact that the bank sold the rest of the loan at an inflated price initially.

Such a setting seems most relevant to securitization deals, where banks were known to sell their equity tranches (i.e., residual claims on securitized loan portfolios) to others either up front or after the securitization took place. It may also be relevant to syndicated loans to especially large and well-known corporate borrowers, so that many participant lenders are involved and the secondary market for the loans is very liquid. Highly levered transactions (HLTs) may be another case, because these loans are increasingly sold to non-bank institutions who do not have close relationships with the initial lender and thus may not be able to track its continuing loan holdings. In these settings, we should expect that participants will look more to reputation per se rather than to retained share, especially because it can be hard for the broader investing public to see whether or not a bank continues to retain part of its loans after the initial sale to investors has taken place.

In this respect, two studies of the relative performance of syndicated loans that were securitized via collateralized loan obligations (CLOs) are especially apt. Benmelech, Dlugosz, and Ivashina (2012) find only limited evidence for underperformance of syndicated loans that were securitized versus those that were not securitized, with the evidence concentrated during the boom years 2005-2007, but their sample identifies CLO ownership via DealScan, which is incomplete on this measure, and includes loans that CLOs acquired on the secondary market. Using a more complete sample that uses information from the Shared National Credit (SNC) program to identify loans sold into CLOs at origination, Bord and Santos (2015) revisit this issue for loans originated during 2004-2008. They find strong evidence that loans sold into CLOs at the time of loan origination subsequently perform worse than unsecuritized loans originated by the same bank. They also find that, compared with unsecuritized loans, the loans sold into CLOs have terms that are less responsive to hard information about the borrower, have higher loan spreads, and have smaller (or in some cases no) retention by lead arrangers at the time of origination. All of these findings suggest a deterioration in standards as banks accommodated the boom in demand for syndicated loans and the simultaneous rise of CLOs as a vehicle for loan securitization.

By contrast, many syndicated loans have restrictions on loan sales; for example, the seller may be required to get the consent of other holders of the loan before a sale is allowed. Pyles and Mullineax (2008) show that such constraints are in fact more likely for smaller borrowers, where bank monitoring is likely to be most critical. Even in the absence of such restrictions, we would hypothesize that smaller syndicates (which should lead to less liquid loan resale markets) may make it easier for banks to effectively commit to retain shares in their loans.

When retention is credible, equilibria are very different. When a high monitoring equilibrium is feasible, banks always monitor intensively, but low-reputation banks must retain more of their loans than high-reputation banks do. As a consequence, incentive compatibility need not hold with equality. The combination of pure monitoring strategies and nonbinding incentive constraints should make these equilibria more robust. Moreover, in this case equilibrium monitoring is *not* undermined by large rare booms in demand—banks simply retain a greater share of their loans. However, such equilibria require that the bank's capital or liquidity needs are high and that loan risk is relatively low. If these conditions are not met, then either a less attractive "low-monitoring" equilibrium holds or there is no way the bank can commit to monitoring.

We conclude with a discussion of our paper's implications for regulatory retention requirements. In the aftermath of the Financial Crisis of 2007-2009, a widespread view was that lenders' lack of "skin in the game" had caused lax standards and ensuing defaults. Indeed, the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 mandated that lenders retain 5% of the credit risk in loans they originate.[13] As we now discuss, in brief, such a rule is likely to be counterproductive, especially if banks can already credibly commit to retaining a fraction of their loans.

Suppose regulators mandate that all lenders retain a fixed fraction α_{skin} of any loans they originate. First, assume that banks can already credibly retain any fraction of their loans, which means that complying with the requirement will not greatly increase lenders' transaction and reporting costs. Nevertheless, such a requirement is likely to reduce welfare. If any HM equilibria are feasible in the absence of the requirement, we know from Corollary 1 that the optimal HM equilibrium is the one where high-reputation banks retain nothing and

[13]However, many classes of loans have been exempted from this requirement. For example, the original act exempted qualified residential mortgages, and a 2018 ruling by the U.S. Court of Appeals for the D.C. Circuit exempted syndicated loans held by managers of Collateralized Loan Obligations (CLOs).

low-reputation banks retain a fraction $m/\delta\Delta\beta P(1)$, and that this equilibrium is indeed feasible. Increasing the minimum retention from 0 to α_{skin} will only reduce welfare: effectively, it shifts from the optimal equilibrium with index $\rho = 1$ to one with $\rho = 1 - (\alpha_{skin}/\alpha_{pr})$, and equilibria with lower ρ have higher expected bank capital or liquidity costs and thus lower welfare. (Note that choosing $\alpha_{skin} > \alpha_{pr}$ merely increases inefficiency, because α_{pr} is sufficient to ensure full monitoring by any lender.)

On the other hand, the new retention requirement does loosen Condition (16), which is the requirement that low-reputation lenders not deviate to not monitoring and retaining a lower amount: this lower amount is now α_{skin} rather than zero, which results in less gain in terms of lower liquidity or capital costs and thus makes shirking less attractive. Thus, if the HM equilibrium is not feasible absent regulation, and if Condition (15) holds, then a regulatory retention requirement might make an HM equilibrium that depends in part on reputation incentives feasible. (If Condition (15) is violated, then lenders only have incentive to monitor if reputation does not play role; i.e., we must have $\alpha(0) = \alpha(1) = \alpha_{pr}$.)

Next, assume that, absent regulation, banks cannot commit to retaining any fraction of their loans. If a reputation equilibrium (as in Proposition 1) exists, then it already achieves the constrained first-best. However, as discussed above, such an equilibrium may be fragile, since it requires mixed strategies and indifference to monitoring for both high- and low-reputation banks. In addition, if monitoring loans creates positive externalities, then something like the HM equilibrium will dominate the baseline case, because banks always monitor fully in the HM equilibrium whereas in the baseline equilibrium the low-reputation bank only monitors with lesser intensity.

If fragility or monitoring externalities are concerns, then a retention requirement may be helpful: such a requirement makes bank retention credible, making some form of HM equilibrium a possibility. However, given that banks cannot commit to retain part of their loans on their own, this regulatory regime will require added compliance efforts for banks and oversight efforts by regulators to make retention credible. Policy makers must weigh these added costs against potential benefits in the form of more stable monitoring equilibria and higher overall monitoring and resulting externalities.

Perhaps the biggest concern with regulatory retention requirements, however, is this: our work makes it clear that one size does not fit all. In the HM equilibrium, optimal retention by the low-reputation bank depends on various bank-specific and loan-specific parameters, such as cost of capital or liquidity β or maximum loan value $P(1)$. These properties will vary not only across banks and loan types but over time as well; for example, our work on HMHL equilibria in the case of stochastic loan demand shows that a change in business conditions can easily shift optimal equilibrium retention shares. It seems unlikely that a blanket requirement such as 5% will be ideal in all cases, and even if it is chosen to do the best job on average across banks and loan types at a given time, changes in financial and economic conditions are likely to drive it away even from such limited optimality.

Figure 1: Description of the stage game in each period

We present below a description of the stage game in each period.

Stage 1

- Borrower obtains a loan of 1 unit from the bank to fund its project.
- Loan rate $R(\overline{q})$ determined based on borrower's conjecture of bank monitoring, $\overline{q}$.

Stage 2

- (Section 4): Bank chooses to retain a fraction α of the loan, and sells the rest in the secondary loan market; $\alpha = 0$ in base model.
- Bank sells loan in the secondary loan market and obtains price $P(\overline{q})$.
- Bank chooses its monitoring, q.

Stage 3

- Project either succeeds w.p. $p + q\Delta$ or fails w.p. $1 - p - q\Delta$.
- Project yields cash flow of $X > 1$ if it succeeds, and bank obtains repayment $R \leq X$.
- Borrower defaults on loan **iff** project fails. In this case, the bank seizes the collateral value $C < 1$.

Appendix

This appendix contains the proofs of all formal results stated in the paper.

Proof of Lemma 1: Suppose $\Lambda > \frac{m}{\delta\Delta} \Rightarrow V_{mon} > V_{shirk}$. Then the bank always strictly prefers to monitor, so $q(0) = q(1) = 1$. But, substituting $q(0) = q(1) = 1$ in the Bellman equation (7) yields $V(0) - V(1) = 0$, which contradicts incentive compatibility. Therefore, it must be that $\Lambda = \frac{m}{\delta\Delta}$.

