RIETI Discussion Paper Series 15-E-121

Loan Monitoring and Bank Risk

Norvald INSTEFJORD
University of Essex

NAKATA Hiroyuki
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 15-E-121
October 2015

Loan Monitoring and Bank Risk*

Norvald INSTEFJORD† and NAKATA Hiroyuki‡

Abstract

We study two issues: the relationship between loan monitoring and loan risk and the effects of regulations on banks' incentives for investments in loan monitoring systems. We describe dynamic monitoring of loans as an optimal stopping problem where the bank stops monitoring loans when it has become sufficiently certain that the loan is of good or bad quality. This process increases the incentive to hold risky loans, which in turn increases the cost of regulatory compliance when the regulator seeks to limit the risk taken by banks. The profitability of improved monitoring must be balanced against the increase in the cost of regulation, and we show that the trade off is always negative. This can explain the trend in banking of switching away from the monitoring of existing loans and instead investing in credit scoring systems, which can improve the initial lending decision, but eliminates certain classes of borrowers.

Keywords: Credit scoring, Dynamic monitoring, Loan risk, Loan sales, Monitoring systems, Optimal stopping.
JEL classification: G21, G28, G32

* This study is conducted as part of research at Research Institute of Economy, Trade and Industry (RIETI).

† Essex Business School, University of Essex
‡ School of Management, University of Leicester, and Research Institute of Economy, Trade and Industry

1 Introduction

This paper studies the effects of loan monitoring systems on bank risk and benefits, and the effects of regulations on banks' incentives for investment in loan monitoring systems.[1] Some attribute bank failures to underinvestment in monitoring. For instance, the Basel Committee asserts "that a significant cause of bank failures is poor credit quality and credit risk assessment. Failure to identify and recognise deterioration in credit quality in a timely manner can aggravate and prolong the problem. Thus, inadequate credit risk assessment policies and procedures, which may lead to inadequate and untimely recognition and measurement of loan losses, undermine the usefulness of capital requirements and hamper proper assessment and control of a bank's credit risk exposure."[2]

It is however a widely held assertion that monitoring of loans reduces loan risk. The following example illustrates this assertion and its limitations. Suppose a bank either makes a safe lending decision or a risky lending decision. Prior to making the decision, the bank can invest in screening of the borrowers. The liability of the bank is 100, and the figure describes the consequences of the firm's decision conditional on the state of nature. These consequences are given at the bottom of the trees by the pair (Loan Value, Bank's Equity Value).

The tree on the left, (a) in Figure 1, describes the situation in which the bank does not acquire any information through screening and makes the lending decision without knowing the true quality of the risky loan. Assuming that the bank is risk neutral, the shareholders

[1] The terms "screening" and "monitoring" appear to be used interchangeably in this context, however screening is sometimes used to describe the information gathering associated with new loans and monitoring the information gathering associated with existing loans. Screening has also a specific meaning in adverse selection models in economics where it means the act of inducing self-selection or separation of hidden types by presenting the decision makers with a choice from a given set of actions, which is not applied here. Finally, some papers use monitoring to mean information gathering about the loan *plus* value enhancing intervention in the risk management of the loan. In this paper monitoring simply means information gathering, as me model separately the risk management decisions associated with the loan given the information gathering process.

[2] Basel Committee on Banking Supervision: Sound credit risk assessment and valuation for loans, June 2006.

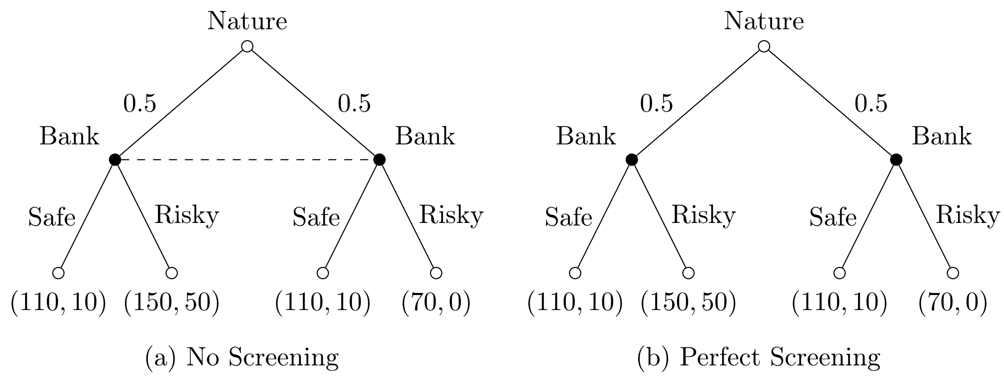

Figure 1: Effects of Loan Screening

prefer the bank lends risky since 25 (the ex ante equity value of risky lending, the average of 50 and 0) is greater than the 10 (the equity value of safe lending). The regulator however prefers the bank makes a safe lending decision since this prevents the bailout of the bank in the bad outcome. The regulator can make the bank buy insurance against the bailout at a cost of 15 (the expected value of the bailout which is one half probability times the bailout cost which represents the shortfall in default, or $100 - 70 = 30$). The shareholder value net of the regulation cost is $25 - 15 = 10$, which is the same as the value of safe lending and the equity holders are indifferent.

The tree on the right, (b) in Figure 1, describes the situation in which the bank acquires information that perfectly identifies the true quality of the risky loan. In this case, the bank will lend risky if and only if the true quality of the risky loan is good, with a loan value of 150. In this case, there are no conflicts of interests among the bank, the regulator and the shareholders, and there is no bailout risk. The shareholder value is 30 (the average of 50 and 10) minus the cost of information. If the cost of the monitoring system is less than 20, therefore, the bank's shareholders prefer to invest in monitoring, and the monitoring system reduces the loan risk.

The crucial feature of the above example is that the investment in monitoring is done before the lending decision is made, and that the relevant information arrives sufficiently rapidly to enable the bank to use the information when the lending decision is made. A credit scoring system is an example of a system of this kind.

It is natural to assume however there would be useful new information about the credit quality of the borrower available also after the lending decision is made, and that the bank can invest in such information as well to make valuable improvements to the risk management of the loan. The systems that collect this type of information we call dynamic monitoring systems. An example of such a system is relationship banking, where the bank builds a relationship with the borrower that enables information to be collected that enables the bank to make improvements to the management of the loan. Another example is the one described in Mester et al (2001), which involves monitoring of the borrower's current account activity and using this information to assess the credit quality of the loan. In particular, the system aims to detect events such as overdraft occurrences or shortfall of income, which may happen independently of whether the borrower maintains loan repayments.[3]

Another aspect glossed over in the example above, but which becomes more striking in a dynamical setting, is that banks have the option to sell loans in the secondary market as a means of manage their balance sheet risk. Several observers have commented on the link between monitoring and loan sales, with a prominent view that a reduction in monitoring activity encourages loan sales and lending activity, and therefore increases risk.[4] The Basel Committee's view that credit risk increases because of the "failure to identify and recognise deterioration in credit quality" is supported by this argument. That loan sales have become substitutes for monitoring is also argued by Parlour and Plantin (2008), and Amir Sufi argues in a recent FT article that banks are no longer "slavvy information gathering lenders, but

[3]Statistics show that for individual borrowers events such as illness, loss of job, and divorce/separation are common causes of subsequent default on a loan. For corporations mismanagement of operations and unexpected loss of main business lines are important.

[4]Cecchetti (2011) for instance makes the observation that "[i]n the years running up to the crisis, mortgage lenders reduced their costly screening of borrowers in order to increase their lending and profit. By reducing screening, they saved costs directly and they also expanded the range of potential borrowers so that they could boost the number of loans. With costs reduced, more loans mean higher profits if – as usually occurred – the mortgage lenders quickly sold the new loans to other intermediaries who would securitize them."

take instead leveraged bets on real estate," implying that banks are failing in their duty to monitor loans.[5]

However, monitoring of loans and loan sales cannot a priori be seen as mutually exclusive activities – by gaining more information about the loan quality before making the decision to sell the loan the bank may improve the profits from loan sales. Bord and Santos (2012) argue that the fractions of loans that were sold/syndicated/securitised among lead US banks from 1990s onwards (the banks that were the most active users of the originate-to-distribute model of banking) were increasing not only for new loans but also for loans that had already been held for some years, suggesting that new information about existing loans obtained through monitoring is a possible trigger for loan sales. The links between regulation, monitoring, and risk remains, therefore, unclear.

Meanwhile, the empirical observation suggests that small banks tend to be more willing to lend to small businesses than larger banks.[6] This has been attributed in part to regulation, as the way risk capital is calculated does not penalise small banks for risky lending in the same way as large banks, and also in part to the fact that small banks do not rely on credit scoring systems to the same extent as larger banks.[7] Regulation is therefore a potential factor explaining why banks switch away from dynamic monitoring systems to credit scoring systems. To add to this point, Dermine (2013) argues directly that loan monitoring is adversely affected by regulation.

The above observations lead us to ask what the benefits of banks' investment in loan monitoring systems are. More specifically, we address the following two questions in the current paper:

(a) Does dynamic loan monitoring reduce bank risk?

[5] FT 13 October 2014: Bernanke's failed mortgage application exposes the flaw in banking.

[6] See Craig and Hardee (2007), and Berger et al. (2005) who argue that large banks often shy away from "informationally challenging" borrowers.

[7] See Stacy Mitchell's article *Why Small Banks Make More Small Business Loans*, Institute for Local Self-Reliance, 10 February 2010.

(b) What are the net benefits of banks' investment in dynamic monitoring systems?

To answer these questions, we describe the bank's problem as an optimal stopping problem. In our model, the bank forms a prior belief about the loan quality (possibly by the use of credit scoring systems) and then keeps revising its belief by acquiring further information through dynamic monitoring until a stopping time is reached. Based on this model, we show that a more intense monitoring activity can increase the incentive to keep riskier loans, which will be met by greater regulatory compliance costs when the regulator seeks to limit the amount of risk the bank takes. Therefore, banks can destroy their monitoring systems and make greater savings in terms of regulatory compliance than they lose by the dismantling of their monitoring ability. This perverse incentive structure should lead to a rethink of the way we regulate banks.

The related literature spans the following three areas: loan monitoring and the use of information in loan management; loan sales; and the dynamic modelling of information release. In the first strand we find Berger et al (2005) and Loranth and Morrison (2009). These papers distinguish between monitoring systems that generate output in the form of "soft" information and "hard" information. While hard information can be transmitted through formal systems, soft information is subjective and is not easily understood out of context. In contrast, we distinguish between monitoring systems on the basis of fast and slow output. The information may be soft or hard in either case – the key difference is the speed of the output.

In the second strand Chemla and Hennessy (2011) and Bester et al. (2013) study the relationship between loan sales and loan screening in a static framework; Parlour and Winton (2009) who look at transfer of control rights to new investors who have some ability to monitor their new loans versus the retention of control rights through insuring the credit risk of the loan; and Chemla and Hennessy (2012) who consider a secondary market consisting of rational speculators who may or may not be informed and rational hedgers who trade for risk sharing reasons. Their focus is on the loan sale market, whereas our focus is on the use of monitoring systems that can lead to loan sales. In particular, they do not study the effects on risk as a result of changes in the monitoring systems.

In the third strand Daley and Green (2012) employ a similar dynamic modelling technology. Daley and Green (2012) consider an informed seller deciding the optimal timing of the sale of a good to an uninformed buyer who learns free of cost the value of the good over time. We consider sales when the seller can invest in costly learning about the value over time, but when the transaction takes place there is no information asymmetry between the seller and the buyer. In this sense the similarity to Daley and Green's (2012) model does not go beyond the modelling technology.

2 The Basic Model

We look at the problem of intermediating loans between borrowers and sellers. Adverse selection costs prevent direct trade, and the intermediation is provided by a bank with monitoring ability which enables better assessment of loan quality than the investors. The bank decides to hold some loans on its own book, and to sell some loans to the investors by signalling the bank's information about the loan, at a cost, to investors. This decision is the main risk management decision for the bank, as the risk of the bank is a function of the hold or sell decision.

2.1 The Structure of the Model

The bank first operates a credit scoring system that is activated at the time the loan originates. A loan that originates at time 0 promises to pay a constant perpetual coupon flow c and is classified by parameter π_0, which is the prior probability that the loan is riskless – a loan with $\pi_0 = 0$ will default with probability 1 and one with $\pi_0 = 1$ will never default.[8] The discount rate is r. A high quality loan is risk free and has value $\frac{c}{r}$. Note that for the bank's portfolio, the credit scoring system will produce an initial distribution of the prior beliefs π_0, and the more effective the scoring system is the more probability mass will accumulate near the extreme points of 0 and 1. The shape of this distribution is not relevant to our model, so we assume that the priors are uniformly distributed on $[0, 1]$.

