PRE-CRISIS CREDIT STANDARDS AND THE SAVINGS GLUT

ADRIAN PENALVER

Date: August 20, 2018.

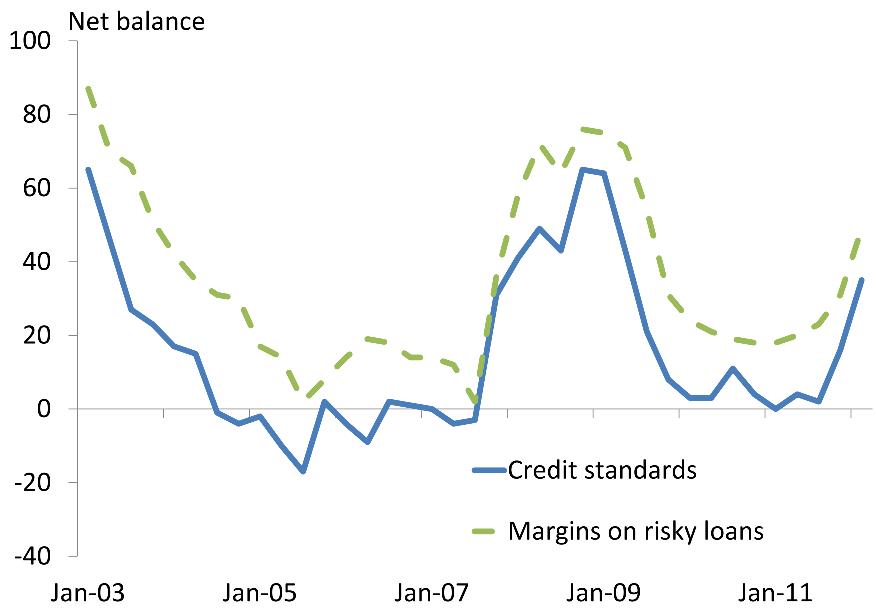
Address: Banque de France, 31 rue Croix-des-Petits-Champs, 75049 Paris CEDEX, FRANCE adrian.penalver@banque-france.fr.

ABSTRACT: This paper presents a theoretical model of how banks set their credit standards based on the firm dynamics framework of Hopenhayn (1992). The model examines how a bank sets its monitoring intensity in order to manage credit risk when it makes long duration loans to borrowers who have private knowledge of their project's stochastic profitability. An exogenous permanent increase in deposits in the model - a savings glut shock - replicates many of the stylised facts of credit conditions prior to the financial crisis: lower deposit rates, lower lending margins, weaker credit standards and more lending. A particularly interesting result of this simulation is that default rates initially fall before rising to a higher steady state value [and banks earn temporary excess profits] along this transition.

KEYWORDS: credit standards, credit risk, monitoring, savings glut.

JEL CLASS.: G21

FIGURE 1. Bank Credit Standards and Credit Spreads



Note: Net balance of respondents reporting a tightening or loosening of credit standards and widening or narrowing of credit spreads on risky loans from the Bank Lending Survey of the European Central Bank.

1. INTRODUCTION

The global financial crisis which began in the summer of 2007 and intensified in 2008 triggered the deepest economic downturn since the Great Depression. Why did it happen? There is, of course, no single or simple answer but a crucial element was that in the years leading up to the crisis, major western banks were lending on easier terms to less creditworthy borrowers so their balance sheets were more vulnerable when the crisis struck. This was exemplified by the rise in sub-prime mortgages in the United States but was also evident in the residential real estate booms in Spain and Ireland and in the corporate sector. Survey measures of credit standards fell in both the United States and the Euro-area from around 2003 and remained low right up until the crisis began. At the same time, however, margins on risky loans were falling. In the lead-up to the crisis, banks were taking greater risk for less compensation (see Figure 1)[1]

Why did they do it? Here, too, there is unlikely to be a unique explanation. A substantial literature is devoted to the risk-taking channel of monetary policy of Borio and Zhu (2012), including numerous empirical studies showing that the creditworthiness of new borrowers deteriorated when policy interest rates were low (see for example Jimenez, Ongena, Peydro, and Saurina (2014) and Ioannidou, Ongena, and Peydro (2015)) [2]

[1]The Senior Loan Officer Survey of the Federal Reserve Board exhibits a similar pattern.

[2] A number of similar papers have replicated these findings across a range of countries including Bonfim and Soares (2014), Paligorova and Sierra (2012), and Gerl, Jakubik, Kowalczyk, Ongena, and Peydro (2015).

An alternative explanation, and the one explored in this paps, argues that banks took more risk for less return because of a relative increase in *ex ante* savings. The 'savings glut' hypothesis contends that the supply of funds to the western financial system increased strongly during the early 2000s, particularly from emerging market economies (Bernanke (2005), Bernanke (2010) and King (2010)).[3]

Emerging market central banks bought government bonds from private investors who then needed alternative uses for their funds. This set up a 'search for the marginal borrower' by western banks. Other emerging markets were reluctant to borrow given their recent experience of crises. Instead, willing borrowers were found in real estate markets, particularly those who had previously been denied credit because of high default risk. With lower borrowing rates and rising asset prices, these look like better credit risk than previously. Mian and Sufi (2014) (chapter six) provide systematic evidence of the rise in lending to less creditworthy borrowers in the US. In particular they document that home ownership rates, which had stayed roughly steady since the mid-1960s, rose by around 4 percentage points during the boom (Mian and Sufi (2014) p. 77).

The objective of this paper is to present a model of bank credit standard setting in general equilibrium. The model described below will generate endogenously determined deposit interest rates, loan interest rates (and thus a loan spread), monitoring intensity, credit risk distribution and default risk. In response to a permanent positive shock to bank deposits come from outside the model, bank lending will increase, deposit rates and the loan spread will fall, monitoring will decrease, the credit distribution will initially improve but eventually deteriorate to a new steady-state in which the default rate is higher.

These features are obtained by adapting and extending the firm dynamics model of Hopenhayn (1992) to incorporate a banking relationship.[4] This unusual approach is based on a simple idea - the cross-sectional distribution of borrowers across the consolidated balance sheet of the banking sector should resemble the cross-sectional distribution of firms in the economy. This is realistic for European economies where bank finance is highly important but should be a reasonable approximation in other economies too. For simplicity, the model in this paper abstracts from the distribution of the size of firms, firm expansion and differences in the leverage of firms by assuming that all firms are the

[3] World foreign exchange reserves tripled in US dollar terms in just six years between 2001 and 2007 (according to IMF International Statistics). Private savings were also strong in Japan and Germany at that time.

[4] The model refers to firms and business lending but nothing would prevent a reader reinterpreting these as households taking out mortgages against a stochastic income process in which episodes of employment and unemployment coincide with buying a house with a mortgage and renting.

same size and borrow the same amount. The focus instead is on the distribution across profitability for which Hopenhayn (1992) is highly suited.

Using this approach, the evolution of the loan portfolio reflects the transition dynamics of firms. Firms enter and exit endogenously (through voluntary liquidation or default) and are simultaneously taking out or repaying loans. Individual firms are subject to persistent idiosyncratic profitability shocks that are privately observed by the owner and which determine how long they survive above various exit thresholds. Firms have limited liability and so below a certain level of losses, they will prefer to default with their losses passed onto the bank. The key idea of the paper is that because of this credit risk, banks will wish to influence the exit and continuation choices of firms. This is achieved in the model by assuming that a single bank sets its own continuation threshold, labelled a covenant, but it will have to undertake costly monitoring in order to enforce this threshold. The stochastic monitoring rate is the key choice variable of the bank.[5] The term credit standards is used hereafter to refer to the combination of the covenant threshold and the monitoring rate. [In general equilibrium the bank is subject to a zero profit condition and so is equivalent to the solution of a fully competitive banking system with no economies or diseconomies of scale.]