Substituting $\delta\Delta\Lambda = m$ in the Bellman equation (7) for $V(0)$ and $V(1)$ and differencing these expressions yields $\Lambda = [q(0) - q(1)] A$. As $\Lambda = \frac{m}{\delta\Delta} > 0$, it must be that $q(0) > q(1)$. ■

Proof of Proposition 1: *(1) Characterizing the feasibility condition.*

We showed in Lemma 1 that $\Lambda = [q(0) - q(1)] A = \frac{m}{\delta\Delta}$ in any reputational monitoring equilibrium. Rearranging this equation yields $q(1) = q(0) - \frac{m}{\delta\Delta A}$. But for the equilibrium to be well-defined it is necessary that $q(1) \geq 0$. Substituting $q(0) \leq 1$ yields the necessity of condition (9). Sufficiency follows by noting that the full monitoring equilibrium with $q(0) = 1$ and $q(1) = \hat{q} \equiv 1 - \frac{m}{\delta\Delta A}$ is feasible when this condition is satisfied.

(2) Characterizing the bank's value function, $V(d)$.

Substituting $q(0) = 1$, $\Lambda = \frac{m}{\delta\Delta}$, and $V(1) = V(0) - \frac{m}{\delta\Delta}$ in equation (7) yields

$$V(0) = A + B + \delta V(0) - \frac{m(1-p)}{\Delta}$$

Solving the above equation for $V(0)$ yields $V(0) = V^*$, which is defined in equation (8). Next, it follows from incentives compatibility that $V(1) = V^* - \frac{m}{\delta\Delta}$. ■.

Proof of Lemma 2: We will prove both statements in the lemma by contradiction.

(1) Suppose there exists an equilibrium in which $q(d) \in (0,1)$. It follows that the IC constraint (14) holds with equality, so that $\delta\Delta\Lambda + \alpha(d)\Delta(R(d) - C) - m = 0$. The bank's value function is

$$\begin{aligned} V(d) &\equiv V(d, \alpha(d)) \\ &= (1 + \beta(1 - \alpha(d))) \cdot P(q(d)) - (1+\beta) - mq(d) + \delta(p + \Delta q(d))\Lambda + \delta V(1). \end{aligned}$$

Now suppose the bank deviates by retaining $\alpha' = \alpha(d) + \varepsilon$, where $\varepsilon > 0$. Because the loan rate is already set at $R(d)$, the IC constraint (14) now holds strictly, and hence, the bank's monitoring choice is $q(d,\alpha')=1$. The expression for bank value from the deviation is

$$V(d,\alpha') = (1+\beta(1-\alpha'))\cdot \overline{P}(1,R(d)) - (1+\beta) - m + \delta(p+\Delta)\Lambda + \delta V(1).$$

The bank will strictly prefer to deviate if $V(d,\alpha') > V(d)$, which is equivalent to

$$[1+\beta-\beta\alpha(d)]\cdot\left[\overline{P}(1,R(d)) - P(q(d))\right] - \beta\varepsilon\overline{P}(1,R(d)) + (\delta\Delta\Lambda - m)\cdot[1-q(d)] > 0$$

Substituting $\overline{P}(1,R) - P(q(d)) = [1-q(d)]\cdot\Delta(R(d)-C)$ and $m - \delta\Delta\Lambda = \alpha(d)\Delta(R(d)-C)$, the above condition may be rewritten as

$$(1+\beta)(1-\alpha(d))\cdot\Delta(R(d)-C)\cdot[1-q(d)] > \beta\varepsilon\overline{P}(1,R(d))$$

If $q(d)<1$, the expression on the left-hand side of the above condition is strictly positive, which means that the condition will hold for sufficiently small ε. But this means that the bank strictly prefers the deviation for sufficiently small ε, which contradicts the existence of an equilibrium with $q(d)\in(0,1)$.

(2) Suppose there exists an equilibrium in which $q(d)=0$ and $\alpha(d)>0$. But, then the bank will strictly prefer deviating to retaining $\alpha'=0$ because it can lower its liquidity costs without altering investors' belief of its monitoring, which leads to a contradiction. ■

Proof of Lemma 3: We prove this result by contradiction. Suppose there exists a monitoring equilibrium in which $\alpha(0)\geq\alpha(1)$, $q(0)\leq q(1)$, and $\alpha(1)<\alpha_{pr}$. Based on Lemma 2, there are only two possible monitoring equilibria in which $q(0)\leq q(1)$: either $q(0)=q(1)=1$ or $q(0)=0$ and $q(1)=1$. In either case, it is clear that

$$\begin{aligned} V(1) &= (1+\beta(1-\alpha(1)))\cdot P(1) - (1+\beta) - m + \delta(p+\Delta)\Lambda + \delta V(1), \text{ and} \\ V(0) &\leq (1+\beta(1-\alpha(0)))\cdot P(1) - (1+\beta) - m + \delta(p+\Delta)\Lambda + \delta V(1). \end{aligned}$$

If $\alpha(0)\geq\alpha(1)$, then

$$\Lambda = V(0) - V(1) \leq \beta[\alpha(1)-\alpha(0)]\cdot P(1) \leq 0.$$

But then the IC constraint (14) will fail for $d=1$ because

$$\delta\Delta\Lambda + \alpha(1)\cdot(R(1)-C) < \alpha(1)\cdot(R(1)-C) < m,$$

where the last inequality follows because $\alpha(1) < \alpha_{pr}$. Hence, $q(1) = 0$, which is a contradiction. ■

Proof of Proposition 2: As per Lemma 2, there are only three possible combinations of $q(0)$ and $q(1)$ that can arise in any reputational monitoring equilibrium: (a) $q(0) = 1, q(1) = 0$; (b) $q(0) = q(1) = 1$; or (c) $q(1) = 1, q(0) = 0$. (The fourth possible combination $q(0) = q(1) = 0$ does not involve any monitoring in either reputation state).

(1) We will first prove by contradiction that the combination $\{q(1) = 1, q(0) = 0\}$ *cannot occur in equilibrium.*

Suppose there exists a reputational equilibrium with $q(1) = 1$ and $q(0) = 0$. Now, as per Lemma 2, $q(0) = 0 \Rightarrow \alpha(0) = 0$. Hence, we only need to consider two possibilities with respect to $\alpha(1)$: either $\alpha(1) > 0$ or $\alpha(1) = 0$. We consider each of these cases separately below:

(a): Suppose $\alpha(1) > 0$. The incentive compatibility conditions are $m > \delta\Delta\Lambda$ for the high-reputation bank and $\delta\Delta\Lambda + \alpha(1)\Delta[R(1) - C] \geq m$ for the low-reputation bank. As $\alpha(1) \leq \alpha_{pr}$ we require that $\Lambda > 0$. We can write the high- and low-rep banks' value functions as follows:

$$\begin{aligned} V(0) &= (1+\beta) \cdot P(0) - (1+\beta) + \delta p\Lambda + \delta V(1), \text{ and} \\ V(1) &= (1+\beta-\beta\alpha(1)) \cdot P(1) - (1+\beta) - m + \delta(p+\Delta)\Lambda + \delta V(1), \end{aligned}$$

where $P(0) = p(X - C) - u + C$ and $P(1) = (p + \Delta)(X - C) - u + C$.

Note that if the low-reputation bank deviates to $\{\alpha'(1) = 0, q'(1, \alpha'(1)) = 0\}$, it gets a value of

$$V'(1) = (1+\beta) \cdot \overline{P}(0, R(1)) - (1+\beta) + \delta V(1_{low}).$$

It follows from equation (12) that $V(1_{low}) \geq B + \delta p\Lambda + \delta V(1) = V(0)$. Substituting $V(1_{low}) \geq V(0)$ in the expression for $V'(1)$, it follows that

$$V'(1) \geq (1+\beta) \cdot \overline{P}(0, R(1)) - (1+\beta) + \delta V(0).$$

But note that

$$
\begin{aligned}
V(0) &= (1+\beta)\cdot P(0) - (1+\beta) + \delta p\Lambda + \delta V(1) \\
&= (1+\beta)\cdot P(0) - (1+\beta) + \delta V(0) - \delta\Lambda(1-p) \\
&< V'(1),
\end{aligned}
$$

where the second expression is obtained after substituting $V(1) = V(0) - \Lambda$, and the inequality follows by noting that $\overline{P}(0, R(1)) = p\left(X - C - \frac{u}{p+\Delta}\right) + C > P(0)$. For the low-reputation bank to not prefer this deviation, it must be that $V(1) \geq V'(1) > V(0)$, which implies that $\Lambda < 0$, thus leading to a contradiction.