The bank can sell loans to the investors at competitive prices, which depend on the ability of the investors to bear credit risk, represented by k. The bank is regulated and has a lower ability to bear credit risk, which is represented by $k + \kappa$, where κ is set by the regulator. When the bank sells a loan to the investors, it incurs a signalling cost g (> 0). If the probability the loan is of high quality is π, then the value of the loan held by the bank is $V_I(\pi) = \frac{c}{r} - (k + \kappa)(1 - \pi)$. The market value of the loan is $V_S(\pi) = \frac{c}{r} - k(1 - \pi)$ at which the bank receives net proceeds of $V_S(\pi) - g$ from selling the loan. We assume that $\kappa > g$; thus, some gains from trade can be made for the lowest quality loans. If this assumption were violated the bank would never sell loans. Thus, there is a trade-off – the bank seeks to balance the gains from trade against the cost of the loan sale transaction. For high risk loans the gains from trade will typically dominate the signalling cost, whereas for low risk loans the reverse will be true.

The bank can continue monitoring the loan using its dynamic monitoring system at a constant cost m, and this monitoring process results in a posterior probability process $\{\pi_t | t \geq 0, \pi_0\}$. The monitoring generates an information signal which is represented by x_t, x_0 given. This process represents an observable characteristic associated with the loan, which follows a Brownian motion

$$dx_t = \theta \mu dt + \sigma dB_t,$$

where $\mu, \sigma > 0$ and $\theta \in \{0, 1\}$. Although the drift rate $\theta\mu \in \{0, \mu\}$ is the primary determinant

[8] Since the date of origination is immaterial to this problem, we normalise by setting $t_0 = 0$ for all loans.

of loan quality, it is not directly observable. The loan has high credit quality and is risk free if $\theta = 1$, and it has low credit quality and is maximally risky if $\theta = 0$. While the drift rate is not observable, the path of $\{x_t\}_t$ is, and this allows the bank to make inferences about loan quality in real time. Monitoring leads, therefore, to a posterior process $\{\pi_t | t \geq 0, \pi_0\}$.

2.2 Bank's Optimal Stopping Rules

The solution to the monitoring problem is to make a final decision of whether to hold or sell the loan at an optimal stopping time t^*. Thus, we are investigating trigger strategies of the type where if $\pi_{t^*} = \pi^*$ for some π^* the bank decides to hold the loan at time t^*, and if $\pi_{t^*} = \pi^{**}$ for some $\pi^{**} \leq \pi^*$ the bank sells the loan at time t^*. The value of the loan that is held at π^* is $V_I(\pi^*) = \frac{c}{r} - (k + \kappa)(1 - \pi^*)$. The net value to the bank of the loan that is sold at π^{**} is $V_S(\pi^{**}) - g = \frac{c}{r} - k(1 - \pi^{**}) - g$. The value of a loan that is monitored $V_M(\pi_t)$ is expressed as follows: For $\pi_t \in [\pi^{**}, \pi^*]$,

$$V_M(\pi_t) = \mathbb{E}\left\{ \int_0^{t^*} e^{-ru}(c - m)du + e^{-rt^*} \max\left\{V_S(\pi_\tau) - g, V_I(\pi_\tau)\right\} \right\}$$

It is clear from the expression above that the value of a monitored loan satisfies $V_M(\pi_t) \geq \max\left\{V_S(\pi_t) - g, V_I(\pi_t)\right\}$ for $\pi_t \in [\pi^{**}, \pi^*]$, so by investing in internal monitoring the bank will always be better off than by sorting loans immediately by naively choosing to keep loans in-house if $V_I(\pi_0) \geq V_S(\pi_0) - g$ and sell loans otherwise.[9] The dynamic monitoring system is described by its cost m and its efficiency as measured by the signal-to-noise ratio $\frac{\mu}{\sigma}$ (or its inverse).

The posterior probability process is $\pi_t = \mathbb{P}(\theta = 1 | \mathcal{F}_t^x)$ (with respect to the σ-algebra generated by $\{x_\tau\}_{\tau=0}^t$). From Peskir and Shiryaev (2006) we find that the likelihood ratio

[9] By setting $t^* = 0$ we find a lower bound for V_M equal to $\max\{V_I, V_S - g\}$, and by choosing t^* freely we may improve on this valuation.

process φ_t, defined by the Radon-Nikodyn derivative, is given by

$$\varphi_t = \frac{d(\mathbb{P}_1|\mathcal{F}_t^x)}{d(\mathbb{P}_0|\mathcal{F}_t^x)},$$

where $\mathbb{P}_1$ is the probability measure for the process x_t when $\theta = 1$ and $\mathbb{P}_0$ is the probability measure for the process x_t when $\theta = 0$. The process admits the representation

$$\varphi_t = \exp\left(\frac{\mu}{\sigma^2}\left(x_t - \frac{\mu}{2}t\right)\right),$$

and we can use this representation to derive the posterior probability process

$$\pi_t = \left(\frac{\pi_0}{1-\pi_0}\varphi_t\right) \Big/ \left(1 + \frac{\pi_0}{1-\pi_0}\varphi_t\right),$$

where π_0 is the prior belief at the time the loan was originated. This process solves the stochastic differential equation

$$d\pi_t = \frac{\mu}{\sigma}\pi_t(1-\pi_t)d\tilde{B}_t, \quad \pi_0 = \pi, \quad \tilde{B}_t = \frac{1}{\sigma}\left(x_t - \mu\int_0^t \pi_s ds\right), \tag{1}$$

where $\tilde{B}_t$ is a standard Brownian motion.[10]

The problem we investigate can be taken to be a problem of optimal stopping in π_t rather than x_t because the process x_t is not known by the decision maker.[11] The time of origination does not play a role here since the system is Markovian and the history leading up to a posterior belief π_t does not matter. One of the key differences between our model and Daley and Green (2012) is that in their model the decision maker is informed about the drift term of the x_t process and is waiting for the optimal time when a transaction with a buyer can take place. In our framework the decision maker is uninformed and invests in

[10]The derivation of (1) is given in the appendix.

[11]The sample path leading up to time t from origination is observable, but the drift term in the process generating this sample path is not.

monitoring which leads to better information about the process. This makes the π_t process the natural state variable.

The infinitesimal generator of $\{\pi_t, t \geq 0\}$ is

$$\mathbb{L} = \frac{\mu^2}{2\sigma^2}\pi^2(1-\pi)^2\frac{\partial^2}{\partial\pi^2} - r\mathbb{I},$$

which applies to any value function V defined over the cash flow of the loan and the credit quality π_t of the loan, and where $\mathbb{I}$ represents the identity operator. Monitoring of the loan will make x_t observable and this leads to changes in π_t under the $\tilde{B}_t$ process. The change in loan value $V_M(\pi_t)$ net of the cost of capital $rV_M(\pi_t)$ is captured by the term $\mathbb{L}(V_M(\pi_t))$, and together with the net gain $c - m$ (which we can interpret as a dividend stream), the optimality condition is that the sum of the two equals zero. For loans that are not monitored there is no change in the loan value because the bank cannot infer new information about their loan quality, and the value of these loans is simply $V_I(\pi)$ for loans that are kept in-house and $V_S(\pi)$ for loans that are sold, when the belief about loan quality is π.

Definition 1: *Bank's optimal stopping rule is a pair* (π^*, π^{**}) *that satisfies the following conditions*:

(a) *The value function* V_M *satisfies* $\mathbb{L}(V_M(\pi_t))+(c-m) = 0$ *for* $\pi_t \in [\pi^{**}, \pi^*]$ *and* $V_M(\pi_t) \geq \max\{V_I(\pi_t), V_S(\pi_t) - g\}$;

(b) *The trigger point* π^* *satisfies* $V_M(\pi^*) = V_I(\pi^*)$ *and* $\frac{d}{d\pi}V_M(\pi^*) = \frac{d}{d\pi}V_I(\pi^*)$;

(c) *The trigger point* π^{**} *satisfies* $V_M(\pi^*) = V_S(\pi^*) - g$ *and* $\frac{d}{d\pi}V_M(\pi^*) = \frac{d}{d\pi}V_S(\pi^{**})$.

Condition (a) follows from standard continuous time techniques, and the only feature to look out for here is the condition that V_M must be greater at any interior point that both V_I and $V_S - g$. This condition impose additional constraints on the value functions, since it necessitates that V_M is strictly convex at the thresholds π^* and π^{**}. Conditions (b) and (c) are value matching and standard smooth pasting conditions.

Lemma 1: *The following function satisfies* $\mathbb{L}(V_M(\pi_t))+(c-m)=0$, *the first of the conditions in Definition 1 (a) above*:

$$V_M(\pi_t) = \bar{A}\pi_t^{\frac{1}{2}(1-\xi)}(1-\pi_t)^{\frac{1}{2}(1+\xi)} + \bar{B}\pi_t^{\frac{1}{2}(1+\xi)}(1-\pi_t)^{\frac{1}{2}(1-\xi)} + \frac{c-m}{r},$$

where $\xi = \sqrt{1+8r\,(\sigma/\mu)^2}$ *and* $\bar{A}$ *and* $\bar{B}$ *are arbitrary real constants.*

Lemma 1 is a general solution of the conditions that are imposed on the value function by condition (a) in Definition 1. The constants $\bar{A}$ and $\bar{B}$ depend on the boundary conditions that arise from the other conditions (b) and (c) in Definition 1, and they determine the option value of deferring the optimal risk management decision for the loan. We expect that the option value of delaying the optimal risk management is high when the optimal risk management decision is unclear, i.e. when the credit quality of the loan is such that the in-house value is close to the sale value net of the sale cost. Define $f_1(\pi)$ and $f_2(\pi)$ by

$$\begin{aligned} f_1(\pi) \quad &:= \pi^{\frac{1}{2}(1-\xi)}(1-\pi)^{\frac{1}{2}(1+\xi)}; \\ f_2(\pi) \quad &:= \pi^{\frac{1}{2}(1+\xi)}(1-\pi)^{\frac{1}{2}(1-\xi)}. \end{aligned}$$

Also define $g_1(\pi)$ and $g_2(\pi)$ by

$$\begin{aligned} g_1(\pi) \quad &:= \tfrac{1}{2}(1-\xi)\pi^{-1} - \tfrac{1}{2}(1+\xi)(1-\pi)^{-1}; \\ g_2(\pi) \quad &:= \tfrac{1}{2}(1+\xi)\pi^{-1} - \tfrac{1}{2}(1-\xi)(1-\pi)^{-1}. \end{aligned}$$

Note that $f_1(\pi)$, $f_2(\pi)$, $g_2(\pi) > 0$ and $g_1(\pi) < 0$, for all π, and $f'(\pi) = f_i(\pi)g_i(\pi)$, $i = 1, 2$.

Also, it is easy to verify that $f_i''(\pi) = f_i(\pi)\left[(g_i(\pi))^2 + g_i'(\pi)\right] > 0$, $i = 1, 2$—so, f_i is strictly convex. Moreover, define the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ by

$$\mathbf{M}_1 := \begin{pmatrix} f_1(\pi^*)g_1(\pi^*) & f_2(\pi^*)g_2(\pi^*) \\ f_1(\pi^{**})g_1(\pi^{**}) & f_2(\pi^{**})g_2(\pi^{**}) \end{pmatrix}, \qquad \mathbf{M}_2 := \begin{pmatrix} f_1(\pi^*) & f_2(\pi^*) \\ f_1(\pi^{**}) & f_2(\pi^{**}) \end{pmatrix},$$

of which we make use in the following result.

Proposition 1: *The points π^* and π^{**}, which are the solutions to the following system:*

$$\mathbf{M}_1^{-1} \begin{pmatrix} k+\kappa \\ k \end{pmatrix} = \begin{pmatrix} \bar{A} \\ \bar{B} \end{pmatrix} = \mathbf{M}_2^{-1} \begin{pmatrix} -(k+\kappa)(1-\pi^*) + \frac{m}{r} \\ -k(1-\pi^{**}) - g + \frac{m}{r} \end{pmatrix}, \tag{2}$$

are the optimal trigger points describing the dynamic monitoring region $[\pi^{**}, \pi^*]$, provided that $V_M(\pi) > \max\{V_I(\pi), V_S(\pi)\}$ for all $\pi \in (\pi^{**}, \pi^*)$. A necessary and sufficient condition for the inequality is that V_M is strictly convex over the region $[\pi^{**}, \pi^*]$.

The system of equations (2) is essentially standard value-matching and smooth-pasting conditions, where (a) the bank will stop the monitoring process and sell the loan when both the value and the marginal value of the loan are the same for either decision, and similarly (b) the bank will stop the monitoring process and hold the loan when the value of and the marginal value of the loan are the same for either decision. If the values or the marginal values were different, it would be profitable to either speed up or defer the decision to hold or sell the loan. The option value of deferring the decision further, therefore, is exactly zero at the optimal stopping points.

We now examine the effects of parameters m and κ on the thresholds π^* and π^{**}.

Proposition 2: *Thresholds π^* and π^{**} have the following properties.*

(a) π^* *is strictly decreasing and* π^{**} *is strictly increasing in* m.

(b) π^* *and* π^{**} *are both strictly increasing in* κ.

Proposition 2 (a) states that the range of monitoring $[\pi^{**}, \pi^*]$ is wider when the cost of monitoring m is lower. Meanwhile, Proposition 2 (b) implies that the bank would keep fewer low quality loans when the regulatory cost of risk capital κ is higher, since bank would only keep loans that is equal to or better than π^{**} on its own balance sheet, and sell out all other loans. Moreover, Proposition 2 (b) also states the bank would be keener to keep monitoring the loans when κ is higher. Before proceeding, we make the following comments related to the model and the optimality conditions.