Different credit standards result in different equilibrium distributions of firms. Intuitively, tighter credit standards filter out a greater proportion of weaker firms and thereby improve the average credit quality of the loan portfolio. But tighter credit standards will be costly because the bank will have to pay to enforce them and because they transfer continuation rights from the owner of the firm to the bank. This reduces the utility of borrowers and thus the equilibrium loan interest rate the bank can charge. Different combinations of credit standards and market-clearing loan interest rates deliver different invariant distributions of firms and credit risk and different levels of profits. The bank is assumed to choose the profit-maximising set of loan conditions. Competition for deposits determines the deposit interest rate that satisfies the zero profit condition under profit maximisation.The model can also be extended to endogenise the covenant threshold and apply more sophisticated state-contingent monitoring - see Penalver (2016) and the main results are unaffected. Code to verify these statements is available online.

The model is closed by assuming that there is a fixed measure of agents in the economy each of whom has a fixed unit of capital. Running a firm is assumed to require two units so market clearing requires that half the agents lend their unit of capital to the other, intermediated through the banking system. This equilibrium condition constrains the

[5] Evidence that the frequency of monitoring is a deliberate choice for small business loans is presented in Minnis and Sutherland (2017).

choices the bank can make. Those not currently running a firm are researching an idea for a new firm, similar to the wage search model of McCall (1970).

The rest of the set-up of the model establishes the conditions under which the bank has an incentive to set credit standards and actively manage the composition of its balance sheet rather than rely entirely on the exit and entry choices of the agents. There are four key assumptions.

The first assumption is that borrowers are protected by limited liability. This caps downside risk to the entrepreneurs and creates default risk for the bank. Limited liability is not an option the bank would voluntarily offer but is imposed exogenously as a realistic feature of credit risk.

The second assumption is that the bank must pay a cost to seize and liquidate the assets of bankrupt entrepreneurs. This drives a wedge between the value of the bankruptcy option to the borrower and the cost of the option to the bank.

The third assumption is that there are entry and exit costs (as in Dixit and Pindyck (1994)). Without these switching frictions, competition for capital would ensure that only the most profitable projects would be in operation at any point in time. This would also be the lowest credit risk distribution, *ceteris paribus*, so there would be no need for the bank to act. With switching frictions, the entry threshold is above the exit thresholds. Naturally, projects at the lower end of the distribution are the ones with the highest subsequent default probability. These three assumptions ensure that the bank has an interest in the profitability distribution and thus the turnover rate of projects.

The fourth assumption is that it is costly for the bank to observe the profitability state of individual projects. With costly monitoring, the bank will generally choose to be less than fully informed about the state of its loan portfolio and will unknowingly roll over some loans that it would not if it knew the true state.

It is the interaction between these four assumptions that motivates the bank to manage credit risk. Without monitoring, borrowers will make individually rational continuation choices which expose the bank to under-compensated credit losses. Exit costs and limited liability give borrowers an incentive to gamble for resurrection. But the cost to the bank of a failed gamble by the borrower is greater than the value of bankruptcy to the borrower. A higher monitoring frequency can reduce credit risk but the loan contract is less valuable to the borrower and thus commands a lower loan interest rate in equilibrium. At the core of the model, therefore, is a trade-off between credit risk, the loan interest margin and monitoring costs. Equilibrium in the model is a deposit interest rate, a loan interest rate, a monitoring intensity and a profitability distribution where the equilibrium choice is constrained by the need to have the same measure of depositors as borrowers.

This equilibrium can then be shocked to consider the effect on monitoring intensity and loan interest rates of the savings glut (an external deposit supply shock).

The setting for this model is clearly quite different from the established literature on banking and credit risk. The bank is not faced by borrowers of fixed probability of default - the typical "high" or "low" risk project analysed first by Holmstrom and Tirole (1997) or independent shocks as in Haubrich (1989). Borrowers, instead, have state contingent default probabilties that vary persistently but stochastically through time. Persistent shocks to productivity and profitability and selective exit are robust features of the empirical literature on firm dynamics (see Foster, Haltiwanger, and Krizan (2006) and Foster, Haltiwanger, and Syverson (2008) for US evidence). This model is thus a first step in understanding the role of banks as "delegated monitors" in a realistic firm-dynamics setting.

Although the life expectancy of a new entrant is increasing in its initial state, firms can be in states most at risk of default at any age. Indeed, firms exit almost surely either through voluntary liquidation or bankruptcy and default. The bank therefore has no special interest in knowing the initial state of the firm and so does not screen or attempt to induce the firms to signal their initial state. In the extended model with state-contingent monitoring, Penalver (2016), knowledge of the initial state conditions the monitoring rate in early periods. This improves monitoring efficiency without changing the qualitative results of the model. It is also only in special circumstances that a bank would choose to be ignorant of the state of an individual borrower until default as in the costly state verification literature (such as Diamond (1984), Gale and Hellwig (1985) and Williamson (1986)). (This is equivalent to zero probability of interim monitoring.) In this framework, firms make state-contingent continuation decisions that transfer default risk to the bank. So the purpose of monitoring in this setting is to identify borrowers below the covenant threshold, withdraw the loans and thereby reduce the amount of risk shifting.

This different perspective allows focus to be drawn on new issues, in particular how bank credit standards shape the entry and exit decisions of firms and thus the allocation of capital. The bank here is engaged in an ongoing process of credit assessment rather than a one-off *ex ante* screening of initial applicants or a final *ex post* audit of failed projects.[6]. The relationship is not finite so that final payoffs can be used to discipline interim behaviour (as in Gertler (1992) or reward and punish to extract truthful revelation. Nor is the relationship infinite so the bank cannot expect to use statistical likelihood to check the veracity of reports.

[6]For example Broecker (1990), Ruckes (2004), and Dell'Ariccia, Marquez, and Laeven (2010), *ex post* auditing

This approach also draws attention to the relationship between the life cycle of firms and default risk. In this model all firms shut down eventually, almost surely, no matter how profitable they become at any point in their life. So what matters for credit risk in equilibrium is the manner and circumstances in which firms depart. This is important when one considers that the majority of firms exit in an orderly way and only a minority of firms (or households) end lending arrangements through default. Thus what is important is not that firms experience adverse shocks but why they choose one exit method over another. The purpose of the model is to show how the policies of the bank influence these choices in equilibrium.

An important feature of the model is that it is set in general rather than partial equilibrium. All the agents in the economy interact with the bank, either as depositors or borrowers. To achieve balance sheet equilibrium, the bank must offer a contract which induces half the agents to choose to be on each side. This is also equivalent to a capital clearing condition. Thus loan contract terms affect the supply of savings as well as the demand for loans. If one believes that credit standards are a macroeconomic phenomenon, then one ought to examine them in an aggregate setting.

The remainder of the paper is organized into four sections as follows. Section 2 sets out the assumptions in more detail and Section 3 solves for the optimal behavior of individual agents for any given contract terms. To assist in understanding the subsequent choice of the bank, an illustrative numerical example is presented. Section 4 explains equilibrium bank behavior and the profit-maximising choice of credit standards and loan interest rate. Section 5 discusses the equilibrium properties of the model and Section 6 uses the model to consider the effect of a global savings glut on equilibrium interest rates, credit standards and credit risk. Section 7 concludes.