(b): Suppose $\alpha(1) = 0$. For the high-reputation bank's IC constraint to be satisfied, we need $\delta\Delta\Lambda \geq m$. But for the low-reputation bank's IC constraint to fail, we require $\delta\Delta\Lambda < m$, which is a contradiction. Thus, we have proved that there cannot be a reputational monitoring equilibrium in which $q(1) = 1$ and $q(0) = 0$.

(2) Next, we describe the reputational monitoring equilibria that may be feasible with retention. There are only two possible combinations of $\{q(0), q(1)\}$ to consider, which are as follows:

(a) $q(0) = 1$ and $q(1) = 0$. We refer to this as the "low monitoring" equilibrium. As per Lemma 2, $q(1) = 0 \Rightarrow \alpha(1) = 0$. The incentive compatibility conditions are $\delta\Delta\Lambda + \alpha(0)\Delta[R(1) - C] \geq m$ and $m > \delta\Delta\Lambda$ for the high-reputation and low-reputation banks, respectively. These conditions can be met simultaneously only if $\alpha(0) > 0$.

(b) $q(0) = q(1) = 1$. We refer to this as the "high monitoring" equilibrium. Then, as per Lemma 3, it must be that $\alpha(0) < \alpha(1)$. ■

Proof of Proposition 3: Suppose we have an equilibrium in which $q(1) = q(0) = 1$ and $\alpha(0) < \alpha(1) \leq \alpha_{pr}$. The value function of bank with reputation d is

$$
V(d) = A + B - \alpha(d)\beta P(1) - m + \delta(p+\Delta)\Lambda + \delta V(1). \tag{26}
$$

Hence,

$$
\Lambda = V(0) - V(1) = [\alpha(1) - \alpha(0)] \cdot \beta P(1).
$$

Step 1: We describe the IC and non-deviation constraints.

Because $\alpha(1) > \alpha(0)$, the IC constraint will hold for the low-reputation bank if it holds for the high-reputation bank. Hence, we only need to ensure that

$$\delta\Delta[\alpha(1)-\alpha(0)]\cdot\beta P(1)+\alpha(0)\Delta[R(1)-C]\geq m. \tag{27}$$

We also need to check that neither bank will deviate from its retention strategy. Given that $q(d)=1$, neither bank has an incentive to increase its retention. Therefore, we only need to examine deviations to $\alpha'<\alpha(d)$. But such a deviation will cause the bank to end up in the d_{low} reputation state next period with certainty. Then, as shown above, $q(d,\alpha')=0$ for any $\alpha'<\alpha(d)$ because monitoring with positive probability requires that $\alpha'\Delta[R(1)-C]\geq m$, which cannot hold for $\alpha'<\alpha(d)\leq\alpha_{pr}$. Hence, if the bank wants to deviate to $\alpha'<\alpha(d)$, it may as well deviate to $\alpha'=0$, and get a value of

$$\begin{aligned} V' &= (1+\beta)\left(\overline{P}(0,R(1))-1\right)+\delta V(d_{low}) \\ &= (1+\beta)\left(\overline{P}(0,R(1))-1\right)+\delta\left[\bar{A}+B+\delta p\Lambda+\delta V(1)\right]. \end{aligned}$$

Let $S\equiv A+B-\alpha(1)\beta P(1)-m$ denote the current-period surplus, net of monitoring costs, for the low-reputation bank so that $V(1)=S+\delta(p+\Delta)\Lambda+\delta V(1)$. Then, after substituting $\delta p\Lambda+\delta V(1)=V(1)-S-\delta\Delta\Lambda$, we can rewrite V' as

$$V'=(1+\beta)\left(\overline{P}(0,R(1))-1\right)+\delta\bar{A}+\delta B+\delta V(1)-\delta S-\delta^2\Delta\Lambda$$

Because $V(1)<V(0)$, the high-rep bank will not deviate if the low-rep bank finds it unprofitable to do so. Hence, we require $V(1)\geq V'$, which is equivalent to

$$(1+\delta)(A-\alpha(1)\beta P(1)-m)\geq\delta\bar{A}+\frac{u\Delta(1+\beta)}{p+\Delta}-\delta p\Lambda-\delta(1+\delta)\Delta\Lambda \tag{28}$$

Step 2: We show that any equilibrium in which constraint (27) holds strictly is dominated by an alternative equilibrium in which the constraint binds.

Suppose there exists a feasible equilibrium in which the IC constraint (27) holds strictly. Then, consider an alternative equilibrium with $\tilde{\alpha}(0)=\alpha(0)-\varepsilon$ and $\tilde{\alpha}(1)=\alpha(1)-\varepsilon$ for some $\varepsilon>0$ so that constraint (27) still holds, and hence, $\tilde{q}(0)=\tilde{q}(1)=1$. By construction, $\tilde{\Lambda}=\tilde{V}(0)-\tilde{V}(1)=\Lambda$. Moreover, $\tilde{V}(1)>V(1)\geq V'$, where the first inequality holds because $\tilde{\alpha}(1)<\alpha(1)$ and second inequality holds because the original equilibrium is feasible. Hence, the alternative equilibrium is also feasible and delivers higher bank value in both reputation states.

Step 3: We characterize the feasibility conditions for an equilibrium in which constraint (27) binds with equality.

Let us denote $\alpha(0) = (1-\rho)\alpha_{pr}$ where $0 < \rho \leq 1$, so that $\alpha(0)\Delta[R(1) - C] = (1-\rho)m$. Then, if the IC constraint (27) is to bind, it must be that $\delta\Delta\Lambda = \rho m$. After substituting for Λ and solving for $\alpha(1)$, it follows that:

$$\alpha(1) = \frac{\rho m}{\delta\Delta\beta P(1)} + (1-\rho)\alpha_{pr}.$$

Note that the requirement that $\alpha(1) \leq \alpha_{pr}$ can be satisfied only if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, which (after substituting for α_{pr} and $P(1)$, and rearranging terms) is equivalent to the feasibility condition (15).

Next, substituting $\delta\Lambda = \frac{\rho m}{\Delta}$ and the expression for $\alpha(1)$ in condition (28), and simplifying, we obtain the feasibility condition (16).

If both the feasibility conditions are satisfied, then the high monitoring equilibrium described in the proposition becomes feasible. We obtain the expression for $V(0)$ after substituting $V(1) = V(0) - \Lambda$ in equation (26), and solving the resulting equation for $V(0)$ after substituting $\Lambda = [\alpha(1) - \alpha(0)] \cdot \beta P(1) = \frac{\rho m}{\delta\Delta}$. ■

Proof of Corollary 1: *(1) We first show that the HM equilibrium is more likely to be feasible for higher values of ρ.*

Note that the first feasibility condition (15) does not depend on ρ. Suppose condition (15) is satisfied, so that $\alpha_{pr} \cdot \beta P(1) \geq \frac{m}{\delta\Delta}$. Then it is easy to show that the second feasibility condition (16) is more likely to be satisfied for higher values of ρ. To see this consider the derivative of the LHS of this condition with respect to ρ, which is

$$\begin{aligned} \frac{d(LHS)}{d\rho} &= (1+\delta)\alpha_{pr} \cdot \beta P(1) - \frac{m[(1+\delta)(1-\delta\Delta) - \delta p]}{\delta\Delta} \\ &\geq \frac{m[(1+\delta)\Delta + p]}{\Delta}, \end{aligned}$$

where the inequality follows if $\alpha_{pr} \cdot \beta P(1) \geq \frac{m}{\delta\Delta}$.

On the other hand, it is clear that $\frac{d(RHS)}{d\rho} \leq \delta m$. Hence, it follows that

$$\frac{d(LHS - RHS)}{d\rho} \geq \frac{m[(1+\delta)\Delta + p]}{\Delta} - \delta m = \frac{m(\Delta + p)}{\Delta} > 0.$$

That is, condition (16) is most likely to hold for ρ as high as possible; i.e., for $\rho = 1$.

Since condition (15) does not depend on ρ, this proves that the HM equilibrium with $\rho = 1$ is most likely to be feasible.

(2) Next, we show that $V(0)$ is increasing in ρ whenever the HM equilibrium is feasible.