Mandatory use of credit scoring systems: In an earlier version of the model we operated with a three-stage monitoring process: an initial screening of borrowers that lead to the formation of the priors π_0; then subsequently the use of credit scoring systems and/or dynamic monitoring systems. The optimality conditions would ensure the use of credit scoring if the cost of the use of credit scoring is sufficiently low. Therefore, the problem becomes essentially a choice of deciding when to stop the dynamic monitoring process. Therefore, in the formulation above we have simplified the three-stage process to a two-stage process: the initial screening process includes the use of credit scoring systems at zero cost. In this formulation, the investment in a more effective credit scoring system simply influences the distribution of prior beliefs, putting more probability mass in the regions close to zero or one. This is of course profitable as it makes it easier for the bank to separate the loans that can be sold or held immediately from those that require dynamic monitoring, but crucially the dynamic monitoring problem is still the same as it depends only on the prior belief of the loan. This simplification is, therefore, taking very little away from the generality of the model.

Interpretation of loan sales: The optimal risk management of bad loans in our model is to sell them. Alternatively, we can interpret loan sales as an intervention policy by the bank which leads to reduced losses in credit events. If we consider the value of a loan of quality π kept in-house without any intervention to mitigate losses to be $\frac{c}{r} - (k + \kappa)(1 - \pi)$, and the value of a similar loan with intervention to be $\frac{c}{r} - k(1 - \pi) - g$, the net value of intervention is $\kappa(1 - \pi) - g$, where $\kappa(1 - \pi)$ is the value of intervention, and g is the cost. The bank invests in information, therefore, to separate the loans that need intervention from those that do not. This interpretation may be more natural for small banks that find it difficult to invest in signalling technology that enables it to sell loans at a fair value. The model will not distinguish loan sales from this kind of intervention.

Extreme Stopping Rule Thresholds: It is possible that the equilibrium values of the optimal stopping problem get close to zero or one, i.e. $\pi^{**} \to 0$ or $\pi^{*} \to 1$. These cases are problematic in the sense that the state variable π_t is "arrested" at the boundary points 0 and 1, i.e. the diffusion term $\frac{\mu}{\sigma}\pi_t(1 - \pi_t) \to 0$ for $\pi_t \to 0$ or $\pi_t \to 1$. When the state variable is near 0, therefore, the choice is between $V_I(0) = \frac{c}{r} - (k + \kappa)$ and $V_S(0) = \frac{c}{r} - k - g$, and under sensible parameter values we will always choose $V_S(0)$. When the state variable is near 1, the choice is between $V_I(1) = \frac{c}{r}$ and $V_S(1) = \frac{c}{r} - g$, and we will always choose $V_I(1)$. There can be no economic value associated with a continuation of monitoring. If we try to calculate π^{*} or π^{**} numerically, however, we will easily get into difficulties because extreme values of π^{*} or π^{**} make the matrices $\mathbf{M}_1$ or $\mathbf{M}_2$ nearly singular, so the optimality condition is badly behaved.

2.3 Steady State

Since we are building a dynamic model we shall assume the bank is operating at steady state which necessitates, when loans are of infinite duration, that there is some exogenous entry of new loans and exit of existing loans that are repaid early.[12] We assume that a loan of any quality $\pi_0 \in [0,1]$ originates randomly at some time s, and conditional on origination will be repaid at a fair price (and disappear) randomly, with an exponential distribution, at some time $t \geq s$. The advantage of the exponential distribution is that it is a memoryless distribution in the sense that the probability that the loan will be repaid over the next instant, conditional on not being repaid up to that time, is independent of the passage of time since origination. Therefore, for time $t > s$ and $\rho_1 > 0$, we assume the probability that a new loan (of any kind) arrives is

$$\mathbb{P}(\text{New type-}\pi \text{ loan arrives between } s \text{ and } s+dt) = \int_s^{s+dt} \rho_1 e^{-\rho_1(t-s)} dt.$$

Conditional on the likelihood $\mathbb{P}_s^\pi$ of a loan of type π exists at time s, and that n such loans exist, the probability that a loan is withdrawn at time s is

$$\mathbb{P}(\text{Existing type-}\pi \text{ loan withdrawn between } s \text{ and } s+dt|\mathbb{P}_s^\pi, n) = \int_s^{s+dt} n\mathbb{P}_s^\pi \rho_0 e^{-\rho_0(t-s)} dt.$$

Therefore, the net new arrival of loans of type π is

$$nd\mathbb{P}_s^\pi = \int_s^{s+dt} \left(\rho_1 e^{-\rho_1(t-s)} - n\mathbb{P}_s^\pi \rho_0 e^{-\rho_0(t-s)}\right) dt.$$

The system is in steady state when $nd\mathbb{P}_s^\pi = 0$, and in this case the bank receives an inflow of capital from the withdrawn loans that exactly matches the outflow of capital into new loans. Working out the integrals, we find that

$$nd\mathbb{P}_s^\pi = (\rho_1 - n\mathbb{P}_s^\pi \rho_0)\, dt = 0,$$

[12] This is done for tractability here but this assumption is not without realism even when loans are not of infinite maturity.

which implies the steady state condition $\mathbb{P}_s^\pi = \frac{\rho_1}{n\rho_0}$ which links the likelihood of any loan existing to the arrival and withdrawal rates and the number of loans existing. If the arrival and withdrawal rates are the same, therefore, the steady state condition implies that the probability that a loan of a given type exists is exactly the inverse of the loans in existence. We will assume a steady state with $n\mathbb{P}_t^\pi = \frac{\rho_1}{\rho_0}$ for the remaining parts of this paper.

Steady State in the Dynamic Monitoring Region: The steady state condition is fairly obvious for the regions $[0, \pi^{**}]$ and $[\pi^*, 1]$ since these loans are never monitored and there is no change in π_t over time. However, for loans in the interior of the monitoring region the state variable π_t will change over time and the steady state condition should recognise this. The number of new loans of type $\bar{\pi}$ at some point in time t equals the new loans arriving plus "old" loans arriving at $\bar{\pi}$ from above, i.e. $\pi_{t^-} > \pi_t = \bar{\pi}$, and plus "old" loans arriving at π_t from below i.e. $\pi_{t^-} < \pi_t = \bar{\pi}$. The number of withdrawn loans is similarly the withdrawn loans plus the loans of type $\bar{\pi}$ exiting to a type above, i.e. $\pi_{t^-} = \bar{\pi} < \pi_t$, and plus the loans of type π_t exiting to a type below, i.e. $\pi_{t^-} = \bar{\pi} > \pi_t$. With Brownian uncertainty this problem can be ignored, and we provide some technical details in the appendix to that effect.

3 Regulation and Intermediation Surplus

In this section we extend the model by endogenising κ, the additional cost of risk capital imposed by the regulator on the bank, and analyse the relationship between bank's risk taking behaviour, bank's intermediation surplus, and regulation.

3.1 Regulation

The bank accumulates risk through its lending operations which is the primary reason the regulator is concerned with controlling the lending operations of the bank. This risk is associated with the loans the bank is currently keeping—either as loans that are placed in-house permanently or loans that are currently being monitored and awaiting a risk management decision of whether to be sold or placed in-house. If the bank holds a lot of risky loans it may pose a systemic threat to the financial system, which the regulator controls through imposing costs on the bank for holding risk, i.e. κ in our model.

The natural risk measure for a loan of quality π is the discount of a risky loan relative to a risk free one. Specifically, the risk free value of the loan is $\frac{c}{r}$, and the risky value of the loan is $V_S(\pi) = \frac{c}{r} - k(1-\pi)$, so the risk associated with the loan is simply the discount $\frac{c}{r} - V_S(\pi) = k(1-\pi)$. For a loan portfolio with loans of quality π uniformly distributed over $[\underline{\pi}, \overline{\pi}]$, the loan portfolio risk is $\int_{\underline{\pi}}^{\overline{\pi}} k(1-\pi)d\pi = k(\overline{\pi} - \underline{\pi})\left(1 - \frac{\overline{\pi}+\underline{\pi}}{2}\right)$. Since the bank holds loans of type π varying uniformly from π^{**} to 1 by assumption, the risk of the bank's loan portfolio is $\frac{k}{2}(1-\pi^{**})^2$.

Note that the risk measure $\frac{k}{2}(1-\pi^{**})^2$ is a market based measure of the need to hold risk capital, since it depends on the discount of the market value of loans relative to a corresponding risk free loan, i.e. the risk is measured relative to the market's risk bearing ability k. If we made the risk measure based on the internal risk bearing ability $k+\kappa$ it would depend on the regulator's actions and would mean that a more strict regulation of a bank would itself lead to a higher risk in the bank's loan portfolio without any action on the part of the bank. To prevent this from happening we make the risk measure a market based risk measure, broadly in line with risk assessment models used in practice. An implication of the risk measure is that the regulator requires higher cost of bearing risk, for instance more strict capital requirements, to the banks in which the barrier for holding risky loans is low (i.e. π^{**} is low).

With the above observation, we assume that the regulator sets κ^* by the following rule:

Assumption 1: *The regulator knows the bank's optimal stopping rules defined by Definition 1, and sets κ^* so that $\psi(\kappa^*) = \bar{\psi}$ holds for $\bar{\psi}$ given, where $\psi(\kappa) := \frac{k}{2}[1-\pi^{**}(\kappa)]^2$.*

The expression $\pi^{**}(\kappa^*)$ reflects the fact that the bank chooses π^{**} in accord with the optimal stopping rules defined by Definition 1, given κ^*. Namely π^{**} is really a function of κ^*, and Assumption 1 assumes that the regulator possesses full knowledge of the bank's optimal stopping rules. Under Assumption 1, κ^* is set so that the risk level is matched with $\bar{\psi}$, which has a one-to-one relationship with π^{**} given k. Thus, setting $\bar{\psi}$ is equivalent to setting π^{**}, and we can interpret Assumption 1 that the regulator imposes the regulatory cost of risk capital κ^* on the bank so that the bank will set its lower threshold π^{**} at a level targeted by the regulator. In other words, π^{**} is fixed under Assumption 1.

In what follows, we define the intermediation surplus for the bank, which roughly measures the value-added of the bank through its monitoring capability and its keeping of the loans in its own book.

3.2 Intermediation Surplus

A loan of quality π has a value of $V_S(\pi) - g$ for the bank, net of signalling costs g. The bank can improve on the market value in two ways. The bank can either hold the loan in-house at the value $V_I(\pi)$ or keep the loan under observation in its monitoring system with the aim of delaying the decision to hold or to sell, at the value $V_M(\pi)$. The former way saves the signalling cost g, but incurs a higher regulatory cost of risk capital as determined by the regulator's choice of κ. The intermediation surplus is, therefore, simply the maximum of these three values minus the market value, i.e.

$$
\begin{aligned}
V(\pi) &= \max\{V_S(\pi) - g, V_M(\pi), V_I(\pi)\} - [V_S(\pi) - g] \\
&= \max\{0, V_M(\pi) - [V_S(\pi) - g], V_I(\pi) - (V_S(\pi) - g)\}.
\end{aligned}
$$

In the steady state the bank will have a continuous inflow of new loans of quality π_0, and we assume that the inflow is uniformly distributed on $[0, 1]$. Observe however that the functional form of $V_I(\pi)$ depends on κ, while that of $V_M(\pi)$ depends on κ, m and σ/μ. Hence, the functional form of $V(\pi)$ depends on κ, m and σ/μ, too. It follows that the expected value of intermediation of a loan can be written as a function of κ, m and σ/μ,

$$\mathbb{E}V(\pi_0|\kappa, m, \sigma/\mu) = \int_0^1 V(\pi_0|\kappa, m, \sigma/\mu)d\pi_0.$$

Since the regulator's choice of κ has a direct influence on the loans that are kept in-house, and an indirect influence on the loans that are being monitored, the strength of regulation is the prime determinant of intermediation surplus. However, under Assumption 1, to restrict the bank's risk level, the regulator would set κ corresponding to the values of m and σ/μ given. Hence, we define the following value function that defines the intermediation surplus for the bank, which is described as a function of m and σ/μ:

$$W(m, \sigma/\mu) := \mathbb{E}V(\pi_0|\kappa^*(m, \sigma/\mu), m, \sigma/\mu). \tag{3}$$

The expression $\kappa^*(m, \sigma/\mu)$ in (3) means that different banks would not necessarily be imposed the same regulatory cost of risk capital κ^*. The reason is that the bank's risk taking also depends on its internal monitoring activity, which in turn is determined by the parameters m and σ/μ.

3.3 Bank's Incentives and Regulation

We explore what effects a reduction in the cost of monitoring, m, and an increase in the effectiveness of the learning process, i.e. the inverse of σ/μ, have on regulation and intermediation surplus. A reduction in m is expressing cost savings in the information gathering process—for instance, through some form of automation of the monitoring process. A decrease in σ/μ means a higher quality of the signal/information acquired through dynamic monitoring—for instance, the bank invests in systems that can increase the number of signals gathered, or by combining multiple signals in a better way, so that the learning process is speeded up. However, before proceeding, we present the following result, which is useful in understanding the main thrust of the main results reported in Propositions 3 and 4 below.