2. MODEL

2.1. **Agents and production.** The economy contains a measure 1 of infinitely small, *ex ante* identical and infinitely living risk-neutral agents and a representative bank. Time is discrete and future payoffs are discounted at rate β. In each period the agents are defined by two state variables. One state is an endogenously chosen occupation as either an entrepreneur (E) or inventor (I). The second state is an exogenously determined idiosyncratic profitability level, a, which can be thought of as including technical productivity, consumer preferences, degree of market power and managerial talent. a is drawn stochastically every period from the compact set $\{A \in \mathbb{R} : 0 \leq a \leq 1\}$.

Output in the economy is produced by firms which have conditional gross payoffs each period of $q(a)$. Running a firm requires two units of capital and each agent is endowed with only one. It is assumed (to economise on dimensionality) that agents cannot add to or draw down on their units of capital.[7] Thus the endogenous allocation of capital across profitability states is determined by the occupation decisions. In addition, market clearing (or equivalently, the balance sheet condition of the bank) requires that half the agents will have to lend their unit of capital to the other half. The distribution of agents across profitability states is identical to the distribution of capital across profitability and ultimately the distribution of credit risk. Throughout their infinite lives, agents cycle stochastically but endogenously between the two occupations.

During each period inventors receive an "idea" for a new firm with a profitability state drawn from the continuous i.i.d distribution $G(a)$.[8] Inventors weigh up paying start-up costs C to enter production next period based on this profitability draw or waiting for a better idea in the future.[9] Inventors are of measure N (which in equilibrium will equal a $\frac{1}{2}$) and while they wait, they deposit their capital at the bank on which they receive an endogenously determined deposit rate, δ.

The others agents, the entrepreneurs, are currently in production for which they will have borrowed capital from the bank at an endogenously determined loan interest rate, ρ. Entrepreneurs' per period payoffs ("profits" and "losses"), $q(a) - \rho$, are a function of their idiosyncratic profitability states and the loan interest rate. These profitability states evolve according to a first-order Markov process with a conditional cumulative distribution function, $F(a'|a)$. The following assumptions on the transition process and profit function are taken directly from Hopenhayn (1992):

A: (i) $F(a'|a)$ is continuous in a and a'; (ii) Profitability shocks are persistent and so $F(a'|a)$ is strictly decreasing in a. (iii) But profitability shocks eventually die out and the monotone mixing condition is satisfied: $F^n(\epsilon|a) > 0\ \forall \epsilon$ for some n where $F^n(\epsilon|a)$ is the conditional cumulative probability distribution of profitability in n periods time given a. So from any given level of profitability, it is possible to transit to any other profitability interval in a finite number of periods. Since there are exit

[7] It is trivial to extend the model to include heterogeneity in endowment, project size or borrowing limit so these normalisations are uncontroversial. It is a much more serious restriction to rule out time varying investment and time varying credit. Technically the possibility of wealth accumulation creates the well-known problem that some agents could save their way out of the model. More specifically in the context of this model, decisions to apply for credit or pay down debt could provide free information to the creditor about the idiosyncratic state of the project and thereby eliminate the need to undertake costly monitor. This signaling could then by muddied by assuming adjustment costs for investment and debt that create zones of inactivity. Including these aspects would substantially complicate what is already a complex model and it is not clear that the main qualitative results of the paper would be affected.

[8] This is similar to the way wage offers are received in labour search models such as McCall (1970).

[9] Start up costs are large in many countries. In Djankov, Porta, Lopez-De-Silanes, and Shleifer (2002) just the official costs of entry range from 0.5% of per capita GDP in the US to 460% of per capita GDP in the Dominican Republic.

thresholds, this assumption implies that all firms will almost surely close at some future point.

B: (i) $q(a)$ is continuous and; (ii) strictly increasing in a.

In each period, entrepreneurs decide whether to continue in production next period or to exit. If they do not wish to continue, there are two exit options.

- "Orderly" exit occurs if the entrepreneur absorbs current period payoffs and pays a liquidation cost L to close the project. These liquidation costs might be pecuniary such as termination pay, liquidating stock at below cost and administrative costs or non-pecuniary such as lost human capital and reputation.[10]
- "Default" occurs if the entrepreneur pays an exogenous cost B to file for bankruptcy protection in which case current period pay-offs (naturally losses) are excused (including repayment of loan interest). If the entrepreneur defaults, the bank must pay a cost K to seize the capital of the firm and absorbs the firm's losses.[11]

(For notational convenience it is assumed that C, L and B are paid in the following period.)

These options for the entrepreneur will partition the set A into three regions - continuation, orderly exit and default - delineated by threshold values a_X and a_D (exit and default, respectively) and the value of B is assumed to be always calibrated for the sensible ordering $a_X > a_D$. If the model had a continuous time shock process, then agents experiencing negative profitability shocks would always enter the voluntary exit region first and default would never arise. In discrete time, there is default because some firms jump over the orderly exit region $(a_D, a_X]$ so the discrete time assumptions could be considered as a proxy for discontinuous shocks. By assumption A(iii), project profitability will almost surely pass below one of these thresholds and the firm is terminated on the first occasion. When entrepreneurs exit, they become inventors again so the value of an entrepreneur's outside option is the expected value of being an inventor. Defaulting entrepreneurs are indistinguishable from other inventors.[12]

To keep the model as simple as possible and to focus on the main mechanism of interest, all decisions take place at the extensive margin. Aggregate variables are thus simple integrals over measures of agents. Likewise, expectations of payoffs are simple integrals

[10] Ramey and Shapiro (2001) describe the heavy discounts on machinery sold during the closure of aircraft manufacturing plants.

[11] In reality, of course, banks are not responsible for the losses incurred by their borrowers but this is the simplest assumption that ensures that all output is allocated.

[12] It is just convenient to recycle defaulters in this way. Nothing of any substance would change by assuming defaulters are excluded forever but new inventors are born at the same steady-state rate.

over profitability states only. It was to turn off the intensive margin that agents were assumed to be unable to change their capital holdings.[13] Relatedly, it will be assumed that projects require 2 units of capital. Entrepreneurs do not borrow directly from inventors (perhaps for the reasons described by Diamond (1984)) and must get a loan of 1 unit from the bank.[14] These simplifications ensure that when an inventor switches to being an entrepreneur, she simultaneously takes out a bank loan of a size that remains fixed until she decides to exit. As a result, the only endogenous distribution involved in the model is the measure of entrepreneurs over the set of profitability states.

Finally, this is a model based on incentives and it will be (implicitly) assumed that all agents have an exogenous endowment of income every period regardless of their circumstances sufficient to cover any expenses or losses. This simply rules out having to consider situations in which agents want to pay but cannot. Decisions are determined by the profitability state alone.

2.2. **Banking.** Capital in this model is intermediated between inventors and entrepreneurs by a single representative bank. The market will be contestable so that the bank will earn zero profits in equilibrium so the representative bank assumption is equivalent to assuming that there are constant returns to scale in banking. Since this is a model of purely idiosyncratic risk, the steady state distribution faced by the bank is invariant. The bank is assumed to offer the following deposit and loan contracts:

- Deposits earn an interest rate δ and can be withdrawn at the end of any period.
- The loan contract specifies an interest rate ρ, a monitoring intensity φ (where $0 \leq \varphi \leq 1$) and a covenant specifying a minimum profitability level ξ. For notational simplicity, the parameters of the bank contract are summarised by $\psi = \{\rho, \varphi, \xi\}$. The loan is nominally indefinite but both parties have an option to terminate it each period. The borrower has the option to repay the loan if she decides to exit production. And the bank can demand repayment if it discovers that the covenant condition has been breached. The covenant condition, ξ, will correspond to a threshold value, $\xi = q(a_T) - \rho$, at which the bank exercises its right to terminate the loan. Since the bank uses the covenant to protect its interests, it follows that ξ must imply a trigger value of a at least as high as that at which entrepreneurs voluntarily exit or else the covenant would be redundant.