Note that

$$\frac{dV(0)}{d\rho} = (1-\delta)^{-1} \cdot \left[\alpha_{pr} \cdot \beta P(1) - \frac{m(1-p-\Delta)}{\Delta}\right]$$

If $\alpha_{pr} \cdot \beta P(1) \geq \frac{m}{\delta\Delta}$, it follows that

$$\begin{aligned} \frac{dV(0)}{d\rho} &\geq (1-\delta)^{-1} \cdot \left[\frac{m}{\delta\Delta} - \frac{m(1-p-\Delta)}{\Delta}\right] \\ &= (1-\delta)^{-1} \cdot \frac{m}{\delta\Delta}\left[1-\delta(1-p-\Delta)\right] > 0. \end{aligned}$$

This proves the first part of the corollary. Condition (18) is obtained by substituting $\rho = 1$ in condition (16), and simplifying. It is also easily verified that $[V(0)]_{\rho=1} = V^*$. ■

Proof of Lemma 4: Because IC(γ) is stricter than IC(1), there are five cases to consider: (1) $\Lambda > \gamma m/\delta\Delta \Rightarrow$ IC(1) and IC(γ) both hold strictly; (2) $\Lambda = \gamma m/\delta\Delta \Rightarrow$ IC(γ) binds, IC(1) holds strictly; (3) $\gamma m/\delta\Delta > \Lambda > m/\delta\Delta \Rightarrow$ IC(γ) fails, IC(1) holds strictly; (4) $\Lambda = m/\delta\Delta \Rightarrow$ IC(γ) fails, IC(1) binds; (5) $m/\delta\Delta > \Lambda \Rightarrow$ IC(γ) and IC(1) both fail.

The proof of the lemma boils down to showing that Cases (1), (3), and (5) above are inconsistent with any monitoring in equilibrium.

(a) If Case (1) holds, then $q(d,\ell) = 1$ for all d and ℓ, so the bank monitors with certainty regardless of its reputation or its lending volume. In a normal economy, the bank lends 1 unit and sells this for surplus $S(d,1) = A + B$, so its discounted profits are

$$V_n(d) = A + B - m + \delta[(p+\Delta)\Lambda + V(1)] \equiv V_n, \tag{29}$$

regardless of its reputation d. If the economy is in a boom and the bank lends 1 unit, it also gets $V_b(d) = V_n$; if it lends γ, it gets

$$V_b(d) = \gamma(A + B - m) + \delta[(p+\Delta)\Lambda + V(1)] \equiv V_{b,\gamma} > V_n. \tag{30}$$

If so, the bank always lends γ in a boom, so that $V(d) = \phi V_n + (1-\phi)V_{b,\gamma}$ regardless of reputation d. Thus, $\Lambda = V(0) - V(1) = 0$, contradicting the assumption of Case (1).

(b) If Case (3) holds, then $q(d,1) = 1$ and $q(d,\gamma) = 0$ for any reputation d. It follows immediately that, in a normal economy, the bank gets $V_n(d) = V_n$ as in Case (1).

If the economy is in a boom and $\ell_b(d) = 1$, the bank also monitors with certainty, so it gets $V_b(d) = V_n$, too, regardless of reputation. If instead $\ell_b(d) = \gamma$, the bank does not monitor, so its surplus per unit is B, and it gets

$$V_b(d) = \gamma B + \delta[p\Lambda + V(1)] = V_{b,nm} \tag{31}$$

regardless of reputation.

If $\ell_b(0) = \ell_b(1)$, an analysis similar to that in (a) shows that $V(0) = V(1)$, and so $\Lambda = 0$, contradicting the assumption of Case (3). Thus, we must have $\ell_b(0) \neq \ell_b(1)$.

If $\ell(0) = \gamma$ and $\ell(1) = 1$, then $V_b(0) = V_{b,nm}$ and $V_b(1) = V_n$. But if $V_{b,nm} < V_n$, a bank with reputation $d = 0$ strictly prefers to switch to lending 1, breaking the proposed equilibrium. If $V_{b,nm} > V_n$, a bank with reputation $d = 1$ strictly prefers to switch to lending γ, also breaking the equilibrium. Finally, if $V_{b,nm} = V_n$, then $V(d) = \phi V_n + (1-\phi)V_b(d)$ has the same value for both $d = 0$ and $d = 1$, which leads to $\Lambda = 0$, contradicting the assumption of Case (3).

If instead $\ell(0) = 1$ and $\ell(1) = \gamma$, then the analysis is essentially the same.

(c) If Case (5) holds, then the bank never finds it incentive compatible to monitor, so $q(d, \ell) = 0$ for all d and ℓ, and monitoring never takes place in equilibrium. ∎

Proof of Lemma 5: Suppose $IC\,(\gamma)$ binds so that $\Lambda = \gamma m/\delta\Delta$. Then, $IC\,(1)$ has to hold strictly, so that $q\,(d, 1) = 1$ for $d \in \{0, 1\}$, and $V_n\,(0) = V_n\,(1) \equiv V_n$. Recall that $V\,(d) = \phi V_n\,(d) + (1-\phi)\,V_b\,(d)$. Therefore, if $V_n\,(0) = V_n\,(1) = V_n$, then it follows that

$$\Lambda = (1-\phi)\,(V_b\,(0) - V_b\,(1))\,. \tag{32}$$

It is clear that we cannot have an equilibrium in which the bank lends 1 unit in the boom state, regardless of its reputation ($\ell_b\,(0) = \ell_b\,(1) = 1$), because then $V_b\,(0) = V_b\,(1) = V_n$ and $\Lambda = 0$, thus violating the IC constraint. Hence, we only need to consider the following cases:

(1) Suppose $\ell_b\,(0) = \ell_b\,(1) = \gamma$.

By substituting $\delta\Delta\Lambda = \gamma m$, the bank's value function in the boom state may be rewritten as

$$V_b\,(d) = \gamma \cdot [B + q\,(d, \ell_b\,(d)) \cdot A] + (p\gamma m/\Delta) + \delta V\,(1) \tag{33}$$

Therefore,

$$\Lambda = (1-\phi) \cdot \gamma A \cdot [q\,(0,\gamma) - q\,(1,\gamma)] \leq (1-\phi) \cdot \gamma A,$$

where the inequality follows because $[q(0,\gamma) - q(1,\gamma)] \leq 1$. But since $\Lambda = \gamma m/\delta\Delta$, it is necessary that $m \leq \delta\Delta(1-\phi)A$.

(2) Suppose $\ell_b(0) = \gamma$ *and* $\ell_b(1) = 1$.

In this case, $V_b(1) = V_n$ because the low-reputation bank will lend 1 unit in the boom and monitor strictly. But for the low-reputation bank to prefer lending 1 unit instead of γ units, it must be that

$$V_n \geq \gamma \cdot [B + q(1,\gamma) \cdot A] + (p\gamma m/\Delta) + \delta V(1) \tag{34}$$

Hence, it must be that

$$\Lambda \leq (1-\phi) \cdot \gamma A \cdot [q(0,\gamma) - q(1,\gamma)]. \tag{35}$$

Then, by the same logic as in case (1), $\Lambda = \gamma m/\delta\Delta$ requires that $m \leq \delta\Delta(1-\phi)A$.

(3) Suppose $\ell_b(0) = 1$ *and* $\ell_b(1) = \gamma$.

We will prove, by contradiction, that such an equilibrium cannot exist. Suppose it did. Then $V_b(0) = V_n$ and

$$V_b(1) = \gamma \cdot [B + q(1,\gamma) \cdot A] + (p\gamma m/\Delta) + \delta V(1) \tag{36}$$

For the IC constraint to hold, it must be that $V_n > V_b(1)$. But then, it cannot be incentive-compatible for the low-reputation bank to lend γ units in a boom. Thus, the equilibrium in case (3) cannot exist.

Overall, we have proved that any equilibrium in which $IC(\gamma)$ binds cannot exist if $m > \delta\Delta(1-\phi)A$. ■

Proof of Propositions 4 through 7: We examine equilibria in which $IC(1)$ binds and $IC(\gamma)$ fails, so that the bank will not monitor its loans if it lends γ in a boom; $q(d,\gamma) = \overline{q}(d,\gamma) = 0$ for $d \in \{0,1\}$. Hence, if the bank lends γ units, it obtains a value of

$$V_{\gamma,nm} \equiv \gamma B + (pm/\Delta) + \delta V(1)$$

regardless of its reputation d, where we have exploited the fact that $\delta\Lambda = \frac{m}{\Delta}$ if $IC(1)$ binds On the other hand, if the bank lends 1 units, it obtains a value of

$$V_n(d) = Aq(d,1) + B + (pm/\Delta) + \delta V(1).$$

There are four possible equilibria to consider, based on how the bank chooses its lending volume $\ell \in \{1, \gamma\}$ in a boom, in the high and low reputation states. For each equilibrium, we must check three necessary conditions: (a) IC(1) must bind; (b) in a boom, a high-reputation bank must not want to switch its loan volume choice, $\ell_b(0)$; and (c) in a boom, a low-reputation bank must not want to switch its loan volume choice, $\ell_b(1)$.