Lemma 2: *Assume Assumption 1. For two distinct pairs of parameter values* $(m_0, \sigma_0/\mu_0)$ *and* $(m_1, \sigma_1/\mu_1)$*, suppose*

$$\frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0) > \frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_1^*, m_1, \sigma_1/\mu_1), \forall \pi \in [\pi^{**}, \min\{\pi_0^*, \pi_1^*\}], \tag{4}$$

where π_i^* *and* κ_i^* *are corresponding to* $(m_i, \sigma_i/\mu_i)$*,* $i = 0, 1$*. Then,*

(a) $V(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0) \geq V(\pi|\kappa_1^*, m_1, \sigma_1/\mu_1)$ *for all* π*, with* $>$ *for* $\pi > \pi^{**}$*.*

(b) $\pi_0^* < \pi_1^*$.

(c) $\kappa_0^* < \kappa_1^*$.

Thus, the extent of convexity of V_M is closely related to its location – in particular, the monitoring region and the regulatory cost of risk capital κ^*. It follows that once we can establish that the V_M function for a set of parameter values of $(m, \sigma/\mu)$ is 'more convex' than the V_M function for another set of parameter values of $(m, \sigma/\mu)$, we can conclude that the former V_M function is located above the latter V_M function. Also, we would know that the upper threshold π^* is lower for the former set of parameter values than for the latter, and the same applies to the regulatory cost of risk capital κ^*. By using Lemma 2, we claim the following result.

Proposition 3: *Under Assumption 1, monitoring cost* m *has the following effects on the bank's surplus* $W(m, \sigma/\mu)$*, the upper threshold* π^* *and the regulatory cost of risk capital* κ^*

imposed by the regulator.

(a) $W(m, \sigma/\mu)$ *is strictly increasing in* m*.*

(b) π^* *is strictly decreasing in* m*.*

(c) κ^* *is strictly decreasing in* m*.*

Proposition 3 (a) states that the bank's intermediation surplus will be reduced when the monitoring cost m is lowered. Also, Proposition 3 (b) implies that the bank stops monitoring earlier when the monitoring cost is higher. Moreover, by Proposition 3 (c), we now know that the regulator will impose a higher regulatory cost of risk capital κ^* when the monitoring cost m is lower. This means that the bank will be penalised by the regulator for making investments to lower the monitoring costs through a higher regulatory cost of risk capital. The intuition behind these results is that when the bank is equipped with a more cost efficient monitoring process/system, it is capable of monitoring loans of a wider range of quality, and the regulator will impose a higher κ^* in order to prevent the bank from holding loans with a quality lower than π^{**}. A higher regulatory cost of risk capital is a heavy burden to the bank, and is detrimental to its intermediation surplus.

Next, we present the following result concerning the effects of signal quality σ/μ.

Proposition 4: *Under Assumption 1, signal quality* σ/μ *has the following effects on the bank's surplus* $W(m, \sigma/\mu)$*, the upper threshold* π^* *and the regulatory cost of risk capital* κ^* *imposed by the regulator.*

(a) $W(m, \sigma/\mu)$ *is strictly increasing in* σ/μ*.*

(b) π^* *is strictly decreasing in* σ/μ*.*

(c) κ^* *is strictly decreasing in* σ/μ*.*

Proposition 4 shares the same spirit with Proposition 3. Namely, a bank with a better monitoring process/system – in this case, a lower σ/μ, will be imposed a higher regulatory cost of risk capital κ^* by the regulator, and its intermediation surplus will be lower. Also, the bank will be monitoring loans of a wider range of quality when it can collect better quality information through monitoring.

The main implication of Propositions 3 and 4 is that banks would have no incentives to invest in its monitoring process/system to improve the signal quality the system can gather, when investment made by a bank can affect m and σ/μ it will face. Moreover, the predictions of Propositions 3 and 4 are consistent with the observation that larger banks, who are imposed a higher capital requirement or credit costs, tend to offload more loans from its balance sheet and rely more on credit scoring systems, but less on dynamic monitoring, compared to smaller banks.

3.4 Implications for Regulation

The implications for financial regulation from the above results are in some sense straightforward. The standard approach to micro-prudential bank regulation is to allocate weights to the bank's various risk categories and to arrive at a risk-weighted aggregate risk capital requirement. In principle, it does not matter to the capital requirement whether the risk arises from lending activities or non-lending activities. However, our results suggests the regulator should apply a pecking order approach to regulation. First the regulator should control non-lending activities of the bank with the aim of satisfying Assumption 1. If this can be done without increasing the regulatory burden on the bank's lending activities the regulator is able to achieve it regulatory objectives at the same time as the bank retains the competitive advantage in lending that arises from the investment in dynamic monitoring systems. This leads to a reallocation of risk capital away from non-lending activities and into lending activities.

Parameter	Value
Coupon flow c	1
Discount rate r	0.05
Diffusion parameter σ	0.30
Market credit cost k	4
Signalling cost g	2

Table 1: The common parameter values used in the numerical examples.

What should happen when Assumption 1 can no longer be satisfied by controlling non-lending activities is unclear. Our model suggests there is in this case a genuine regulatory trade off between controlling the risk associated with lending and preventing banks from investing in competitive advantage arising from superior monitoring technology. This issue requires rigorous treatment in a dynamic equilibrium model which goes beyond the scope of this paper.

4 Numerical Examples

In this section, we provide numerical examples to illustrate the analytic results. The parameter values in Table 1 are used throughout.

First, we look at numerical solutions to the optimal monitoring region. The first is how the regulatory cost of risk capital (given by the parameter κ) affects the dynamic monitoring behaviour – we do not employ Assumption 1 here, and thus, κ is treated as a parameter. Figure 2 shows the results and uses, in addition to the parameter values above, drift parameter $\mu = 0.10$; monitoring cost $m = 0.30$; and the regulatory cost of risk capital κ varying from 2.5 to 18.5. The monitoring region $[\pi^{**}, \pi^{*}]$ is represented as the grey area in Figure 2.

We find that the threshold values for making the decision to sell the loan, and for making the decision to hold the loan, are both increasing in the regulatory cost of risk capital κ,

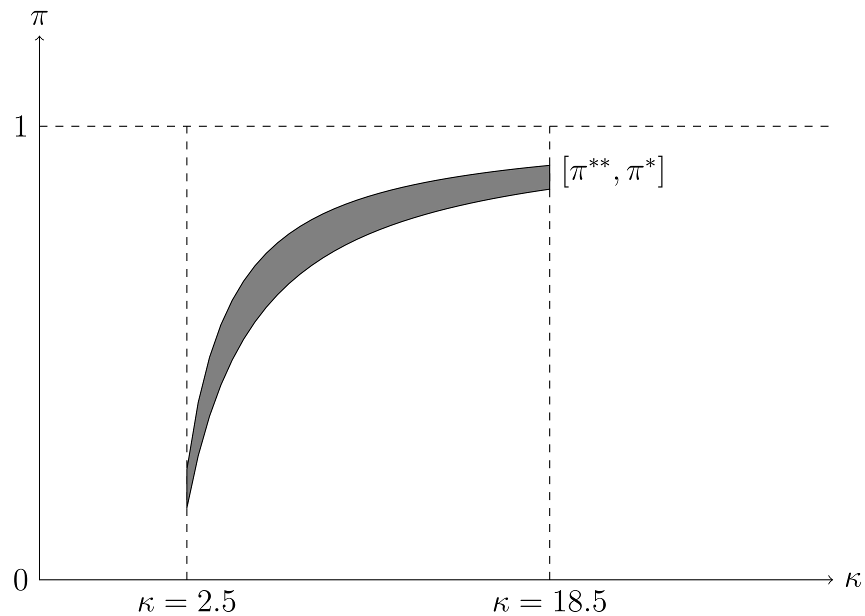

Figure 2: Illustration of how the bank's optimal stopping rules change in response to changes in the regulatory cost of risk capital κ.

confirming Proposition 2 (b). Therefore, the bank holds fewer loans, and sells more loans, the greater the cost of risk capital κ. The regulator has, therefore, the ability to control the bank's behaviour indirectly by influencing the risk management and loan monitoring decisions.

Next, we investigate the relationship between the monitoring cost and bank risk. We apply Assumption 1, and look at how monitoring activity and intermediated loan values vary as we vary the cost of monitoring m. We utilise largely the parameter values in Figure 2. If we take $c = 1$, $k = 4$, $r = 0.05$, and $\bar{\psi} = 0.18$, the riskiest perpetual loan trades at 56bp higher than its risk free counterpart, a 20-year annuity loan will trade at 216bp higher, and a 10-year annuity loan will trade at nearly 670bp higher.[13] Figure 3 shows

[13]These calculations are based on an average discount $\bar{\psi} = 0.18$ which translates into a lower barrier $\pi^{**} = 0.7$. We then solve the yield to maturity i of the equation measuring the discount at the point π^{**}:

$$\psi = \frac{1}{r}\left(1 - (1+r)^{-T}\right) - \frac{1}{i}\left(1 - (1+i)^{-T}\right)$$

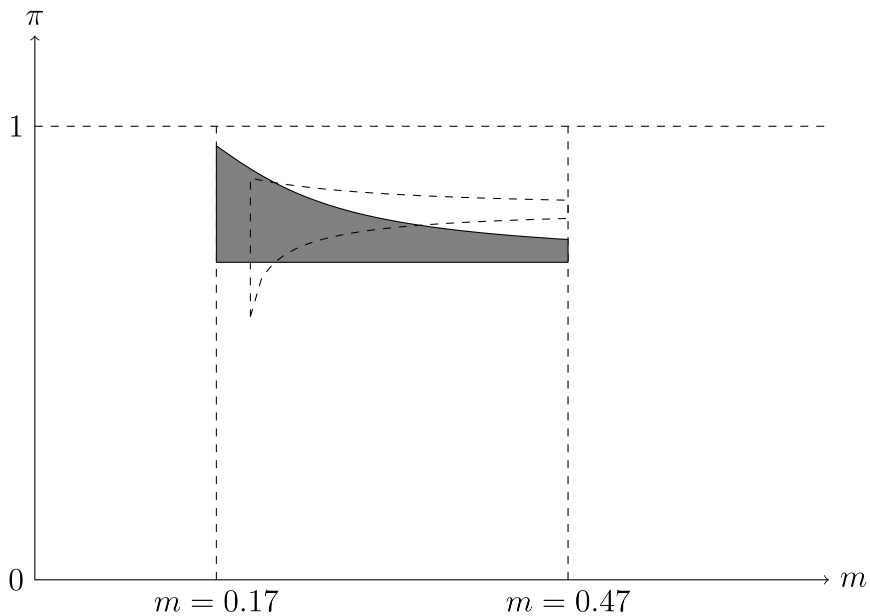

Figure 3: Monitoring region with risk constraint $\bar{\psi} = 0.18$ (grey) and monitoring region with κ fixed at 11 (dashed lines) for varying monitoring cost m.

the impact of the restriction that $\bar{\psi} = 0.18$ on monitoring activity for drift parameter μ fixed at 0.10; and monitoring cost m varying from 0.17 to 0.47. The region with dashed boundary is the corresponding region when κ is fixed at 11, derived for comparison. Under Assumption 1, regulation will prevent the bank from considering loans to hold or monitor below a certain threshold, and this limits the risk taking behaviour of the bank but will not limit the monitoring behaviour for loans of higher quality. In order to discipline the bank the regulator must impose higher capital requirements through the regulatory cost parameter κ when the monitoring cost is low, and this parameter can be reduced gradually as the monitoring cost increases, confirming the predictions of Proposition 3. Also, the dashed boundary with κ fixed at 11 endorses Proposition 2 (a) – the lower threshold π^{**} is strictly

When the time to maturity T is large, the discount in the loan value is very responsive to changes in the yield, and the risk premium of the average loan is small. This suggests that in a more realistic model the risk measure ψ should be calibrated to the maturity of the loan.

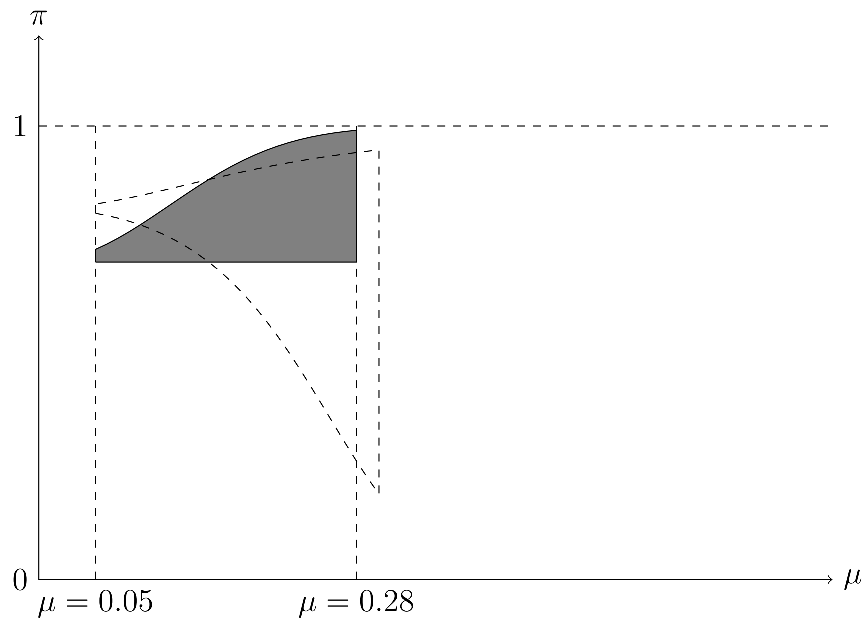

Figure 4: Monitoring region with risk constraint $\bar{\psi} = 0.18$ (grey) and monitoring region with κ fixed at 11 (dashed lines) for varying levels of monitoring efficiency μ.

increasing and the upper threshold π^* is strictly decreasing in m.