There is no theoretical justification for these contract forms. They were chosen because they mimic key features of actual bank contracts and they are very simple and fit

[13] For a similar model with entrepreneurial wealth accumulation which is used to explain firm size dynamics, see Arellano, Bai, and Zhang (2012).

[14] This implied leverage ratio could be set to any constant without loss of generality with an appropriate adjustment to the market clearing condition.

easily into the recursive structure of the overall model. It is important to note that the bank commits to this loan contract. As will be discussed in more detail later, this contract has an element of insurance for the borrower who will be able to survive through low-profitability states. During these low profitability states, the value of the contract could well be negative because of the heightened risk of default and will be certainly less valuable than withdrawing the loan and offering the funds to a potential new entrant. The bank gains because the loan contract is profitable when the borrower is in a high-profitability state and thus far from default. More complex loan contracts that address some of the obvious weaknesses of this simple contract are the subject of ongoing research.

In this version of the model, δ, ρ and φ are endogenous.

3. Equilibrium behaviour of the agents

This is a recursive model and in each period the move order is the following:

(1) Agents enter the period in their previously chosen activity state (inventor or entrepreneur) and then draw their idiosyncratic shocks. The inventors get a new idea from $G(a)$ and entrepreneurs get an update of their profitability according to $F(a'|a)$.
(2) Entrepreneurs decide whether to continue with production next period or to exit either voluntarily or by defaulting. Payoffs are received and loan interest paid by non-defaulting entrepreneurs. Entrepreneurs who exit voluntarily repay their loan. Inventors receive deposit interest and decide whether to enter production next period based on their profitability draw.
(3) The bank monitors continuing loans at the stochastic rate φ and recalls the loans of all entrepreneurs found below the covenant profitability threshold state a_T.
(4) The bank receives deposits from waiting and new inventors and makes additional loans to entering entrepreneurs.

The following two sections formalise the analysis of the choices of the productive and inventors.

3.1. **Entrepreneurs.**

Depending on the idiosyncratic profitability state, a, an entrepreneur chooses between declaring bankruptcy and defaulting, exiting in an orderly fashion or continuing. If she decides to go bankrupt to escape $q(a) - \rho$, she pays B next period and switches to being an inventor. The value of bankruptcy is thus

$$V_B = \beta\{E\left[V_I(a',.)\right] - B\} \tag{1}$$

where the value function of an inventor is denoted $V_I(a, V_E)$ and $E\left[V_I(a',.)\right] = \int_A V_I(a',.)G(da')$.

If she chooses orderly exit from production, she absorbs current losses, pays liquidation costs L next period and also enters next period as an inventor. The value of orderly exit conditional on state a is

$$V_X(a) = q(a) - \rho + \beta\{E\left[V_I(a', .)\right] - L\} \tag{2}$$

The remaining option is to continue in production next period. Naturally the conditional value of continuing in production is to receive current payoffs and the discounted expected value of being an entrepreneur in the next period.

$$V_C(a) = q(a) - \rho + \beta\{E\left[V_E(a', \psi, V_I) \mid a\right]\} \tag{3}$$

where the value function of an entrepreneur is denoted $V_E(a; \psi, V_I)$.

Given the options available, the value of being an entrepreneur at the moment the shock is revealed is:

$$V_E(a, \psi, V_I) = \max\{V_B, V_X(a), V_C(a)\} \tag{4}$$

There is a natural ordering of the choices facing an entrepreneur. Bankruptcy costs will be assumed to be sufficiently large that entrepreneurs only choose this form of exit when facing a very bad profitability state. So the default threshold is determined by a comparison of the value of default and the value of orderly exit. It is straightforward to see from equations (1) and (2) that entrepreneurs will default for all values of $a \leq a_D$ where

$$q(a_D) = \beta(L - B) + \rho \tag{5}$$

The threshold for orderly exit, a_X, results from the comparison of $V_X(a)$ and $V_C(a)$. The only challenging aspect of this problem is the conditional expected value of being an entrepreneur next period at the time the decision is made: $E\left[V_E(a', .) \mid a\right]$. Consider first an entrepreneur with profitability above the loan covenant threshold, $a > a_T$. In this case the entrepreneur faces no risk if the bank randomly selects her loan to monitor, so she can ignore the monitoring option of the bank and

$$E\left[V_E(a', .) \mid a \geq a_T\right] = \int_A V_E(a\prime, .)F(da'|a)$$

The situation is more complex for an entrepreneur with $a_D < a \leq a_T$. In this case, if the entrepreneur decides to continue and escapes monitoring (with probability $1 - \varphi$) then the entrepreneur gets the conditional expected value of being an entrepreneur in the next period. If the entrepreneur tries to continue but is monitored then the loan is recalled by the bank, the project is shut down and the agent involuntarily reverts to being an inventor. Therefore for $a_D < a \leq a_T$

$$E\left[V_E(a';.) \mid a_D \leq a < a_T\right] = (1-\varphi)\int\limits_A V_E(a\prime;.)F(da'|a) + \varphi\left(\int\limits_A V_I(a';.)G(da') - L\right)$$

The voluntary exit threshold a_X is the value of current period profitability at which an entrepreneur is indifferent between continuing or exiting voluntarily.[15] Some simple cancelling defines the voluntary exit threshold implicitly as:

$$\int\limits_A V_E(a\prime,.)F(da\prime|a_X) = \int\limits_A V_I(a',.)G(da') - L \tag{6}$$

3.2. **Inventors.** We turn now to the decision by inventors whether or not to enter production. Unlike in Hopenhayn (1992), inventors are assumed to draw a profitability level *before* they decide whether or not to enter although they cannot begin production until the *following* period. Each period, an inventor gets one idea with a profitability level drawn from $G(a)$. The agent can either decide to pay the cost of starting up a firm, C, and enter production next period or keep his capital on deposit at rate δ for another period. Profitability next period will be subject to an idiosyncratic shock according to the same function F as existing firms. So the expected value of entering production is equal to the expected value of being an existing entrepreneur at the same level of profitability net of C. An inventor this period receives the interest on his deposit for this period and the discounted expected value of the maximum of the choice between entering or remaining as an inventor the following period. The value function for an inventor with a profitability draw of a is therefore:

$$V_I(a, V_E) = \delta + \beta \max\left\{\int\limits_A V_E(a',.)F(da\prime|a) - C, \int\limits_A V_I(a',.)G(da')\right\} \tag{7}$$

The threshold level of profitability at which inventors will choose to enter is where the expected value of being an entrepreneur at that level of profitability net of start up costs matches the expected value of waiting. a_E is thus determined implicitly by:

$$\int\limits_A V_E(a',.)F(da\prime|a_E) - C = \int\limits_A V_I(a',.)G(da') \tag{8}$$

The right hand side is not contingent on the current state and in equilibrium will be a known constant.

The equilibrium behaviour of entrepreneurs and inventors is described in the following propositions:

[15] Since the bank only monitors continuing loans, these are always above the default threshold and so any agent forced to repay the loan will prefer to exit in an orderly fashion rather than default.