Case (1): "Tight credit equilibrium" with $\ell_b(0) = \ell_b(1) = 1$. Condition (a) is $\Lambda = V(0) - V(1) = m/\delta\Delta$. First, it is easy to see that, because loan volumes and thus monitoring choices are the same in normal and boom states, we will have $V_b(d) = V_n(d) = V(d)$. Therefore, $\Lambda = [q(0,1) - q(1,1)]A$. If $IC(1)$ is to bind, it must be that $\Lambda = [q(0,1) - q(1,1)]A = \frac{m}{\delta\Delta} > 0$. Thus, as in the baseline model, we must have $q(0,1) > q(1,1)$ to ensure that $\Lambda > 0$.

To ensure that the high-reputation bank does not switch to $\ell = \gamma$ in the boom state (condition (b)), it must be that $V_n(0) \geq V_{\gamma,nm}$, which is equivalent to $q(0,1)A \geq (\gamma - 1)B$.

Similarly, to ensure that the low-reputation bank does not switch to $\ell = \gamma$ in the boom state (condition (c)), it must be that $V_n(1) \geq V_{\gamma,nm}$, which is equivalent to $q(1,1)A \geq (\gamma - 1)B$.

Because $q(0,1) > q(1,1)$, it is clear that condition (c) is stricter than condition (b). Furthermore, it is most likely to hold when $q(1,1)$ is as high as possible, which corresponds to $q(0,1) = 1$ and $q(1,1) = \hat{q} = 1 - \frac{m}{\delta\Delta A}$ so that $IC(1)$ binds. Substituting $q(1,1) = \hat{q}$ in condition (c), and rearranging terms, yields the feasibility condition $m \leq \delta\Delta(A - (\gamma - 1)B)$. Sufficiency of this condition follows by noting that, when this condition is met, the equilibrium characterized in Proposition 5 is feasible.

Case (2): "Partially tight credit" equilibrium with $\ell_b(0) = 1,\ \ell_b(1) = \gamma$. In this case, $V_b(0) = V_n(0)$ and $V_b(1) = V_{\gamma,nm}$. Note that $\Lambda = \phi[V_n(0) - V_n(1)] + (1-\phi)[V_b(0) - V_b(1)]$. Rearranging terms, it is easily shown that condition (a) is equivalent to

$$q(0,1)A - \phi q(1,1)A - (1-\phi)(\gamma - 1)B = m/\delta\Delta. \tag{37}$$

As in case (1) above, condition (b) requires that $V_n(0) \geq V_{\gamma,nm}$, which is equivalent to $q(0,1)A \geq (\gamma - 1)B$. On the other hand, to ensure that the low-reputation bank does not switch to $\ell = 1$ in the boom state (condition (c)), it must be that $V_n(1) \leq V_{\gamma,nm}$, which is equivalent to $q(1,1)A \leq (\gamma - 1)B$.

Clearly, for condition (a) to be satisfied, it is necessary that $A \geq (1-\phi)(\gamma-1)B + m/\delta\Delta$ (condition "F1") Also, condition (b) requires that $A \geq (\gamma - 1)B$ (condition "F2"). This proves the necessity of these two feasibility conditions.

To prove sufficiency, we show that if conditions F1 and F2 above are met, then we can find monitoring probabilities $q(0,1)$ and $q(1,1)$ that are consistent with conditions (a), (b), and (c). Consider $q(0,1) = \hat{q}_{0,ptc} \equiv \min\left\{1, \left[\frac{(\gamma-1)B}{A} + \frac{m}{\delta\Delta A}\right]\right\}$ and $\hat{q}_{1,ptc} = \phi^{-1}\left[\hat{q}_{0,ptc} - \frac{m}{\delta\Delta A} - \frac{(1-\phi)(\gamma-1)B}{A}\right]$, which have been constructed such that condition (a) is satisfied. Moreover, it is clear that $0 < \hat{q}_{0,ptc} \leq 1$. It only remains to be shown that $\hat{q}_{1,ptc} \in [0,1]$ and that conditions (b) and (c) are satisfied. Consider the following two subcases:

(i) Suppose $A \geq (\gamma-1)B + m/\delta\Delta$ so that $\hat{q}_{0,ptc} = (\gamma-1)B/A + m/\delta\Delta A \leq 1$. Then it must be that $\hat{q}_{1,ptc} = (\gamma-1)B/A < 1$, where the inequality follows from condition F2. Condition (c) is met because $\hat{q}_{1,ptc}A = (\gamma-1)B$. Also, condition (b) is satisfied because $\hat{q}_{0,ptc}A = (\gamma-1)B + m/\delta\Delta > (\gamma-1)B$.

(ii) Suppose $A < (\gamma-1)B + m/\delta\Delta$ so that $\hat{q}_{0,ptc} = 1$. Then $\hat{q}_{1,ptc} > 0$ by condition F1. Also, because $\hat{q}_{0,ptc} < (\gamma-1)B/A + m/\delta\Delta A$ in this case, $\hat{q}_{1,ptc} < (\gamma-1)B/A$, which implies that condition (c) is satisfied. Finally, condition (b) is also satisfied because $\hat{q}_{0,ptc}A = A \geq (\gamma-1)B$ by condition F2.

Case (3): "Loose credit equilibrium" with $\ell_b(0) = \ell_b(1) = \gamma$. In this case, $V_n(d)$ is the same as in Case (1) above, but $V_b(d) = V_{\gamma,nm}$. Then, $\Lambda = V(0) - V(1) = \phi[V_n(0 - V_n(1)] = \phi[q(0,1) - q(1,1)] \cdot A$. If $IC\,(1)$ is to be met, it must be that $\phi[q(0,1) - q(1,1)] \cdot A \geq \frac{m}{\delta\Delta} > 0$, which requires that $q(0,1) > q(1,1)$.

To ensure that the high-reputation bank does not switch to $\ell = 1$ in the boom state (condition (b)), we need $V_n(0) \leq V_{\gamma,nm}$, which is equivalent to $q(0,1)A \leq (\gamma-1)B$.

To ensure that the low-reputation bank does not switch to $\ell = 1$ in the boom state (condition (c)), we need $V_n(1) \leq V_{\gamma,nm}$, which is equivalent to $q(1,1)A \leq (\gamma-1)B$. Because $q(0,1) > q(1,1)$, condition (c) will be satisfied if condition (b) is satisfied.

For condition (a) to be satisfied, it is necessary that $\phi A \geq \frac{m}{\delta\Delta}$, that is, $m \leq \phi\delta\Delta A$. In addition, it is also necessary that $q(0,1)\phi A \geq \frac{m}{\delta\Delta}$; combining this with the requirement that $(\gamma-1)B \geq q(0,1)A$ (condition (b)) yields the necessity of the condition $\phi(\gamma-1)B \geq \frac{m}{\delta\Delta}$, or equivalently, $m \leq \phi\delta\Delta(\gamma-1)B$. These two necessary conditions can be combined into a single necessary condition, $m \leq \phi\delta\Delta \min\{A, (\gamma-1)B\}$.

To prove sufficiency, suppose $m \leq \phi\delta\Delta \min\{A, (\gamma-1)B\}$. Then, consider the monitoring probabilities $q(0,1) = \min\{1, \frac{(\gamma-1)B}{A}\}$ and $q(1,1) = \min\{1, \frac{(\gamma-1)B}{A}\} - \frac{m}{\phi\delta\Delta A}$, which have been constructed so that condition (a) is satisfied. The feasibility condition guarantees that $q(1,1) \geq 0$, which also implies that $q(0,1) > 0$ so that these probabilities are well defined

($q(1,1) < q(0,1) \leq 1$ by construction). Next, both conditions (b) and (c) are satisfied because $q(0,1)A = \min\{A, (\gamma-1)B\} \leq (\gamma-1)B$.