Next we investigate the relationship between risk and the effectiveness of monitoring as measured by μ.[14] The shaded area in Figure 4 shows the results under Assumption 1 with $\bar{\psi} = 0.18$, for monitoring cost m fixed at 0.30, and drift parameter μ varying from 0.05 to 0.28. The faintly shaded region is the monitoring region derived when the parameter κ is fixed at 11, for comparison. The results in Figure 4 pretty much mirror those in Figure 3 – the bank will be constrained to limit its monitoring to loans above a certain threshold. Regulation will therefore have the effect of reducing the risk in the bank's loan portfolio by discouraging the monitoring of loans of quality below this threshold. These results are consistent with Proposition 4.

Next we address the following question. How much should the bank invest in improving

[14]Since we keep the diffusion parameter σ constant, the variation in μ leads to variation in the signal-to-noise ratio $\frac{\mu}{\sigma}$ which is the essential parameter influencing monitoring behaviour.

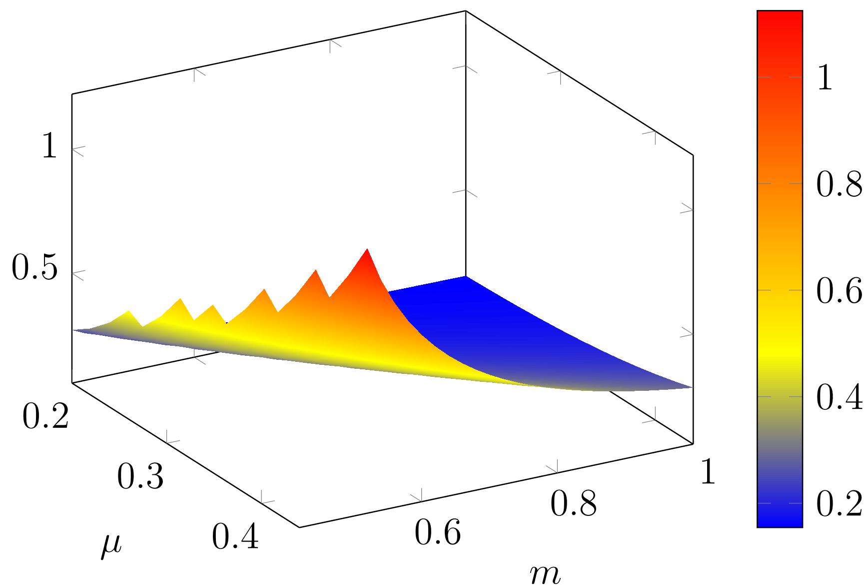

Figure 5: The figure shows the relationship between the monitoring effectiveness μ (on the x-axis), the monitoring cost m (on the y-axis), and the aggregate loan value (on the z-axis), when the bank faces fixed cost of risk capital $\kappa = 11$.

its dynamic monitoring systems? In our model, a natural objective function for the bank's optimisation problem with regard to this would be the intermediation surplus minus the cost of investment, where both depend on the choice of monitoring technology, which is parameterised by the pair (m, μ) with σ fixed. A possible constraint on this problem is the response from the regulator who can impose changes to the regulatory cost parameter κ. However, rather than considering a full optimal problem of investment in monitoring technology, we shall here consider the following two programmes – one where the bank can maximise the intermediation surplus by choosing freely over the space (m, μ) with constant κ, and one where the bank maximises over the same space with the constraint that κ is implied by $\psi(\kappa) = \bar{\psi}$.

Figure 5 shows the results for loss parameter κ fixed at 11; the drift parameter μ varying between 0.10 and 0.28; and the monitoring cost m varying from 0.19 to 0.50. The figure shows the bank's intermediation surplus $W(m, \sigma/\mu)$ for each pair (m, μ). We can see from the figure that the intermediation surplus is increasing in the μ-dimension, reflecting the fact that the bank's loan portfolio becomes more valuable the more effective the monitoring technology gets. The intermediation surplus is decreasing in the m-dimension, reflecting the fact that the bank's loan portfolio becomes more valuable the cheaper the monitoring technology gets. In either case, an investment into more effective or cheaper monitoring technology yields a positive payoff, and we expect that there is an interior solution for the optimal monitoring technology. Thus, depending on the cost of investment that attains each pair of (m, μ), there may be a trade-off between the increased intermediation surplus and the increased cost of investment, assuring the existence of an optimal investment level.

Next, we consider the restriction imposed by the regulator that prevents the bank from acquiring more risk when it gets access to cheaper or more effective monitoring. In this case, the regulation κ is endogenous, satisfying the restriction $\psi(\kappa) = \bar{\psi}$ which we made use of in the preceding subsection. Figure 6 shows the results for regulatory cost of risk capital κ given endogenously by setting $\bar{\psi} = 0.18$; the drift parameter μ varying from 0.10 to 0.28; and the monitoring cost m varying from 0.19 to 0.50. The figure shows the the bank's intermediation surplus $W(m, \sigma/\mu)$ for each pair (m, μ).

The effects on the intermediation surplus measured in both Figures 5 and 6 are exaggerated in the sense that only the top section of the vertical axes are shown. These effects are, however, comparable across the two figures and we can see that there are greater value variation in Figure 6 than in Figure 5. More importantly, the effects go in exactly the opposite directions. The intermediation surplus in Figure 6 is increasing in the m-dimension, and decreasing in the μ-dimension. This means that the bank loses by improving its dynamic intelligence systems even if the improvement comes at zero cost. The bank will in fact have an incentive to pay to make its dynamic intelligence systems more expensive or less effective as it will commit the bank to a risk acquisition strategy that will make the bank's regulatory burden lighter.

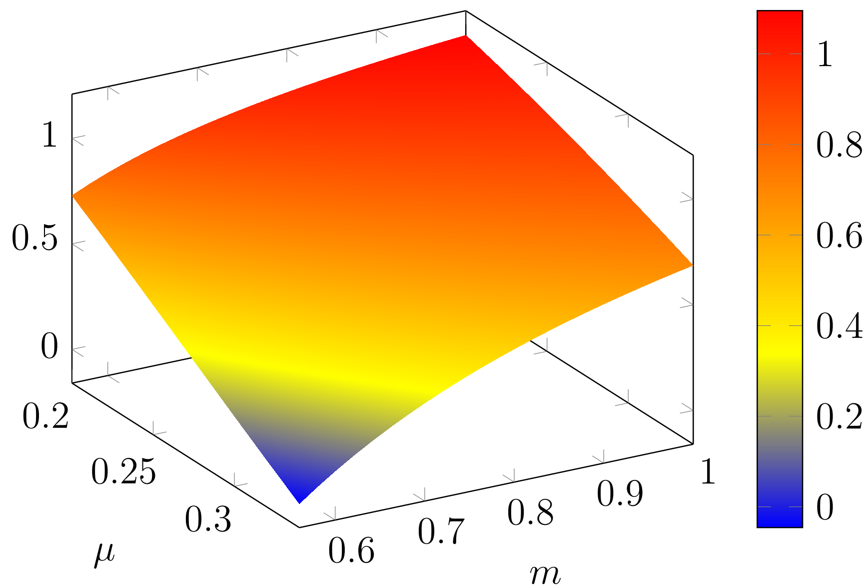

Figure 6: The figure shows the relationship between the monitoring effectiveness μ (on the x-axis), the monitoring cost m (on the y-axis), and the aggregate loan value (on the z-axis), when the bank must adhere to risk controls given by $\psi(\kappa) = \bar{\psi}$.

5 Conclusion

In this paper we have studied the relationship between loan monitoring and bank risk. We find the counter-intuitive result that banks may fail to invest in dynamic monitoring systems – systems that aim to "identify and recognise deterioration in loan quality" in the words of the Basel Committee on Banking Supervision – because these investments may in fact increase bank risk for which the bank is penalised through regulation. The reason is that the learning process through dynamic monitoring systems is slow and will increase the number of risky loans under observation, leading to extra net risk. Even if new and better monitoring technology is available at zero cost, the bank will be worse off. This raises the obvious question whether there is a social trade-off between regulating our banking system such that it can carry out its loan monitoring role efficiently or such that it does not become systemically risky.

Banks have increased incentives to hold risky loans when they can monitor the loans dynamically in real time. What economic implications can we draw from this effect? From a welfare point of view, this risk is not entirely harmful as it improves loan risk management of risky loans – and therefore facilitate lending to borrowers who are typically classified as risky. The question is whether these borrowers are treated in the same way by credit scoring systems and dynamic monitoring systems. In this sense, the credit scoring systems and the dynamic monitoring systems may not be perfect substitutes. The credit risk of some borrowers can be hard to assess with credit scoring systems, the kind of borrowers referred to in Berger et al (2005) as "informationally challenging" borrowers. Therefore, bank regulation that focuses on controlling bank risk only is likely to discriminate against dynamic monitoring systems, and therefore also against the informationally challenging borrowers. The banks will adapt to risk controls imposed by the regulator by underinvesting in dynamic monitoring technology, which in turn makes it harder for the informationally challenging borrowers to obtain bank loans.

Therefore, it is an open question whether the socially optimal lending to these borrowers is likely to be achieved under a regulatory framework that produces a bias towards credit scoring. It may be socially optimal to allow higher bank risk for institutions that target the segment of the loan market where the dynamic monitoring systems are the most effective. The evidence suggests this may be the case. Smaller banks, for instance, are more likely than larger banks to carry regulatory slack, i.e. that they can afford to take on extra bank risk without violating capital requirements than the bigger banks, and they should therefore afford to make investments in the dynamic monitoring systems that the bigger banks do not make. Smaller banks are also typically subject to looser bank regulation because they are less systemic, and therefore more likely to carry regulatory slack. These effects may explain why the smaller banks are often leading lending to small or young businesses which typically are seen as an informationally challenging sector of the loan market.

References

[1] Beltratti, A. and R.M. Stultz, 2009, “Why Did Some Banks Perform Better during the Credit Crisis? A Cross-Country Study of the Impact of Governance and Regulation”, http://ssrn.com/abstract=1433502.

[2] Berger, A.N., N.H. Miller, A. Petersen, R.G. Rajan, and J.C. Stein, 2005, Does Function Follow Organizational Form? Evidence From the Lending Practices of Large and Small Banks, *Journal of Financial Economics* **76**, 237 - 29.

[3] Berndt, A. and A. Gupta, 2009, “Moral Hazard and Adverse Selection in the Originate-to-Distribute Model of Bank Credit,” *Journal of Monetary Economics* **56**, 725 - 743.

[4] Bester, H., T. Gehrig, and R. Stenbacka, 2013, Loan Sales and Screening Incentives, mimeo, Free University of Berlin, Universität Wien, and Hanken School of Economics.

[5] Bord, V.M. and J.A.C. Santos, 2012, “The Rise of the Originate-to-Distribute Model and the Role of Banks in Financial Intermediation,” *RBNY Economic Policy Review*, July 2012.

[6] Borodin, A.N. and P. Salminen, 2002, *Handbook of Brownian Motion – Facts and Formulae*, Birkhäuser Verlag, Basel, Boston, Berlin.

[7] Cecchetti, S., 2011, *Money, Banking and Financial Markets,* Global ed., McGraw-Hill, Europe, Middle East & Africa.

[8] Chemla, G. and C.A. Hennessy, 2011, Originate-to-Distribute and Screening Incentives, mimeo, Imperical College Business School and London Business School.

[9] Daley, B. and B. Green, 2012, “Waiting for News in the Market for Lemons,” *Econometrica*, **80**, 1433-1504.

[10] Dermine, J., 2013, “Bank Regulations after the Global Financial Crisis: Good Intentions and Unintended Evil,” *European Financial Management* **19**, 658 - 674.

[11] Gorton, G.B. and G.G. Pennacchi, 1995, “Banks and Loan Sales: Marketing Nonmarketable Assets,” *Journal of Monetary Economics* **35**, 389 - 411.

[12] Keys, B.J., T. Mukherjee, A. Seru, and V. Vig, 2009, “Financial Regulation and Securitization: Evidence from Subprime Loans,” *Journal of Monetary Economics* **56**, 700 - 720.

[13] Loranth, G. and A. Morrison, 2009, “Internal Reporting Systems, Compensation Contracts, and Bank Regulation,” CEPR Discussion Paper No. 7179.