PROPOSITION 1– *Given Assumptions A and B and a banking contract ψ, unique, bounded and mutually consistent functions $V_E(a, \psi, V_I)$ and $V_I(a, \psi, V_E)$ exist.*

Proof. The details of the proof are in the appendix. □

PROPOSITION 2– *These value functions yield unique and continuous functions in ψ for a_E and a_X. $a_E(\psi)$ and $a_X(\psi)$ are strictly increasing in ρ.*

Proposition 2 states there are unique values of the entry and exit thresholds, a_E, a_D and a_X which, along with the profitability threshold implied by the loan covenant, a_T, completely summarise the equilibrium behaviour of agents for a given loan contract ψ and deposit rate δ. So from here on we can dispense with the value functions.

It turns out that the key condition that proves Proposition 2 is that $\beta < 1$. The intuition for this condition is worth understanding. In this model, agents cycle endogenously but stochastically between being inventors (depositors) and entrepreneurs (borrowers) throughout their infinite lives. The equilibrium cross-sectional distribution across idiosyncratic states is also the proportion of time that each agent will spend in each state throughout their life. As $\beta \to 1$, the agent comes to value each of these infinte future states with equal weight and the current state becomes of vanishing relative importance. The value functions of all agents converge towards a common value (although in the limit this explodes for $\beta = 1$). This value is common between inventors and all entrepreneurs.

As $\beta \to 1$, a rise in ρ decreases all the value functions by the same amount. Without any difference in the relative value of being in different endogenous states, there is no change in behaviour and thus no effect on the switching thresholds. With $\beta < 1$, the current state has more weight in the value function than the common expected future of all agents. In this case, an increase in ρ reduces the common expected future value of all agents but reduces the value of being an entrepreneur in the current period by more because this rate is paid now. This reduces the relative value of being an entrepreneur and raises the threshold value of the idiosyncratic state at which the marginal entrepreneur is just indifferent between continuing or exiting.

This section appears very complicated but is intuitively very simple. Entrepreneurs are choosing whether to continue, exit voluntarily or default. The bank also has the option to monitor each loan and withdraw funding (and thus shut down the firm) for all those found to be below a covenant threshold. The inventors are choosing whether to enter production or not depending on their initial idiosyncratic profitability draw. These various options have a natural ordering so that there are unique values of a at which the agents switch from one option to another. These thresholds, which completely summarise the behaviour of the agents, are influenced by the contract terms of the loan. Intuitively, the less attractive the terms of the loan contract, the higher the various thresholds. For example, if the loan interest rate is higher, then entering becomes less attractive and inventors will only be willing to do so at a higher value of a.

3.3. **Cross-sectional distribution of firms.** Define $H(\psi)$ as the cumulative distribution of firms by current idiosyncratic state at the end of each period conditional on the loan terms ψ.[16]

From period t to t', $H(\psi)$ will evolve into $H'(\psi)$ in three ways. First, there will be the new entrants. These will be inventors who received an idea for a new firm with $a' \geq a_E(\psi)$. Second it will contain the distribution of those firms who evolved into an idiosyncratic state above the voluntary exit threshold (and therefore choose to continue) but are not monitored in period t'. Third, it will contain firms who are monitored but have evolve into a state $a' > a_T(\psi)$ and thus are allowed to continue by the bank.

This evolution of the distribution of firms, $H(\psi)$, can be summarised in the following transition function:

$$H'(\psi) = P(\psi)H(\psi) + Ng(\psi) \tag{9}$$

where $P(\psi)$ is a bounded linear operator on the space of positive bounded measures defined by $PH(\Gamma, \psi) = \int P(a, \Gamma, \psi)H(da)$ for all borel sets Γ in A. $P(\psi)$ is defined as

$$P(a, \Gamma, \psi) = \begin{array}{lll} \int\limits_{\Gamma\cap[a_T(\psi),1]} F(d\gamma|a) & & \text{for } \Gamma \cap [a_T(\psi), 1] \\ (1-\varphi) & \int\limits_{\Gamma\cap[a_X(\psi),a_T(\psi))} F(d\gamma|a) & \text{for } \Gamma \cap [a_X(\psi), a_T(\psi)) \\ 0 & & \text{for } \Gamma \cap [0, a_X(\psi)) \end{array}$$

and

$$g(\Gamma, \psi) = \begin{array}{ll} G(\Gamma) & \text{for } \Gamma \cap [a_E(\psi), 1] \\ 0 & \text{for } \Gamma \cap [0, a_E(\psi)) \end{array}$$

An invariant steady state distribution, $\bar{H}(\psi)$, occurs if

$$\bar{H}(\psi) = P(\psi)\bar{H}(\psi) + Ng(\psi)$$

PROPOSITION 3– *For each ψ and arbitrary N there is a unique invariant distribution, $\bar{H}(\psi)$.*

Uniqueness of the distribution is, of course, crucial for the validity of the comparison of the equilibrium results in Section 6.

[16] Implicitly δ is given too.

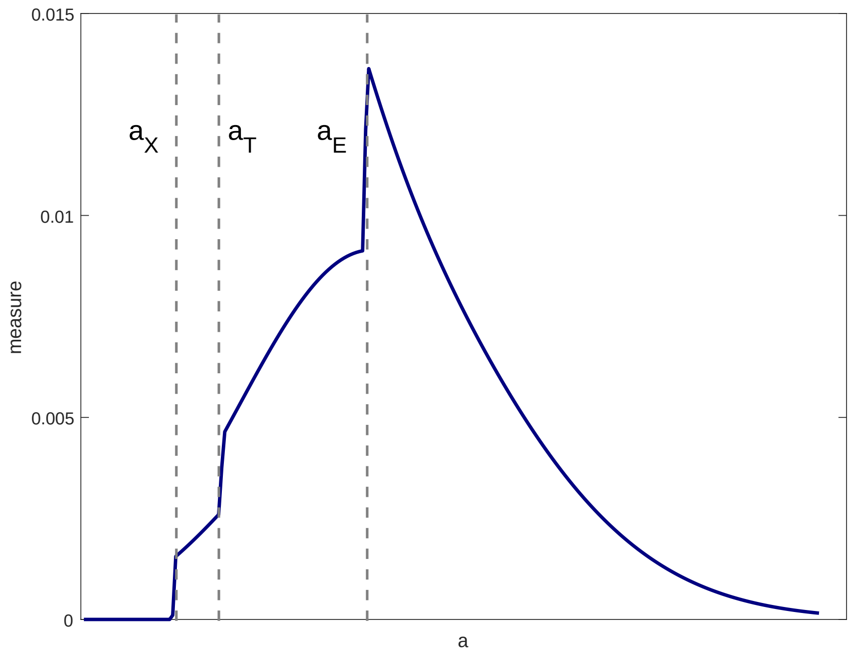


FIGURE 2. Illustrative Invariant Distribution

Note: The figure presents the distribution function across idiosyncratic profitability states a.

4. EQUILIBRIUM CHOICES OF THE BANK

4.1. **Illustrative numerical example.** Before turning to the decision of the bank, it is useful to give an illustrative example of an invariant distribution derived from the model and decompose the transition equation (16). The parameterisation for Figure 2 is entirely illustrative but was chosen to give roughly sensible credit spreads and entry, exit and default rates.[17] The monitoring intensity, which will be endogenised later, is here assumed to be strictly positive but less than 1. The parameters used are described in the Appendix and the code to run the model is available online.