Case (4): $\ell_b(0) = \gamma,\ \ell_b(1) = 1$. In this case, $V_b(0) = V_{\gamma,nm}$ and $V_b(1) = V_n(1)$. By a similar logic as in case (2), condition (a) becomes

$$\phi[q(0,1) - q(1,1)] \cdot A + (1-\phi)[(\gamma-1)B - q(1,1)A] = \frac{m}{\delta\Delta}. \tag{38}$$

To ensure that the high-reputation bank does not switch to $\ell = 1$ in the boom state (condition (b)), it must be that $V_n(0) \leq V_{\gamma,nm}$, which is equivalent to $q(0,1)A \leq (\gamma-1)B$.

Similarly, to ensure that the low-reputation bank does not switch to $\ell = \gamma$ in the boom state (condition (c)), it must be that $V_n(1) \geq V_{\gamma,nm}$, which is equivalent to $q(1,1)A \geq (\gamma-1)B$. But then, condition (a) can only hold if $\phi[q(0,1) - q(1,1)] \cdot A \geq m/\delta\Delta$, which requires $q(0,1) > q(1,1)$. Thus, it must be that $q(0,1)A > q(1,1)A \geq (\gamma-1)B$, where the last inequality follows from condition (c). But this contradicts condition (b). Hence, it follows that Case (4) cannot exist. ■

Proof of Proposition 8: Let $\hat{\gamma}_{tc} \equiv \frac{(A-(m/\delta\Delta))}{B} + 1$ and $\hat{\gamma}_{ptc}(\phi) \equiv \frac{(A-(m/\delta\Delta))}{(1-\phi)B} + 1$ denote the levels of γ at which the feasibility conditions for the TC equilibrium and PTC equilibrium, respectively, bind with equality. Hence, the TC equilibrium is infeasible if $\gamma > \hat{\gamma}_{tc}$ and the PTC equilibrium is infeasible if $\gamma > \hat{\gamma}_{ptc}(\phi)$. Note that $\hat{\gamma}_{ptc}(\phi)$ increases as ϕ increases, and $\hat{\gamma}_{ptc}(\phi) > \hat{\gamma}_{tc}$.

We also make use of the following expressions for $V_{tc}(0)$, $V_{ptc}(0)$ and $V_{lc}(0)$ in the remainder of this proof:

$$V_{tc}(0) = (1-\delta)^{-1} \cdot \left(A + B - \frac{m(1-p)}{\Delta}\right), \tag{39}$$

$$V_{ptc}(0) = (1-\delta)^{-1} \cdot \left(A \cdot \hat{q}_{0,ptc} + B - \frac{m(1-p)}{\Delta}\right), \tag{40}$$

and

$$V_{lc}(0) = (1-\delta)^{-1} \cdot \left(\phi A \hat{q}_{0,lc} + \phi B + (1-\phi)\gamma B - \frac{m(1-p)}{\Delta}\right) \tag{41}$$

We now provide a pair-wise comparison of the TC, PTC, and LC equilibria.

(a) The TC equilibrium vs. the PTC equilibrium: Note that the PTC equilibrium is feasible whenever the TC equilibrium is feasible ($\hat{\gamma}_{tc} < \hat{\gamma}_{ptc}(\phi)$). Comparing the expressions for $V_{tc}(0)$ and $V_{ptc}(0)$, it follows that $V_{tc}(0) \geq V_{ptc}(0)$ and the inequality is strict if $\hat{q}_{0,ptc} < 1$,

i.e., if $(\gamma - 1)B/A + m/\delta\Delta A < 1$. Hence, it follows that: (i) TC equilibrium (weakly) dominates the PTC equilibrium if its feasible, i.e., if $\gamma \leq \hat{\gamma}_{tc}$; and (ii) the PTC equilibrium dominates the TC equilibrium only if the former is feasible but the latter is not, i.e., if $\hat{\gamma}_{tc} < \gamma \leq \hat{\gamma}_{ptc}(\phi)$.

(b) The TC equilibrium vs. the LC equilibrium.

Clearly, the LC equilibrium dominates the TC equilibrium if $\gamma > \hat{\gamma}_{tc}$, because LC equilibrium is feasible in this region whereas the TC equilibrium is not.

Suppose instead the parameter values are such that both these equilibria are feasible.[14] Then it must be that $\hat{q}_{0,lc} = \frac{(\gamma-1)B}{A}$ because feasibility of TC equilibrium requires that $(\gamma - 1)B < A$. Substituting for $\hat{q}_{0,lc}$ in equation (41), it is easily verified that $V_{tc}(0) > V_{lc}(0)$ **iff** $A > (\gamma - 1)B$, which is clearly satisfied if the TC equilibrium is feasible. Hence, whenever it is feasible, the TC equilibrium (weakly) dominates the LC equilibrium.

(c) The PTC equilibrium vs. the LC equilibrium:

Clearly, the LC equilibrium dominates the PTC equilibrium if $\gamma > \hat{\gamma}_{ptc}(\phi)$, because LC equilibrium is feasible in this region whereas the PTC equilibrium is not.

Suppose instead that $\gamma \leq \hat{\gamma}_{ptc}(\phi)$ so that the PTC equilibrium is feasible. Then it must be that $\hat{q}_{0,lc} = \frac{(\gamma-1)B}{A}$ because feasibility of PTC equilibrium requires that $(\gamma - 1)B < A$. Substituting $\hat{q}_{0,lc} = \frac{(\gamma-1)B}{A}$ in equation (41) and comparing with equation (40), it follows that $V_{ptc}(0) \geq V_{lc}(0)$ iff $A\hat{q}_{0,ptc} \geq (\gamma - 1)B$. After substituting for $\hat{q}_{0,ptc}$, this condition is equivalent to $\min\{A, [(\gamma - 1)B + m/\delta\Delta]\} \geq (\gamma - 1)B$, which is clearly satisfied because feasibility of the PTC equilibrium implies that $(\gamma - 1)B \leq A$. Hence, we have shown that, whenever it is feasible, the PTC equilibrium (weakly) dominates the TC equilibrium.

Combining parts (a) and (c), it follows that the TC equilibrium is the dominant equilibrium if $\gamma \leq \hat{\gamma}_{tc}$, the PTC equilibrium is the dominant equilibrium if $\hat{\gamma}_{tc} < \gamma \leq \hat{\gamma}_{ptc}(\phi)$, and the LC equilibrium is the dominant equilibrium only if the TC and PTC equilibria are infeasible. ∎

[14] We verify whether it is possible for both these equilibria to even co-exist. The TC equilibrium is feasible only if $(\gamma - 1)B \leq A - \frac{m}{\delta\Delta}$, whereas the LC equilibrium is feasible only if $(\gamma - 1)B \geq \frac{m}{\phi\delta\Delta}$ (where we are exploiting the fact that $\frac{(\gamma-1)B}{A} < 1$ if the TC equilibrium is feasible). Combining the two conditions, we must have $\frac{m}{\phi\delta\Delta} \leq A - \frac{m}{\delta\Delta}$, which is equivalent to,

$$m \leq \frac{\delta\Delta\phi A}{1 + \phi}. \tag{42}$$

But Assumption 4 requires that $m > \delta\Delta(1 - \phi)A$. Therefore, Assumption 4 and (42) can be jointly satisfied only if $1 - \phi^2 - \phi < 0$ (i.e., if ϕ exceeds the positive root of this quadratic equation, which is approximately 0.618). Hence, there may be a narrow set of parameter values for which both these equilibria may co-exist.

Proof of Proposition 9: Under the HMHL equilibrium, $\ell_b(d) = \gamma$ for all d, and $q(d,\ell) = 1$ for all d and ℓ. Hence,

$$V_n(d) = A + B - m - \alpha(d,1)\beta P(1) + \delta(p+\Delta)\Lambda + \delta V(1),$$

and

$$V_b(d) = \gamma \cdot [A + B - m - \alpha(d,\gamma)\beta P(1)] + \delta(p+\Delta)\Lambda + \delta V(1).$$

Then, we have the following expression for $\Lambda \equiv V(0) - V(1)$:

$$\begin{aligned} \Lambda &= \phi \cdot [V_n(0) - V_n(1)] + (1-\phi) \cdot [V_b(0) - V_b(1)] \\ &= [\phi \Delta_{\alpha,1} + (1-\phi)\gamma \cdot \Delta_{\alpha,\gamma}] \cdot \beta P(1), \end{aligned} \tag{43}$$

where $\Delta_{\alpha,1} \equiv \alpha(1,1) - \alpha(0,1)$ denotes the difference in retention between the low- and high-reputation banks with $\ell = 1$, and $\Delta_{\alpha,\gamma} \equiv \alpha(1,\gamma) - \alpha(0,\gamma)$ denotes the difference in retention between the low- and high-reputation banks when $\ell = \gamma$.