[14] Malekan, S. and G. Dionne, 2012, “Securitization and Optimal Retention under Moral Hazard,” http://ssrn.com/abstract=2038831.

[15] Mester, L.J., L.I. Nakamura, and M. Renault, 2001. “Checking accounts and bank monitoring,” Working Paper 01-3, Federal Reserve Bank of Philadelphia.

[16] Parulour, C.A. and G. Plantin, 2008, “Loan Sales and Relationship Banking,” *Journal of Finance* **63**, 1291 - 1314.

[17] Parlour, C.A. and A. Winton, 2009, “Laying Off Credit Risk: Loan Sales versus Credit Default Swap,” http://ssrn.com/abstract=1262885.

[18] Peskir, G. and A. Shiryaev, 2006, *Optimal Stopping and Free-Boundary Problems*, Birkhäuser, Basel - Boston - Berlin.

[19] Purnanandam, A., 2010, “Originate-to-Distribute Model and the Subprime Mortgage Crisis,” http://ssrn.com/abstract=1167786.

[20] Sufi, A., 2007, “Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans,” *Journal of Finance* **62**, 629 - 668.

A Technical Details

A.1 Derivation of the Posterior Process (1):

With Brownian uncertainty, the increments $x_t - x_0$ are normal with mean $\theta\mu t$ and variance $\sigma^2 t$, with density function

$$f(x|\theta) = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left(-\frac{1}{2}\left(\frac{x-\theta\mu t}{\sigma\sqrt{t}}\right)^2\right). \tag{A.1}$$

Normalise $x_0 = 0$, so conditional on observing x_t, the likelihood of $\theta = 1$ is π_t which according to Bayes Law is

$$\pi_t = \frac{f(x_t|1)\pi_0}{f(x_t|1)\pi_0 + f(x_t|0)(1-\pi_0)}. \tag{A.2}$$

Using the formula for the density functions above, we find

$$\pi_t = \frac{\frac{\pi_0}{1-\pi_0}\exp\left(\frac{1}{2}\left(\frac{x_t}{\sigma\sqrt{t}}\right)^2 - \frac{1}{2}\left(\frac{x_t-\mu t}{\sigma\sqrt{t}}\right)^2\right)}{1+\frac{\pi_0}{1-\pi_0}\exp\left(\frac{1}{2}\left(\frac{x_t}{\sigma\sqrt{t}}\right)^2 - \frac{1}{2}\left(\frac{x_t-\mu t}{\sigma\sqrt{t}}\right)^2\right)} = \frac{\frac{\pi_0}{1-\pi_0}\varphi_t}{1+\frac{\pi_0}{1-\pi_0}\varphi_t}, \tag{A.3}$$

where we have used the definition of φ_t.

Using Ito's lemma and the definition of φ_t, we work out

$$d\varphi_t = \frac{\mu}{\sigma^2}\varphi_t dx_t \tag{A.4.a}$$

$$d\pi_t = \frac{\mu}{\sigma^2}\pi_t(1-\pi_t)dx_t - \frac{\mu^2}{\sigma^2}\pi_t^2(1-\pi_t)dt. \tag{A.4.b}$$

Using the change of probability measure $d\tilde{B}_t = \frac{1}{\sigma}dx_t - \frac{\mu}{\sigma}\pi_t dt$, the process π_t becomes a martingale, $d\pi_t = \frac{\mu}{\sigma}\pi_t(1-\pi_t)d\tilde{B}_t$.

A.2 Steady State in the Monitoring Region $\pi^{**} < \pi_t < \pi^*$:

In steady state the expected number of loans of a given quality $\bar{\pi} \in (\pi^{**}, \pi^*)$ need to remain constant. We argued in the text that for loans that are being monitored, the number of loans of any given type $\bar{\pi}$ will be the new loans that arrive minus the old loans that are withdrawn, plus the loans that are drawn to the barrier $\bar{\pi}$ from above (i.e. $\pi_{t-} > \bar{\pi}$) and below (i.e. $\pi_{t-} < \bar{\pi}$), and the loans on the barrier that go higher or lower. This problem can be posed as a reflecting barrier problem. Consider $\bar{\pi}$ a reflecting barrier from below and above. Then any loan with posterior process π_t that hits the barrier from above will have its sample paths replaced by the sample path $2\bar{\pi} - \pi_t$. Any loan with posterior process π_t that hits the barrier from below will have its sample paths replaced by the sample path $2\bar{\pi} - \pi_t$. The barrier $\bar{\pi}$ absorbs mass from these movements from the amount of time the reflected loans spend on the barrier itself. However, a diffusion spends no time at a reflecting boundary point, a fact that is known in the literature (see, for instance, p.16 of Borodin and Salminen (2002)). The end points π^{**} and π^{*} are, in contrast, absorbing or killing barrier points and the diffusions may spend some time on the boundary itself reflecting the time it takes the bank to transfer the loan from monitored status to being sold (again, see Borodin and Salminen (2002)).

B Proofs

B.1 Proof of Lemma 1:

The particular solution $\frac{c-m}{r}$ can be verified directly: since it is a constant we find $\mathbb{L}\frac{c-m}{r} = -(c-m)$ and it satisfies the ODE implied by $\mathbb{L}(V) + (c-m) = 0$. The two functions that span the solution space (i.e. the solutions to the homogeneous part of the ODE $\mathbb{L}V = 0$), are $\pi_t^{\frac{1}{2}(1-\xi)}(-1+\pi_t)^{\frac{1}{2}(1+\xi)}$ and $\pi_t^{\frac{1}{2}(1+\xi)}(-1+\pi_t)^{\frac{1}{2}(1-\xi)}$. This can be verified by straightforward differentiation.

These functions take values in the complex number space. A general solution is $A\pi_t^{\frac{1}{2}(1-\xi)}(-1+\pi_t)^{\frac{1}{2}(1+\xi)} + B\pi_t^{\frac{1}{2}(1+\xi)}(-1+\pi_t)^{\frac{1}{2}(1-\xi)}$ for arbitrary constants A and B, which according to the problem at hand must take values on the real number line only. We can write a general real solution, therefore, as

$$\begin{aligned}
&\bar{A}i^{-(1+\xi)}i^{(1+\xi)}\pi_t^{\frac{1}{2}(1-\xi)}(1-\pi_t)^{\frac{1}{2}(1+\xi)} + \bar{B}i^{-(1-\xi)}i^{(1-\xi)}\pi_t^{\frac{1}{2}(1+\xi)}(1-\pi_t)^{\frac{1}{2}(1-\xi)} \\
&= \bar{A}\pi_t^{\frac{1}{2}(1-\xi)}(1-\pi_t)^{\frac{1}{2}(1+\xi)} + \bar{B}\pi_t^{\frac{1}{2}(1+\xi)}(1-\pi_t)^{\frac{1}{2}(1-\xi)}
\end{aligned}$$

where the coefficients $\bar{A}$ and $\bar{B}$, as well as the sum itself, are real. This builds on the Euler's formula $e^{ix} = \cos(x) + i\sin(x)$, which for $x = \frac{\pi}{2}$ yields $e^{i\frac{\pi}{2}} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i$. Then $i^x = e^{i\frac{\pi}{2}x}$ and $i^{-x}i^x = e^{i\frac{\pi}{2}(x-x)} = 1$. $\square$

B.2 Proof of Proposition 1:

The system of equations can be directly derived from the value matching and smooth pasting conditions (b) and (c) in Definition 1. The condition that

$$V_M(\pi) \geq \max\left\{\frac{c}{r} - K(1-\pi), \frac{c}{r} - k(1-\pi) - g\right\}$$

ensures that the value function of the optimal stopping problem is indeed $V_M(\pi)$ for all $\pi \in [\pi^{**}, \pi^*]$. The functions $\pi^{\frac{1}{2}(1-\xi)}(1-\pi)^{\frac{1}{2}(1+\xi)}$ and $\pi^{\frac{1}{2}(1+\xi)}(1-\pi)^{\frac{1}{2}(1-\xi)}$ are both strictly convex, so it is possible to find coefficients $\bar{A}$ and $\bar{B}$ such that the function is strictly convex over the monitoring region, and strict convexity of V_M is clearly sufficient for V_M to strictly dominate both V_I and V_S since the latter two functions are both linear. It is also necessary, because if V_M is not convex over the whole range, it will be non-convex in the neighbourhood of π^* or π^{**}, and in this case the value of V_M will dip below $V_S - g$ if it is non-convex near π^{**} or below V_I if it is non-convex near π^*. $\square$

B.3 Proof of Proposition 2:

Claim: π^* is strictly decreasing in m.

Let $m_1 < m_2$. The subscript 1 indicates that the variable is corresponding to m_1 and 2 corresponding to m_2, except for functions f_i and g_i, $i = 1, 2$ defined in the main text—e.g. π_1^* and π_2^* correspond to the optimal upper threshold for m_1 and m_2 respectively. We prove the claim by contradiction by supposing $\pi_1^* < \pi_2^*$.

Since the functional form of $V_M(\pi)$ is affected by m, we let $V_M(\pi|m_i)$ represent $V_M(\pi)$ when $m = m_i$, $i = 1, 2$. The fact that $V_M(\pi|m_1)$ and $V_M(\pi|m_2)$ are both strictly increasing and strictly convex in π and the assumption that $\pi_1^* < \pi_2^*$ imply $V_M(\pi_1^*|m_1) < V_M(\pi_1^*|m_2)$.

It follows that

$$\begin{aligned} A_1 f_1(\pi_1^*) + B_1 f_2(\pi_1^*) - \frac{m_1}{r} &< A_2 f_1(\pi_1^*) + B_2 f_2(\pi_1^*) - \frac{m_2}{r} \\ &< A_2 f_1(\pi_1^*) + B_2 f_2(\pi_1^*) - \frac{m_1}{r}. \end{aligned}$$

Thus,

$$A_1 f_1(\pi_1^*) + B_1 f_2(\pi_1^*) < A_2 f_1(\pi_1^*) + B_2 f_2(\pi_1^*). \tag{B.1}$$

Also, by assumption $\pi_1^* < \pi_2^{**}$; thus, we have $\frac{\partial V_M}{\partial \pi}(\pi_1^*|m_1) > \frac{\partial V_M}{\partial \pi}(\pi_1^*|m_2)$. It follows that

$$A_1 f_1(\pi_1^*) g_1(\pi_1^*) + B_1 f_2(\pi_1^*) g_2(\pi_1^*) > A_2 f_1(\pi_1^*) g_1(\pi_1^*) + B_2 f_2(\pi_1^*) g_2(\pi_1^*). \tag{B.2}$$

Then, rearranging (B.1) yields

$$(A_1 - A_2) f_1(\pi_1^*) < -(B_1 - B_2) f_2(\pi_1^*). \tag{B.3}$$

Moreover, (B.2) yields

$$(A_1 - A_2) f_1(\pi_1^*) g_1(\pi_1^*) > -(B_1 - B_2) f_2(\pi_1^*) g_2(\pi_1^*). \tag{B.4}$$

From (B.4), if $A_1 > A_2$, then $B_1 > B_2$ must hold, which contradicts with (B.3). Hence, $A_1 < A_2$ must hold.

We now consider two cases: (i) $\pi_1^{**} \geq \pi_2^{**}$, and (ii) $\pi_1^{**} < \pi_2^{**}$. In case (i), there exists $\hat{\pi}$ such that $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_1) = \frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_2)$, or

$$A_1 f_1(\hat{\pi}) g_1(\hat{\pi}) + B_1 f_2(\hat{\pi}) g_2(\hat{\pi}) = A_2 f_1(\hat{\pi}) g_1(\hat{\pi}) + B_2 f_2(\hat{\pi}) g_2(\hat{\pi}).$$

Rearranging the equation, we have

$$(A_1 - A_2)f_1(\hat{\pi})g_1(\hat{\pi}) = -(B_1 - B_2)f_2(\hat{\pi})g_2(\hat{\pi}).$$

Since $A_1 < A_2$, and we know that $f_1, f_2, g_2 > 0$ and $g_1 < 0$, $B_1 < B_2$ must hold. Also, strict convexity of $f_2(\pi)$ in π implies that

$$f_2(\hat{\pi})g_2(\hat{\pi}) = f_2'(\hat{\pi}) < f_2'(\pi_1^*) = f_2(\pi_1^*)g_2(\pi_1^*).$$

Hence,

$$(A_1 - A_2)f_1(\hat{\pi})g_1(\hat{\pi}) = -(B_1 - B_2)f_2(\hat{\pi})g_2(\hat{\pi}) < -(B_1 - B_2)f_2(\pi_1^*)g_2(\pi_1^*).$$

This combined with (B.4) leads us to

$$(A_1 - A_2)\left[f_1(\hat{\pi})g_1(\hat{\pi}) - f_1(\pi_1^*)g_1(\pi_1^*)\right] < 0.$$

However, strict convexity of $f_1(\pi)$ in π implies that $f_1(\hat{\pi})g_1(\hat{\pi}) - f_1(\pi_1^*)g_1(\pi_1^*) < 0$, which is a contradiction.