It is easy to see the influence of the three behavioural thresholds on the end of period distribution. Below a_X, there are no entrepreneurs in the distribution because they have either defaulted or exited voluntarily. Between a_X and a_T we have entrepreneurs that are in breach of the loan covenant but have escaped monitoring. There is a concentration of entrepreneurs just above the entry threshold, a_E. There is a long right tail to this distribution. These are entrepreneurs who have either entered with a very high initial profitability level or entered and subsequently experienced predominantly positive profitability shocks.

[17] The example uses a normal distribution for $G(a)$ and an AR(1) process for $F(a'|a)$ using a Tauchen matrix approximation (see Tauchen (1986)).

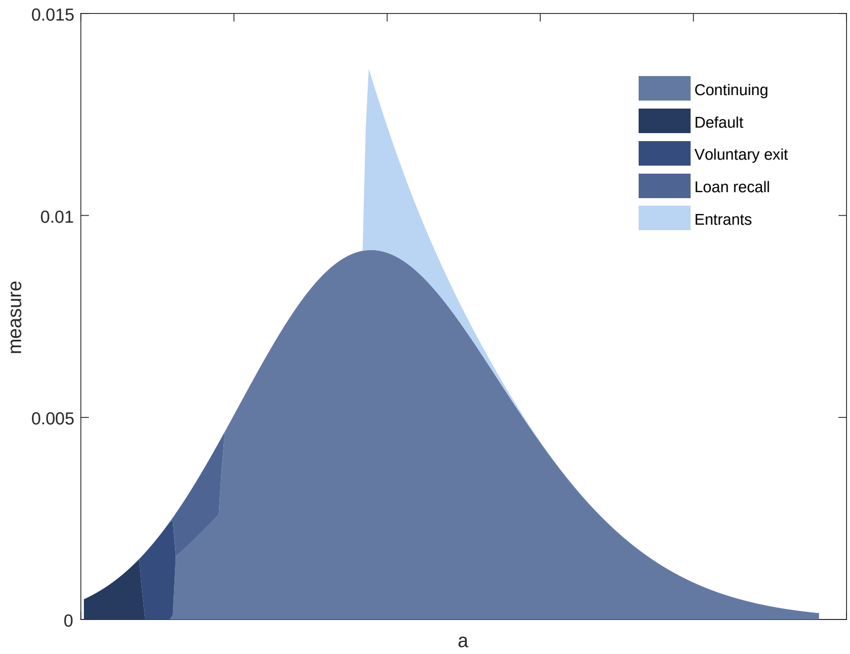


FIGURE 3. One Period Transition of the Invariant Distribution

Figure 2 also helps draw out the importance of start-up and liquidation costs in driving the results of the model. If we combine equations (6) and (8) which determine the entry and exit thresholds then we obtain

$$\int_A V_E(a\prime, .)F(da\prime|a_X) = \int_A V_E(a', .)F(da\prime|a_E) - C - L \tag{10}$$

from which it can be easily seen that if $C = L = 0$, then $a_E = a_X$. Since a covenant threshold is only relevant in the interval $(a_X, a_E]$, if we have $a_E = a_X$, then a_T is redundant and so is bank monitoring. In this case, even though there is a positive probability that profitability falls below the default threshold, a_D, there is no incentive for bank monitoring. This occurs because in the frictionless entry and exit case, borrower behaviour is completely aligned with the interests of the bank. Borrowers only continue in situations in which they would also wish to enter. Frictionless entry and exit always selects the most profitable firms given the profitability processes and thus the lowest possible credit risk. So asymmetry of information has no bite when exit and entry is costless. This ability to rely on borrower behaviour breaks down when there are entry and exit costs because private choice by borrowers no longer selects the lowest credit risk portfolio. By setting a covenant threshold and monitoring stochastically, the bank can alter the distribution of credit risk.

Figure 3 illustrates the one period transition of the distribution in Figure 2. Looking from right to left, one can see that the upper tail of the distribution is entirely driven by the presence of a small number of existing entrepreneurs experiencing positive shocks.

Since on average entrepreneurs with positive profitability experience a reversion towards the mean (of zero), there is a noticeable deterioration in the average quality of continuing entrepreneurs. The distribution is refreshed by the entry of new entrepreneurs clustered above the entry threshold. Moving further to the left, a number of entrepreneurs fall below the threshold a_T but above a_X. These are the entrepreneurs who want to continue but are in breach of the loan covenant. φ proportion of these entrepreneurs are monitored, have their loans revoked and exit and $1 - \varphi$ are able to continue. Moving further to the left, there are entrepreneurs who fall below a_X but above a_D and exit voluntarily. Finally, there is a portion of the distribution who falls below a_D and defaults.

The model structure and Figures 2 and 3 are broadly consistent with the empirical evidence on firm dynamics. Studies using US data, for example Doms and Bartelsman (2000), Baily, Bartelsman, and Haltiwanger (2001) and Foster, Haltiwanger, and Krizan (2006), show that there are wide distributions of profitability and productivity within industry classifications and that firm-level shocks are highly persistent. Farinas and Ruano (2005) use Spanish manufacturing data and show that the productivity distribution of exiting firms is stochastically dominated by the distribution of continuing firms and that the productivity of entering firms is stochastically dominated by continuing firms.

4.2. **Equilibrium bank behaviour.** We can now turn to the bank's choice of parameters ρ and φ in the loan contract $\psi = \{\rho, \varphi, \zeta\}$. Proposition 3 asserted that there is a unique invariant distribution for any loan contract, $\psi = \{\rho, \varphi, \zeta\}$ and measure of inventors, N. The bank, however, is constrained in its choice of loan contract by the need to finance its loans by deposits.[18] Using the simplifying assumption made earlier that all agents have a fixed unit of capital but projects require 2 units, it follows that there must be as many borrowers as depositors. With measure 1 of agents, the funding constraint faced by the bank in steady state is:

$$\int_A \bar{H}(da, \psi) = \frac{1}{2} \tag{11}$$

Although choosing an optimal $\{\rho, \varphi\}$ combination is a joint decision, for ease of explanation (and proof) it will be assumed that the bank uses the loan rate to equilibrate its balance sheet and then uses the monitoring rate, φ, to maximise profits.

PROPOSITION 4– *There is a unique value $\tilde{\rho}$ that satisfies the balance sheet constraint, equation* (11), *for given monitoring rate φ and exogenous ζ.*

This is a very intuitive proposition. From Proposition 3, there is a unique invariant distribution given φ for any ρ. The integral size of that unique distribution is continuous

[18] The level of equity funding is not relevant in this model but could be added without changing anything of substance.

in ρ so if it is too big - ie the bank faces an excess demand for loans - then raising the borrowing rate simultaneously reduces the demand for new loans (by increasing a_E), increases the incentive for existing borrowers to repay and exit production voluntarily (an increase in all the a_X simultaneously) and effectively tightens the loan covenants (by increasing a_T). These effects work on both sides of the balance sheet by reducing loans and increasing deposits. Uniqueness follows from continuity and monotonicity of the behavioural functions. So to satisfy the balance sheet constraint, ρ is a function of φ and we can denote the subset of invariant balance sheets, $\bar{H}$, that satisfy the balance sheet constraint as $\hat{H}(\rho(\varphi), \varphi)$. In other words, the invariant balance sheets satisfying the balance sheet constraint are now solely a function of the monitoring intensity.