Non-deviation constraint for retention strategies: We need to check that the bank does not deviate to a no-retention strategy with zero monitoring after the loan rate has been set at $R(1)$. Since deviation is more likely in the low reputation state, it is sufficient that these non-deviation constraints are satisfied in the low reputation state. In the normal economy, the non-deviation constraint requires that

$$V_n(1) \geq (1+\beta)\left(\bar{P}(0, R(1)) - 1\right) + \delta V(d_{low}).$$

where, using a similar logic used in Section 5, it follows that

$$V(d_{low}) = \bar{A} + \gamma_\phi B + \delta p \Lambda + \delta V(1).$$

Note that we can write $V(1)$ as

$$V(1) = Y + \gamma_\phi B + \delta(p+\Delta)\Lambda + \delta V(1),$$

where

$$Y \equiv \phi[A - m - \alpha(1,1)\beta P(1)] + (1-\phi)\gamma \cdot [A - m - \alpha(1,\gamma)\beta P(1)]. \tag{44}$$

Hence, we can rewrite the expression for $V\left(d_{low}\right)$ as

$$V\left(d_{low}\right)=\bar{A}+V\left(1\right)-Y-\delta\Delta\Lambda$$

After substituting for $V_n\left(1\right)$ and $V\left(d_{low}\right)$ in the non-deviation condition, and simplifying, we obtain the following condition:

$$A-m-\alpha\left(1,1\right)\beta P\left(1\right)+\delta Y\geq\delta\bar{A}+\frac{u\Delta\left(1+\beta\right)}{p+\Delta}-\delta p\Lambda-\delta\left(1+\delta\right)\Delta\Lambda \tag{45}$$

Similarly, in the boom state with $\ell=\gamma$, the non-deviation constraint requires that

$$V_b\left(1\right)\geq\gamma\left(1+\beta\right)\left(\bar{P}\left(0,R\left(1\right)\right)-1\right)+\delta V\left(d_{low}\right).$$

After substituting for $V_b\left(1\right)$ and $V\left(d_{low}\right)$, and simplifying, we can rewrite this condition as

$$\gamma\cdot\left[A-m-\alpha\left(1,\gamma\right)\beta P\left(1\right)\right]+\delta Y\geq\delta\bar{A}+\gamma\frac{u\Delta\left(1+\beta\right)}{p+\Delta}-\delta p\Lambda-\delta\left(1+\delta\right)\Delta\Lambda \tag{46}$$

Non-deviation constraint for the lending volume strategy: We also need to verify that the bank will not deviate to $\ell=1$ and $\alpha=\alpha\left(d,1\right)$ in the boom state (we do not have to worry about deviation to $\ell=1,\alpha=0$, and $q=0$ because such a deviation is dominated by a deviation to $\ell=\gamma,\alpha=0$ and $q=0$, which we have considered above and ruled out through condition (46)). This requires that $V_b\left(d\right)\geq V_n\left(d\right)$, which is equivalent to

$$\left(\gamma-1\right)\left(A+B-m\right)+\left[\alpha\left(d,1\right)-\gamma\alpha\left(d,\gamma\right)\right]\beta P\left(1\right)\geq0\text{ for }d\in\{0,1\} \tag{47}$$

Characterizing $\alpha(d,\ell)$**:** By the same logic as in Section 4, it follows that the most efficient equilibria are those in which the IC constraints bind with equality in the high reputation state and hold strictly in the low reputation state (if not, it should be possible to lower $\alpha\left(0,\ell\right)$ and $\alpha\left(1,\ell\right)$ by a small $\varepsilon>0$ so that all the feasibility conditions are met and the bank value is higher). Hence, it must be that

$$\delta\Delta\left[\phi\Delta_{\alpha,1}+\left(1-\phi\right)\gamma\Delta_{\alpha,\gamma}\right]\cdot\beta P\left(1\right)+\alpha\left(0,1\right)\Delta\cdot\left[R\left(1\right)-C\right]=m, \tag{48}$$

and

$$\delta\Delta\left[\phi\Delta_{\alpha,1}+\left(1-\phi\right)\gamma\Delta_{\alpha,\gamma}\right]\cdot\beta P\left(1\right)+\gamma\alpha\left(0,\gamma\right)\Delta\cdot\left[R\left(1\right)-C\right]=\gamma m \tag{49}$$

Suppose we set $\alpha(0,1) = (1-\rho)\,\alpha_{pr}$ so that $\delta\Delta\Lambda = \rho m$. Then it follows from equation (49) that

$$\alpha(0,\gamma) = \frac{(\gamma-\rho)\,\alpha_{pr}}{\gamma} > \alpha(0,1)\,. \tag{50}$$

Without loss of generality, we can set $\Delta_{\alpha,1} = \gamma\Delta_{\alpha,\gamma} \equiv \Delta_\alpha$ so that $\Lambda = \beta P(1)\,\Delta_\alpha$. Since $\delta\Delta\Lambda = \rho m$, it must be that $\Delta_\alpha = \frac{\rho m}{\delta\Delta\beta P(1)}$, which in turn implies that

$$\alpha(1,1) = \frac{\rho m}{\delta\Delta\beta P(1)} + (1-\rho)\,\alpha_{pr}, \tag{51}$$

and

$$\alpha(1,\gamma) = \frac{1}{\gamma}\left(\frac{\rho m}{\delta\Delta\beta P(1)} + (\gamma-\rho)\,\alpha_{pr}\right) \tag{52}$$

Characterizing the feasibility conditions: It is easily shown that the feasibility condition $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$ is necessary to ensure that $\alpha(1,1) \leq \alpha_{pr}$ and $\alpha(1,\gamma) \leq \alpha_{pr}$.

Next, we need to verify that the above values of $\alpha(d,\ell)$ satisfy the non-deviation constraints (45) and (46). After substituting for $\alpha(1,1)$ and $\alpha(1,\gamma)$, the expression for Y simplifies to

$$Y = \gamma_\phi \cdot (A - m - \alpha_{pr}\beta P(1)) + \rho\left(\alpha_{pr}\beta P(1) - \frac{m}{\delta\Delta}\right) \tag{53}$$

Hence, the feasibility condition (45) can be rewritten as the following condition on ρ:

$$\left[\begin{array}{c} (1+\delta\gamma_\phi)\cdot(A - m - \alpha_{pr}\beta P(1)) - \frac{u\Delta(1+\beta)}{p+\Delta} \\ -\delta\bar{A} + (1+\delta)\,\rho\alpha_{pr}\beta P(1) - \frac{\rho m[(1+\delta)(1-\delta\Delta)-\delta p]}{\delta\Delta} \end{array}\right] \geq 0 \tag{54}$$

Similarly, condition (46) can also be rewritten as a condition on ρ as follows:

$$\left[\begin{array}{c} (\gamma+\delta\gamma_\phi)\cdot(A - m - \alpha_{pr}\beta P(1)) - \gamma\frac{u\Delta(1+\beta)}{p+\Delta} \\ -\delta\bar{A} + (1+\delta)\,\rho\alpha_{pr}\beta P(1) - \frac{\rho m[(1+\delta)(1-\delta\Delta)-\delta p]}{\delta\Delta} \end{array}\right] \geq 0 \tag{55}$$

Note that if $A - m - \alpha_{pr}\beta P(1) \geq \frac{u\Delta(1+\beta)}{p+\Delta}$, then condition (54) is the more binding constraint. Moreover if $A - m - \alpha_{pr}\beta P(1) \geq \frac{u\Delta(1+\beta)}{p+\Delta}$ and $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, it is easily shown that condition (54) is automatically met because the right-hand side expression is positive.[15]

[15] This is because $\delta\gamma_\phi A > \alpha A > \delta\bar{A}$ (where the last inequality follows because $\bar{\alpha} > \alpha_{pr}$), and if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, then

$$\begin{aligned} (1+\delta)\,\rho\alpha_{pr}\beta P(1) - \frac{\rho m\left[(1+\delta)(1-\delta\Delta)-\delta p\right]}{\delta\Delta} &> (1+\delta)\,\rho\cdot\left(\alpha_{pr}\beta P(1) - \frac{m}{\delta\Delta}\right) \\ &\geq 0. \end{aligned}$$

On the other hand, condition (55) is more binding if $A - m - \alpha_{pr}\beta P(1) < \frac{u\Delta(1+\beta)}{p+\Delta}$. In this case, condition (55) is more likely to be met for high γ and high ρ (because $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$ implies that the RHS is increasing in ρ)

Finally, for both $d \in \{0,1\}$, it is easily verified that condition (47) simplifies to the following condition:

$$A + B - m - \alpha_{pr}\beta P(1) \geq 0. \tag{56}$$

This condition is automatically met if $A - m - \alpha_{pr}\beta P(1) \geq \frac{u\Delta(1+\beta)}{p+\Delta}$.