Next, in case (ii), there exists $\hat{\pi}$ such that $V_M(\hat{\pi}|m_1) = V_M(\hat{\pi}|m_2)$. At $\hat{\pi}$, we must have $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_1) < \frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_2)$, or

$$A_1 f_1(\hat{\pi})g_1(\hat{\pi}) + B_1 f_2(\hat{\pi})g_2(\hat{\pi}) < A_2 f_1(\hat{\pi})g_1(\hat{\pi}) + B_2 f_2(\hat{\pi})g_2(\hat{\pi}).$$

Rearranging this, we have

$$(A_1 - A_2)f_1(\hat{\pi})g_1(\hat{\pi}) < -(B_1 - B_2)f_2(\hat{\pi})g_2(\hat{\pi}).$$

Since $A_1 < A_2$ and $f_1, f_2, g_2 > 0$ and $g_1 < 0$, $B_1 < B_2$ must hold. By a similar argument we had for case (i) leads us to a contradiction. This completes the proof of the first part of (a).

Claim: π^{**} is strictly increasing in m.

Again, let $m_1 < m_2$, and we show the claim by contradiction by supposing $\pi_1^{**} > \pi_2^{**}$.

The fact that $V_M(\pi|m_1)$ and $V_M(\pi|m_2)$ are both strictly increasing and strictly convex in π and the assumption that $\pi_1^{**} > \pi_2^{**}$ imply $V_M(\pi_1^*|m_1) < V_M(\pi_1^*|m_2)$. Thus,

$$\begin{aligned} A_1 f_1(\pi_1^{**}) + B_1 f_2(\pi_1^{**}) - \frac{m_1}{r} &< A_2 f_1(\pi_1^{**}) + B_2 f_2(\pi_1^{**}) - \frac{m_2}{r} \\ &< A_2 f_1(\pi_1^{**}) + B_2 f_2(\pi_1^{**}) - \frac{m_1}{r}. \end{aligned}$$

It follows that

$$A_1 f_1(\pi_1^{**}) + B_1 f_2(\pi_1^{**}) < A_2 f_1(\pi_1^{**}) + B_2 f_2(\pi_1^{**}),$$

and rearranging this, we have

$$(A_1 - A_2) f_1(\pi_1^{**}) < -(B_1 - B_2) f_2(\pi_1^{**}). \tag{B.5}$$

Also, we know that $\frac{\partial V_M}{\partial \pi}(\pi_1^{**}|m_1) < \frac{\partial V_M}{\partial \pi}(\pi_1^{**}|m_2)$, or

$$A_1 f_1(\pi_1^{**}) g_1(\pi_1^{**}) + B_1 f_2(\pi_1^{**}) g_2(\pi_1^{**}) < A_2 f_1(\pi_1^{**}) g_1(\pi_1^{**}) + B_2 f_2(\pi_1^{**}) g_2(\pi_1^{**}),$$

which is equivalent to

$$(A_1 - A_2) f_1(\pi_1^{**}) g_1(\pi_1^{**}) < -(B_1 - B_2) f_2(\pi_1^{**}) g_2(\pi_1^{**}). \tag{B.6}$$

From (B.5), if $B_1 > B_2$, then $A_1 < A_2$ must hold, which contradicts with (B.6). Thus, $B_1 < B_2$ must hold.

Now recall that $\pi_1^* > \pi_2^*$ from the first part of part (a). Thus, there exists $\hat{\pi} \in [\pi_1^{**}, \pi_2^*]$

such that $V_M(\hat{\pi}|m_1) = V_M(\hat{\pi}|m_2)$, or

$$(A_1 - A_2)f_1(\hat{\pi}) = -(B_1 - B_2)f_2(\hat{\pi}),$$

and $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_1) > \frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_2)$. Since $B_1 < B_2$, $A_1 > A_2$ must hold.

Now, $\pi_2^* < \pi_1^*$ implies that $\frac{\partial V_M}{\partial \pi}(\pi_2^*|m_1) < \frac{\partial V_M}{\partial \pi}(\pi_2^*|m_2)$, since $V_M(\pi|m_1)$ and $V_M(\pi|m_2)$ are strictly convex in π. With $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_1) > \frac{\partial V_M}{\partial \pi}(\hat{\pi}|m_2)$ for $\hat{\pi} \in [\pi_1^{**}, \pi_2^*]$, there must exist $\bar{\pi}$ such that $\frac{\partial V_M}{\partial \pi}(\bar{\pi}|m_1) = \frac{\partial V_M}{\partial \pi}(\bar{\pi}|m_2)$, or

$$(A_1 - A_2)f_1(\bar{\pi})g_1(\bar{\pi}) = -(B_1 - B_2)f_2(\bar{\pi})g_2(\bar{\pi}),$$

which contradicts with $A_1 > A_2$ and $B_1 < B_2$. This completes the proof of the second part of (a).

Next we prove part (b).

Claim: π^{**} is strictly increasing in κ.

Let $\kappa_1 < \kappa_2$, and just like the notation for part (a), all variables with subscript 1 correspond to κ_1, and those with 2 correspond to κ_2. We prove the claim by contradiction, so suppose $\pi_1^{**} > \pi_2^{**}$. Moreover, similar to the notation we used for $V_M(\pi)$ function above, we let $V_M(\pi|\kappa_i)$ represent $V_M(\pi)$ when $\kappa = \kappa_i$, $i = 1, 2$. Then, there exists $\hat{\pi}$ such that $V_M(\hat{\pi}|\kappa_1) = V_M(\hat{\pi}|\kappa_2)$, or

$$A_1 f_1(\hat{\pi}) + B_1 f_2(\hat{\pi}) = A_2 f_1(\hat{\pi}) + B_2 f_2(\hat{\pi}),$$

which is equivalent to

$$(A_1 - A_2)f_1(\hat{\pi}) = -(B_1 - B_2)f_2(\hat{\pi}).$$

Thus, $A_1 > A_2$ iff. $B_1 < B_2$. Also, at $\hat{\pi}$, $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|\kappa_1) > \frac{\partial V_M}{\partial \pi}(\hat{\pi}|\kappa_2)$, or

$$A_1 f_1(\hat{\pi})g_1(\hat{\pi}) + B_1 f_2(\hat{\pi})g_2(\hat{\pi}) > A_2 f_1(\hat{\pi})g_1(\hat{\pi}) + B_2 f_2(\hat{\pi})g_2(\hat{\pi}),$$

and rearranging, we obtain

$$(A_1 - A_2)f_1(\hat{\pi})g_1(\hat{\pi}) > -(B_1 - B_2)f_2(\hat{\pi})g_2(\hat{\pi}). \tag{B.7}$$

Since $A_1 > A_2$ iff. $B_1 < B_2$, to satisfy (B.7) it must be that $A_1 - A_2 < 0$ and $B_1 - B_2 > 0$. However, this requires that

$$(A_1 - A_2)f_1(\pi)g_1(\pi) > -(B_1 - B_2)f_2(\pi)g_2(\pi), \forall \pi \in [\pi_1^{**}, \min\{\pi_1^*, \pi_2^*\}],$$

or $\frac{\partial V_M}{\partial \pi}(\pi|\kappa_1) > \frac{\partial V_M}{\partial \pi}(\pi|\kappa_2)$ for all $\pi \in [\pi_1^{**}, \min\{\pi_1^*, \pi_2^*\}]$. We know that $V_M(\pi_1^{**}|\kappa_1) < V_M(\pi_1^{**}|\kappa_2)$, however. Contradiction.

<u>Claim: π^* is strictly increasing in κ.</u>

We prove by contradiction. Suppose $\pi_1^* > \pi_2^*$. We know that $\pi_1^{**} < \pi_2^{**}$. Then, there exists $\hat{\pi}$ such that $V_M(\hat{\pi}|\kappa_1) > V_M(\hat{\pi}|\kappa_2)$ and $\frac{\partial V_M}{\partial \pi}(\hat{\pi}|\kappa_1) = \frac{\partial V_M}{\partial \pi}(\hat{\pi}|\kappa_2)$, or

$$\begin{aligned} (A_1 - A_2)f_1(\hat{\pi}) &> -(B_1 - B_2)f_2(\hat{\pi}), && \text{(B.8)} \\ (A_1 - A_2)f_1(\hat{\pi})g_1(\hat{\pi}) &> -(B_1 - B_2)f_2(\hat{\pi})g_2(\hat{\pi}). && \text{(B.9)} \end{aligned}$$

It follows from (B.9) that $A_1 > A_2$ iff. $B_1 > B_2$. This together with (B.8) leads us to conclude that $A_1 > A_2$ and $B_1 > B_2$.

Now, we observe that $\frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_1) < \frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_2)$ for some $\pi \in [\pi_2^{**}, \pi_2^*]$, or

$$(A_1 - A_2)f_1(\pi)\left[(g_1(\pi))^2 + g_1'(\pi)\right] + (B_1 - B_2)f_2(\pi)\left[(g_2(\pi))^2 + g_2'(\pi)\right] < 0, \text{ for some } \pi \in [\pi_2^{**}, \pi_2^*].$$

However, this inequality contradicts with $A_1 > A_2$ and $B_1 > B_2$, since $f_1(\pi), f_2(\pi) > 0$, $(g_1(\pi))^2 + g_1'(\pi) > 0$ and $(g_2(\pi))^2 + g_2'(\pi) > 0$ hold. Contradiction. This completes the proof of part (b). □

B.4 Proof of Lemma 2:

Smooth pasting conditions at π^{**} require the following:

$$\frac{\partial V_M}{\partial \pi}(\pi^{**}|\kappa_0^*, m_0, \sigma_0/\mu_0) = k = \frac{\partial V_M}{\partial \pi}(\pi^{**}|\kappa_1^*, m_1, \sigma_1/\mu_1).$$

It follows from (4) that

$$\frac{\partial V_M}{\partial \pi}(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0) > \frac{\partial V_M}{\partial \pi}(\pi|\kappa_1^*, m_1, \sigma_1/\mu_1), \forall \pi \in (\pi^{**}, \min\{\pi_0^*, \pi_1^*\}). \tag{B.10}$$

Also, value matching conditions at π^{**} require

$$V_M(\pi^{**}|\kappa_0^*, m_0, \sigma_0/\mu_0) = V_S(\pi^{**}) = V_M(\pi^{**}|\kappa_1^*, m_1, \sigma_1/\mu_1).$$

Hence, we have

$$V_M(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0) > V_M(\pi|\kappa_1^*, m_1, \sigma_1/\mu_1), \forall \pi \in (\pi^{**}, \min\{\pi_0^*, \pi_1^*\}). \tag{B.11}$$

We now show that $\pi_0^* < \pi_1^*$ by contradiction. So, suppose instead $\pi_0^* \geq \pi_1^*$. Then, (B.11) implies

$$V_M(\pi_1^*|\kappa_0^*, m_0, \sigma_0/\mu_0) > V_M(\pi_1^*|\kappa_1^*, m_1, \sigma_1/\mu_1), \tag{B.12}$$

and (B.10) leads to

$$\frac{\partial V_M}{\partial \pi}(\pi_1^*|\kappa_0^*, m_0, \sigma_0/\mu_0) > k + \kappa_1^* = \frac{\partial V_M}{\partial \pi}(\pi_1^*|\kappa_1^*, m_1, \sigma_1/\mu_1). \tag{B.13}$$

Since $V_M(\cdot|\kappa_0^*, m_0, \sigma_0/\mu_0)$ is strictly convex, $\kappa_0^* > \kappa_1^*$ must hold. It follows that $V_I(\pi|\kappa_0^*) < V_I(\pi|\kappa_1^*)$ for all π by construction of V_I. Strict convexity of V_M together with (B.12) and

(B.13) imply

$$V_I(\pi|\kappa_0^*) < V_I(\pi|\kappa_1^*) < V_M(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0),$$

contradicting with the value matching condition at π_1^*. Hence, $\pi_0^* < \pi_1^*$ must hold, proving (b).

Next, we show $\kappa_0^* < \kappa_1^*$ by contradiction. So, suppose instead $\kappa_0^* \geq \kappa_1^*$. Then, $V_I(\pi|\kappa_0^*) \leq V_I(\pi|\kappa_1^*)$ for all $\pi < 1$. Since we know that $\pi_0^* < \pi_1^*$,

$$V_M(\pi_0^*|\kappa_0^*, m_0, \sigma_0/\mu_0) = V_I(\pi_0^*|\kappa_0^*) \leq V_I(\pi_0^*|\kappa_1^*).$$

Strict convexity of V_M however implies that $V_M(\pi_0^*|\kappa_1^*, m_1, \sigma_1/\mu_1) > V_I(\pi_0^*|\kappa_1^*)$, contradicting with (B.11). Thus, $\kappa_0^* < \kappa_1^*$ follows, completing the proof of (c).