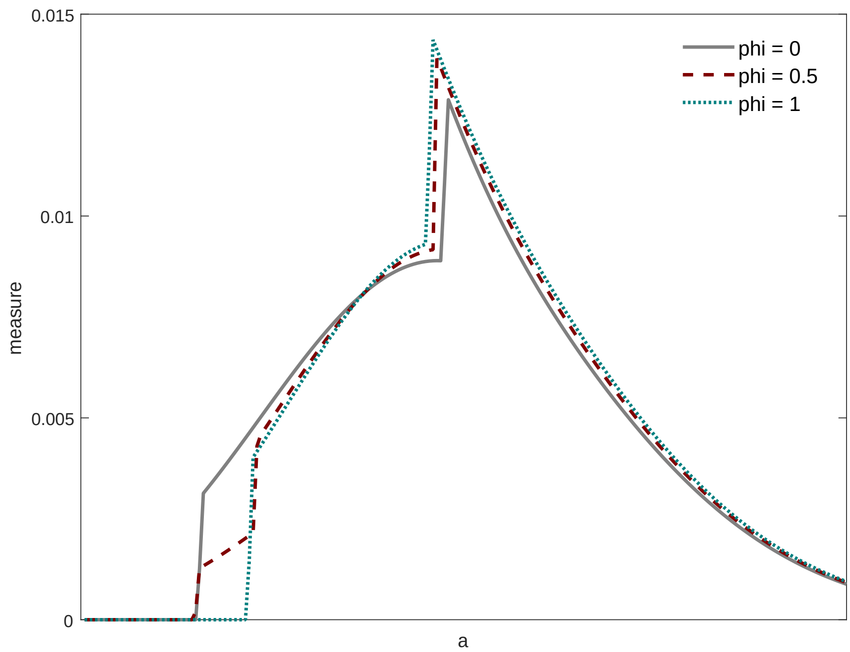


FIGURE 4. Equilibrium Effect of Higher or Lower Monitoring

Crucially, not all invariant distributions that satisfy the balance sheet constraint are equivalent from the point of view of the bank. Figure 4 illustrates what is at stake by plotting the distributions that satisfy the balance sheet constraint for the extreme cases of no monitoring ($\varphi = 0$) and complete monitoring ($\varphi = 1$) and an intermediate monitoring rate ($\varphi = 0.5$). To satisfy the balance sheet constraint, $\rho(0) > \rho(0.5) > \rho(1)$. The most stark difference between the distributions is in the left tail. When there is complete monitoring, firms are forced to exit when they fall below the covenant threshold. When there is no monitoring, firms are free to choose when they exit so the distribution is lower truncated at the voluntary exit threshold which is further to the left. When $\varphi = 0.5$, half the firms that would be shut-down by the bank but are above the voluntary exit threshold survive. The balance sheet constraint requires that the integral measure of firms across the three distributions is the same, so the higher survival rate of less profitable firms when monitoring is low has to be compensated by a smaller mass of firms with higher profitability. This occurs through a *lower* entry threshold as a result of the lower interest rate charged on loans with higher monitoring. Although the marginal entrants when the monitoring rate is high have *lower* profitability than when monitoring is low, they have higher profitability than the marginal exiting firm. This higher entry mass also increases the mass of incumbent firms at higher profitability levels.

Different balance sheet consistent invariant distributions, $\hat{H}(\varphi)$, generate different levels of steady state profits for the bank, represented in the following equation:

$$\begin{aligned}
\Pi(\varphi, \delta) \quad = \quad & \rho(\varphi) \int \int_{a_D(\rho(\varphi),\varphi)}^{1} F(da'|a)\hat{H}(da, (\rho(\varphi), \varphi)) \\
& - \int \int_{0}^{a_D(\rho(\varphi),\varphi)} (K - (q(a') - \rho(\varphi)))F(da'|a)\hat{H}(da, (\rho(\varphi), \varphi)) \\
& - \varphi m \int \int_{a_X(\rho(\varphi),\varphi)}^{1} F(da'|a)\hat{H}(da, (\rho(\varphi), \varphi)) \\
& - \delta \frac{1}{2}
\end{aligned} \tag{12}$$

The first line represents the loan interest payments $\rho(\varphi)$ the bank will receive from all firms who evolve into profitability states above the default threshold $a_D(\rho(\varphi), \varphi)$. The second line represents the loss given default bank arising from the assumption that the bank has to pay K to seize the assets of the firm and is the residual claimant. The third line represents the monitoring costs m on all non-defaulting firms in proportion to the monitoring rate φ. Finally, the bank will have to pay $\delta\frac{1}{2}$ in equilibrium as deposit interest. The dependence of profits on the endogenous deposit interest, δ, is noted explicitly.

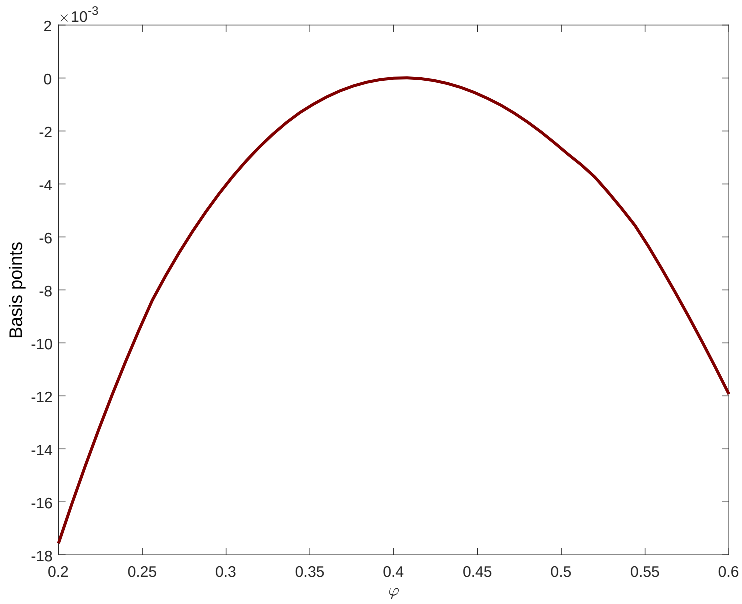
The objective of the bank is to

$$\max_{\varphi} \Pi(\delta) \tag{13}$$

Again the mathematical complexity obscures a relatively intuitive problem. The bank is trying to pick how frequently to monitor its loans. The balance sheet constraint ensures that having picked the monitoring rate, the loan interest rate is also determined. On the one hand, the higher the monitoring rate, the lower the default risk and as a result a lower loss given default but on the other hand the loan interest rate is lower and monitoring is costly. Π is a continuous function on a bounded support and therefore attains a maximum. Natural curvature in the model yields a unique profit-maximising solution. Note that there is no uncertainty faced by the bank.

Notwithstanding the relative simplicity of the problem, finding a solution is difficult and requires numerical methods. Why it is so difficult can be seen by observing how many times φ appears in equation (12).

FIGURE 5. Equilibrium Profits for Different Choices of Monitoring Intensity



In the final step of the solution, we require that $arg \max_{\varphi} \Pi(\delta) = 0$. $arg \max_{\varphi} \Pi(\delta)$ is a continuous and monotonic function of δ and profits fall if δ increases. There is a unique δ that satisfies the zero-profit condition.