Characterizing the value functions: Note that $V(0) = \phi V_n(0) + (1-\phi) V_b(0)$. After substituting all the equilibrium parameters, and simplifying, we obtain the following expression for $V(0)$:

$$V(0) = \gamma_\phi \cdot [A + B - m - \alpha_{pr}\beta P(1)] + \rho\alpha_{pr}\beta P(1) + \delta(p+\Delta)\Lambda + \delta V(1).$$

After substituting $V(1) = V(0) - \Lambda$ and $\Lambda = \frac{\rho m}{\delta\Delta}$ in the above equation, and solving for $V(0)$, we obtain the expression in equation (25). ■

Proof of Corollary 2: The corollary follows by noting that if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, then $V(0)$ is increasing in ρ. Moreover, we show in the proof of Proposition 9 that the more general feasibility conditions for the existence of the HMHL equilibria are more likely to be satisfied for higher values of ρ.

Next, we will prove that if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$ and $A - m - \alpha_{pr}\beta P(1) \geq \frac{u\Delta(1+\beta)}{p+\Delta}$, then condition (18) is also satisfied, so that the HM equilibrium is feasible.

First, note that if $\alpha_{pr} \geq \frac{m}{\delta\Delta\beta P(1)}$, then $A - m - \alpha_{pr}\beta P(1) \leq A - \frac{m(1+\delta\Delta)}{\delta\Delta}$. Hence, it must be that

$$A - \frac{m(1+\delta\Delta)}{\delta\Delta} \geq \frac{u\Delta(1+\beta)}{p+\Delta}, \tag{57}$$

which can only be met if $\delta\Delta A > m$. But,

$$\begin{aligned} A - \frac{m(1+\delta\Delta)}{\delta\Delta} &= A - \frac{m(1+\delta(1-p))}{\delta\Delta} + \frac{m(1-p-\Delta)}{\Delta} \\ &< A - \frac{m(1+\delta(1-p))}{\delta\Delta} + \frac{m}{\Delta} \qquad (58) \\ &< A - \frac{m(1+\delta(1-p))}{\delta\Delta} + \delta A, \qquad (59) \end{aligned}$$

where the first inequality follows by noting that $1 - p - \Delta < 1$ and the second inequality follows because $\delta\Delta A > m$ (as shown above).

Consider the following two cases:

(1) Suppose $A \leq \beta\bar{\alpha}\overline{P}(1, R(0))$. Then, combining conditions (57) and (59), it follows that condition (18) is met in this case.

(2) Suppose $A > \beta\bar{\alpha}\overline{P}(1, R(0))$. In this case, condition (18) requires that

$$A - \frac{m(1+\delta-\delta p)}{\delta\Delta} + \delta\beta\bar{\alpha}\overline{P}(1, R(0)) \geq \frac{u\Delta(1+\beta)}{(p+\Delta)}$$

But,

$$\begin{aligned} \beta\bar{\alpha}\overline{P}(1, R(0)) &= \beta\frac{m}{\Delta}\left[(p+\Delta) + \frac{C}{\left(X - C - \frac{u}{p}\right)}\right] \\ &> \alpha_{pr} \cdot \beta P(1) \\ &\geq \frac{m}{\delta\Delta}, \end{aligned}$$

if condition (15) is met. Therefore, it follows from conditions (57) and (58) that condition (18) is met in this case as well. ■

References

Abreu, D. 1986. Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* 39:191–225.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, and D. Evanoff. 2011. The role of securitization in mortgage renegotiation. *Journal of Financial Economics* 102:559–578.

Bar-Isaac, H., and J. Shapiro. 2013. Ratings quality over the business cycle. *Journal of Financial Economics* 108:62–78.

Bebchuk, L., and I. Goldstein. 2011. Self-fulfilling credit market freezes. *Review of Financial Studies* 24:3519–3555.

Benmelech, E., J. Dlugosz, and V. Ivashina. 2012. Securitization without adverse selection: The case of CLOs. *Journal of Financial Economics* 106:91–113.

Bolton, P., X. Freixas, and J. Shapiro. 2007. Conflicts of interest, information provision, and competition in the financial services industry. *Journal of Financial Economics* 85:297–330.

Bolton, P., X. Freixas, and J. Shapiro. 2012. The credit ratings game. *Journal of Finance* 67:85–112.

Bord, V., and J. Santos. 2015. Does securitization of corporate loans lead to riskier lending? *Journal of Money, Credit, and Banking* 47:415–444.

Boyd, J. H., and E. C. Prescott. 1986. Financial intermediary-coalitions. *Journal of Economic Theory* 38:211–232.

Chen, Z., A. D. Morrison, and W. J. Wilhelm, Jr. 2015. Traders vs. relationship managers: Reputational conflicts in full-service investment banks. *Review of Financial Studies* 28:1153–1198.

Dellarocas, C. 2005. Reputation mechanism design in online trading environments with pure moral hazard. *Information Systems Research* 16:209–230.

DeMarzo, P., and D. Duffie. 1999. A liquidity-based model of security design. *Econometrica* 67:65–99.

Dewatripont, M., and J. Tirole. 1995. *The prudential regulation of banks.* MIT Press.

Diamond, D. 1984. Financial intermediation and delegated monitoring. *Review of Economic Studies* 51:394–414.

Dinc, S. 2000. Bank reputation, bank commitment, and the effects of competition in credit markets. *Review of Financial Studies* 13:781–812.

Gopalan, R., V. Nanda, and V. Yerramilli. 2011. Does poor performance damage the reputation of financial intermediaries? Evidence from the loan syndication market. *Journal of Finance* 66:2083–2120.

Gorton, G., and G. Pennacchi. 1995. Banks and loan sales: Marketing non-marketable assets. *Journal of Monetary Economics* 35:389–411.

Green, E. J., and R. H. Porter. 1984. Noncooperative collusion under imperfect price information. *Econometrica* 52:87–100.

Hartman-Glaser, B. 2017. Reputation and signaling in asset sales. *Journal of Financial Economics* 125:245–265.

Holmström, B., and J. Tirole. 1997. Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics* 112:663–691.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2010. Did securitization lead to lax screening? Evidence from subprime loans. *Quarterly Journal of Economics* 125:307–362.

Klein, B., and K. B. Leffler. 1981. The role of market forces in assuring contractual performance. *Journal of Political Economy* 89:615–641.

Kreps, D. M., and R. Wilson. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27:253–279.

Leland, H. E., and D. H. Pyle. 1977. Informational asymmetries, financial structure and financial intermediation. *Journal of Finance* 32:371–387.

Mailath, G. J., and L. Samuelson. 2006. *Repeated games and reputations*. Oxford University Press.

Milgrom, P., and J. Roberts. 1982. Predation, reputation, and entry deterrence. *Journal of Economic Theory* 27:280–312.

Parlour, C. A., and A. Winton. 2013. Laying off credit risk: Loan sales versus credit default swaps. *Journal of Financial Economics* 107:25–45.

Pennacchi, G. 1998. Loan sales and the cost of bank capital. *Journal of Finance* 43:375–396.

Piskorski, T., A. Seru, and V. Vig. 2010. Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. *Journal of Financial Economics* 97:369–397.

Purnanandam, A. 2011. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies* 24:1881–1915.

Pyles, M. K., and D. J. Mullineax. 2008. Constraints on loan sales and the price of liquidity. *Journal of Financial Services Research* 33:21–36.

Ramakrishnan, R., and A. Thakor. 1984. Information reliability and a theory of financial intermediation. *Review of Economic Studies* 51:415–432.