Now, we know that $\pi_0^* < \pi_1^*$. Thus, $V_M(\pi|\kappa_0^*, m_0, \sigma_0/\mu_0) > V_M(\pi|\kappa_1^*, m_1, \sigma_1/\mu_1)$ for all $\pi \in (\pi^{**}, \pi_0^*]$. Also, we know that $V_I(\pi|\kappa_0^*) > V_I(\pi|\kappa_1^*)$ for all $\pi < 1$ while the value function for $\pi \leq \pi^{**}$ does not depend on $(m, \sigma/\mu)$. This establishes (a). $\square$

B.5 Proof of Proposition 3:

We first show that $\frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa(m, \sigma/\mu), m, \sigma/\mu)$ is strictly decreasing in m, for all σ/μ fixed. Suppose $0 < m_0 < m_1$. Since $\bar{\psi} = \frac{k}{2}(1-\pi^{**})^2$ is fixed by assumption, π^{**} is fixed for both m_0 and m_1. By letting A_0 and B_0 associated with m_0 and A_1 and B_1 with m_1 for every σ/μ fixed, the optimality conditions outlined in Proposition 1 imply:

$$A_0 f_1(\pi^{**}) + B_0 f_2(\pi^{**}) = \frac{m_0}{r} - k(1-\pi^{**}) - g$$
$$A_1 f_1(\pi^{**}) + B_1 f_2(\pi^{**}) = \frac{m_1}{r} - k(1-\pi^{**}) - g$$

which implies that $(A_0 - A_1)f_1(\pi^{**}) + (B_0 - B_1)f_2(\pi^{**}) = \frac{m_0 - m_1}{r}$. The optimality conditions also imply:

$$A_0 f_1(\pi^{**})g_1(\pi^{**}) + B_0 f_2(\pi^{**})g_2(\pi^{**}) = k$$
$$A_1 f_1(\pi^{**})g_1(\pi^{**}) + B_1 f_2(\pi^{**})g_2(\pi^{**}) = k$$

which implies that $(A_0 - A_1)f_1(\pi^{**})g_1(\pi^{**}) + (B_0 - B_1)f_2(\pi^{**})g_2(\pi^{**}) = 0$. Combining both we can find expressions for $(A_0 - A_1)$ and $(B_0 - B_1)$:

$$(A_0 - A_1) = \frac{m_0 - m_1}{r} \frac{1}{1 + g_2(\pi^{**})} \frac{1}{f_2(\pi^{**})}$$
$$(B_0 - B_1) = \frac{m_0 - m_1}{r} \frac{g_2(\pi^{**})}{1 + g_2(\pi^{**})} \frac{1}{f_1(\pi^{**})}$$

Since f_i, $i = 1, 2$, and g_2 are strictly positive, we find that $\text{sign}(A_0 - A_1) = \text{sign}(B_0 - B_1) = \text{sign}(m_0 - m_1)$, which implies that for $m_1 > m_0$, $A_1 > A_0$ and $B_1 > B_0$. Since f_i, $i = 1, 2$, are strictly convex, we have for every σ/μ fixed,

$$\begin{aligned} \frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_0^*, m_0, \sigma/\mu) &= A_0 \frac{\partial^2 f_1}{\partial \pi^2}(\pi) + B_0 \frac{\partial^2 f_2}{\partial \pi^2}(\pi) \\ &< A_1 \frac{\partial^2 f_1}{\partial \pi^2}(\pi) + B_1 \frac{\partial^2 f_2}{\partial \pi^2}(\pi) \\ &= \frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_1^*, m_1, \sigma/\mu), \end{aligned}$$

or in short, for every σ/μ fixed,

$$\frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_0^*, m_0, \sigma/\mu) < \frac{\partial^2 V_M}{\partial \pi^2}(\pi|\kappa_1^*, m_1, \sigma/\mu). \tag{B.14}$$

By Lemma 2, (B.14) implies that for every σ/μ fixed, $V(\pi_0|\kappa_0^*, m_0, \sigma/\mu) < V(\pi_0|\kappa_1^*, m_1, \sigma/\mu)$ for all π_0. By integrating over π_0, this implies $W(m_0, \sigma/\mu) < W(m_1, \sigma/\mu)$ for all σ/μ fixed. This completes the proof of part (a). Also, by Lemma 2, (b) and (c) directly follow from

(B.14). $\square$

B.6 Proof of Proposition 4

We will use the same strategy and show that an increase in σ/μ leads to a decrease in the convexity of the function $V_M(\pi)$. First, since σ/μ enters the equilibrium conditions only through the parameter ξ, and since an increase in σ/μ leads to an increase in ξ, we will consider a change in the equilibrium conditions from ξ to $\xi+\Delta$ (where obviously $\Delta > 0$). Corresponding to the case above, therefore, we find the following.

$$\begin{aligned} A_0 f_1(\pi^{**}|\xi) + B_0 f_2(\pi^{**}|\xi) &= \frac{m}{r} - k(1-\pi^{**}) - g, && \text{(B.15)} \\ A_1 f_1(\pi^{**}|\xi+\Delta) + B_1 f_2(\pi^{**}|\xi+\Delta) &= \frac{m}{r} - k(1-\pi^{**}) - g, && \text{(B.16)} \end{aligned}$$

and

$$\begin{aligned} A_0 f_1(\pi^{**}|\xi) g_1(\pi^{**}|\xi) + B_0 f_2(\pi^{**}|\xi) g_2(\pi^{**}|\xi) &= k, && \text{(B.17)} \\ A_1 f_1(\pi^{**}|\xi+\Delta) g_1(\pi^{**}|\xi+\Delta) + B_1 f_2(\pi^{**}|\xi+\Delta) g_2(\pi^{**}|\xi+\Delta) &= k. && \text{(B.18)} \end{aligned}$$

We can use the fact that $f_1(\pi^{**}|\xi+\Delta) = f_1(\pi^{**}|\xi)\pi^{**-\frac{\Delta}{2}}(1-\pi^{**})^{\frac{\Delta}{2}}$ and $f_2(\pi^{**}|\xi+\Delta) = f_2(\pi^{**}|\xi)\pi^{**\frac{\Delta}{2}}(1-\pi^{**})^{-\frac{\Delta}{2}}$, and also that $g_1(\pi^{**}|\xi+\Delta) = g_1(\pi^{**}|\xi) - \frac{\Delta}{2\pi^{**}(1-\pi^{**})}$ and $g_2(\pi^{**}|\xi+\Delta) = g_2(\pi^{**}|\xi) + \frac{\Delta}{2\pi^{**}(1-\pi^{**})}$. Rewriting (B.16) and (B.18) using these expressions we find the following.

$$A_1 f_1(\pi^{**}|\xi)\left(\frac{1-\pi^{**}}{\pi^{**}}\right)^{\frac{\Delta}{2}} + B_1 f_2(\pi^{**}|\xi)\left(\frac{\pi^{**}}{1-\pi^{**}}\right)^{\frac{\Delta}{2}} = \frac{m}{r} - k(1-\pi^{**}) - g, \qquad \text{(B.19)}$$

$$
\begin{aligned}
A_1 f_1(\pi^{**}|\xi) \left(\tfrac{1-\pi^{**}}{\pi^{**}}\right)^{\frac{\Delta}{2}} \left[g_1(\pi^{**}|\xi) - \tfrac{\Delta}{2\pi^{**}(1-\pi^{**})}\right] + & \\
B_1 f_2(\pi^{**}|\xi) \left(\tfrac{\pi^{**}}{1-\pi^{**}}\right)^{\frac{\Delta}{2}} \left[g_2(\pi^{**}|\Delta) + \tfrac{\Delta}{2\pi^{**}(1-\pi^{**})}\right] = & \quad k.
\end{aligned} \tag{B.20}
$$

Taking the difference between (B.15) and (B.19) and eliminating terms, we find

$$
\left[A_0 - A_1 \left(\frac{1-\pi^{**}}{\pi^{**}}\right)^{\frac{\Delta}{2}}\right] \left(\frac{1-\pi^{**}}{\pi^{**}}\right)^{\xi} + \left[B_0 - B_1 \left(\frac{\pi^{**}}{1-\pi^{**}}\right)^{\frac{\Delta}{2}}\right] = 0. \tag{B.21}
$$

Taking the difference between (B.17) and (B.20) and eliminating terms and using additionally (B.21), we find

$$
\left(\frac{1-\pi^{**}}{\pi^{**}}\right)^{\xi} \left[A_0\xi - A_1(\xi-\Delta) \left(\frac{1-\pi^{**}}{\pi^{**}}\right)^{\frac{\Delta}{2}}\right] = B_0\xi - B_1(\xi-\Delta) \left(\frac{\pi^{**}}{1-\pi^{**}}\right)^{\frac{\Delta}{2}}. \tag{B.22}
$$

Clearly, for $\Delta = 0$ the term inside the bracket on the left hand side of (B.22) becomes $A_0\xi - A_1\xi$ and the right hand side $B_0\xi - B_1\xi$, so we need that $A_1 = A_0$ and $B_1 = B_0$. We expand, therefore, the effects on both sides for small changes $d\Delta$ around the point $\Delta = 0$. Taking the point $A_1\xi$ as the starting point we find the expansion $\xi dA_1 - A_1\left(1 - \frac{\xi}{2}\ln\frac{1-\pi^{**}}{\pi^{**}}\right)d\Delta = 0$. Rearranging we find that $\frac{dA_1}{d\Delta} = \frac{A_1}{\xi}\left(1 - \frac{\xi}{2}\ln\frac{1-\pi^{**}}{\pi^{**}}\right)$. Taking the point $B_1\xi$ as the starting point we find the expansion $\xi dB_1 - B_1\left(1 - \frac{\xi}{2}\ln\frac{\pi^{**}}{1-\pi^{**}}\right)d\Delta = 0$. Rearranging again we find that $\frac{dB_1}{d\Delta} = \frac{B_1}{\xi}\left(1 - \frac{\xi}{2}\ln\frac{\pi^{**}}{1-\pi^{**}}\right)$.

Next, we need to look at the effect that a change in $\frac{\sigma}{\mu}$, or equivalently, ξ, has on the functions f_1 and f_2. We use the same idea as above to study the impact that the change in ξ has on the convexity of the value function V_M. Namely, by showing $\frac{\partial^3 V_M}{\partial\xi\partial\pi^2}(\pi;\xi) > 0$, we prove $V_M(\pi;\xi_0) < V_M(\pi;\xi_1)$ for all $\pi \in (\pi^{**}, \min\{\pi_0^*, \pi_1^*\}]$ when $\xi_0 < \xi_1$. Note that we are slightly abusing notation to simplify the expression – $V_M(\pi;\xi)$ is more precisely $V_M(\pi|\kappa^*(m, h^{-1}(\xi)), m, h^{-1}(\xi))$, where $h^{-1}(\xi)$ is the inverse function of $h(\sigma/\mu) = \sqrt{1 + 8r\frac{\sigma^2}{\mu^2}}$.

A change in the parameter ξ will lead to a change of the second derivative $\frac{\partial^2 V_M}{\partial \pi^2}(\pi;\xi)$,

$$\frac{\partial^3 V_M}{\partial \xi \partial \pi^2}(\pi;\xi) = \frac{\partial A}{\partial \xi}(\xi)\frac{\partial^2 f_1}{\partial \pi^2}(\pi|\xi) + \frac{\partial B}{\partial \xi}(\xi)\frac{\partial^2 f_2}{\partial \pi^2}(\pi|\xi) + A\frac{\partial^3 f_1}{\partial \xi \partial \pi^2}(\pi|\xi) + B\frac{\partial^3 f_2}{\partial \xi \partial \pi^2}(\pi|\xi).$$

We know from the above that

$$\begin{aligned}
\frac{\partial A}{\partial \xi}(\xi) &= \frac{dA}{d\Delta} = \frac{A_1}{\xi}\left(1 - \frac{\xi}{2}\ln\frac{1-\pi^{**}}{\pi^{**}}\right), \\
\frac{\partial B}{\partial \xi}(\xi) &= \frac{dB}{d\Delta} = \frac{B_1}{\xi}\left(1 - \frac{\xi}{2}\ln\frac{\pi^{**}}{1-\pi^{**}}\right).
\end{aligned}$$

We now need to work on the last two terms. By straightforward differentiation we find

$$\begin{aligned}
\frac{\partial^3 f_1}{\partial \xi \partial \pi^2}(\pi|\xi) &= \frac{f_1(\pi|\xi)}{2}\ln\frac{1-\pi}{\pi}\left(\frac{\xi^2-1}{4}\right)\left[\frac{1}{\pi(1-\pi)}\right]^2 + \frac{f_1(\pi|\xi)}{2}\xi\left[\frac{1}{\pi(1-\pi)}\right]^2, \\
\frac{\partial^3 f_2}{\partial \xi \partial \pi^2}(\pi|\xi) &= \frac{f_2(\pi|\xi)}{2}\ln\frac{\pi}{1-\pi}\left(\frac{\xi^2-1}{4}\right)\left[\frac{1}{\pi(1-\pi)}\right]^2 + \frac{f_2(\pi|\xi)}{2}\xi\left[\frac{1}{\pi(1-\pi)}\right]^2.
\end{aligned}$$

Using these results, some mechanical calculations will lead us to the following:

$$\frac{\partial^3 V_M}{\partial \xi \partial \pi^2}(\pi;\xi) = [Af_1(\pi|\xi) + Bf_2(\pi|\xi)]\,\frac{3\xi^2-1}{4\xi}\left[\frac{1}{\pi(1-\pi)}\right]^2 > 0.$$

Since an increase in $\frac{\sigma}{\mu}$ leads to an increase in ξ, the results follow immediately by applying Lemma 2. $\square$