Figure 5 illustrates the general case of an interior solution to the model. Figure 5 plots profits per loan in percentage points on the vertical axis against the monitoring intensity, φ. For interior solutions of the model, this is humped shaped with a unique profit maximising value of φ. Corner solutions are also possible with respect to the monitoring intensity under different parameterisations. If monitoring is very expensive, the bankruptcy cost is so high that the default rate is negligible or entry and exit costs are small (or any combination of the three), then $\varphi = 0$ can be optimal. Under the opposite conditions, full monitoring with $\varphi = 1$ can be optimal. Unique interior solutions occur over all intermediate ranges of these parameters.

What is the economic intuition behind the hump shape? With no monitoring, the equilibrium distribution of loans has the lowest credit quality, the highest default risk and the highest loss given default. But it is also the most valuable contract for the borrower since she has all the continuation right and therefore this contract commands the highest loan interest rate. At $\varphi = 0$, there is high risk and a high loan premium but no monitoring cost. As monitoring increases, credit quality improves, loan interest rates fall and monitoring costs increase. For interior solutions, the reduction in equilibrium credit losses initially more than compensates for the lower loan spread and higher monitoring expenses but eventually this is reversed. The trade-off reverses because the marginal improvement in credit risk is decreasing.

This reduction in the marginal effectiveness of monitoring occurs because of effects at opposite ends of the credit risk distribution. As the monitoring rate increases and the balance-sheet consistent loan interest rate decreases, the default threshold falls because it is the point of indifference between *net* profits and a fixed bankruptcy cost. Similarly, the bank's covenant threshold, which is also based on a net profit condition, is gradually falling. For interior solutions, there is initially a reinforcing benefit of monitoring as it leads to a smaller mass of at-risk firms that are further from the default threshold. But as more mass of the equilibrium distribution moves to the right, the relative importance of the firms that *currently* fail the covenant threshold as a proportion of all future defaulting firms diminishes. Monitoring has less marginal impact on firms that are increasingly already at little risk of default.

At the top end of the distribution, a higher monitoring rate induces a higher overall exit rate which must be matched in equilibrium by a higher entry rate. Since the marginal new entrants are of higher credit quality than the marginal exiting firm, this increases the average credit quality of the distribution, which is a benefit, but these new entrants are at low risk of default. As average distance to default declines, there is less marginal benefit to monitoring.

As alluded to in the introduction, the model can be extended to endogenise the covenant threshold, ζ, and allow state-contingent monitoring. These improve the efficiency of the monitoring process without changing the main qualitative feature of the solution - a unique profit maximising set of credit terms with zero profits in equilibrium. Full details can be found in Penalver (2016).

5. General equilibrium properties

The previous two sections have described the optimal choices of the agents and the decision by the bank. This section analyses the properties of equilibrium, explains the importance of several assumptions in deriving the results and hopefully provides economic intuition about how the various features of the model interact.

The first thing to note is that the qualitative features of the model are unaffected by the deposit interest rate. The equilibrium variables depend of the deposit rate because the loan contract terms have to change in response to changes in the deposit rate in order that the balance sheet condition is satisfied but the general properties of the model are unaffected.

The bank makes a profit (excluding monitoring costs and loss given default) in this model because there is a positive loan spread. This margin can be divided into three components. First, there is an intermediation rent received by the bank because of the assumptions that inventors cannot lend directly to entrepreneurs and that the project size is fixed.[19] This rent occurs because there is a utility gain in switching from being an inventor to being an entrepreneur because only the marginal entrant is indifferent between the two situations. In order to meet the balance sheet constraint, the bank has to change the relative attractiveness of being a depositor or borrower. The bank changes the loan rate (in conjunction with the other credit terms) and the deposit rate to equilibrate the balance sheet and earns the difference.

Second, borrowers pay for the option to declare bankruptcy. They do not do this explicitly but the fact that downside losses are capped increases the attractiveness of borrowing and must be constrained by a higher loan interest rate than if bankruptcy were not available. If the bank didn't have to pay to seize the assets of the firm, then this bankruptcy option would be equally valued by both parties and the bank would be fairly compensated for the transfer of losses during bankruptcy. If loss given default is fully compensated, then the bank has no incentive to reduce equilibrium default risk in which case, since monitoring is costly and paid for by the bank, it would be optimal to set the monitoring rate to zero. Turning this around, only if bankruptcy creates excess losses to the bank can there be an incentive to monitor. But these additional costs imply that the cost of the bankruptcy option to the bank is greater than the benefit to the borrower and therefore what the borrower is willing to pay will not compensate the bank for credit risk. This, of course, would seem to imply that risky lending loses money and that a private bank would not fund risky projects.

This explains the importance of the third component of the loan spread which is the insurance borrowers are willing to pay to avoid "premature" liquidation. To understand this premium, consider the situation in which entry and exit from production is costless ($L = C = 0$). From (10) if $C = L = 0$, then $a_E = a_X$ (because monitoring is irrelevant in this case). Firms would simply start up when there is an expected profit for at least one period and shut down again whenever there is an expected loss the following period. Production would be completely opportunistic. But in the model, as in reality, firms cannot jump in and out of production like this. Start-up costs are a deterrent to entry and liquidation costs provide an incentive to absorb mild losses if there is sufficient prospect of a future return to profitability. The voluntary exit threshold, $a_X(.)$, is the point at which the prospect of recovery becomes too low. Exiting above this point entails a utility loss. Yet, as explained in the previous section, if the bank is to protect itself against credit risk with a loan covenant, then this has to bite at above these voluntary exit thresholds and the greater the rate of monitoring, the less attractive the loan is to the borrower. Or put

[19] Standard assumptions like positive and continuously decreasing returns to capital, flexible loan size through time, free entry and exit and direct lending would together give all borrowers and lenders the same expected utility.

another way, the borrower is willing to pay an insurance premium against premature liquidation through a higher interest rate for a lower monitoring rate.

There is an important relationship between this premature liquidation insurance premium and the bankruptcy premium. The insurance premium, *per se*, does not involve any additional risk to the bank. It would exist even if bankruptcy imposed no additional cost on the bank and credit risk was fully compensated. But the more insurance the bank offers, through a lower monitoring rate, the lower the credit quality of the distribution (as was illustrated in Figure 4). So the insurance premium can only be earned by taking on default risk at the same time. It was explained above that if bankruptcy is inefficient, then the bankruptcy premium does not cover the credit risk. So it is the presence of the insurance premium that explains the willingness to offer risky loan contracts.

Joining up the threads of this discussion, monitoring and covenants are only relevant when bankruptcy is costly and when there are entry and exit frictions from production. It is the interplay between these two effects that determines equilibrium credit standards.

In the more general version of the model with an endogenous covenant threshold, the covenant is always set at a value below the entry threshold so the bank will knowingly allow firms to continue with profits below that of a potential entrant. The model, therefore, is consistent with voluntary forebearance on the part of banks although this is simply a different way of interpreting the insurance premium against premature liquidation.

6. THE SAVINGS GLUT

The key trade-offs can be observed by using the model to shed light on how the savings glut might have caused the low credit standards and narrow credit spreads observed in the years before it erupted.

Figure 6 presents a summary of credit conditions in Ireland in the years leading up to its banking crisis as an illustration of the issues in play. Annual loan growth to Irish residents, which was already strong in 2003, accelerated in 2004 and 2005 reaching a peak of over 30% at the start of 2006. This very strong loan growth coincided with a rapid inflow of foreign resident deposits into the Irish banking system. Over the same period, the 5-year interest rate on loans to non-financial corporations was steadily, albeit gently, declining. Yet despite the relaxation of a range of measures of credit standards presented in the right hand panel - certainly relative to what happened as funding conditions tightened in 2007 and 2008 - the default rate (on total loans) is fairly flat and if anything,

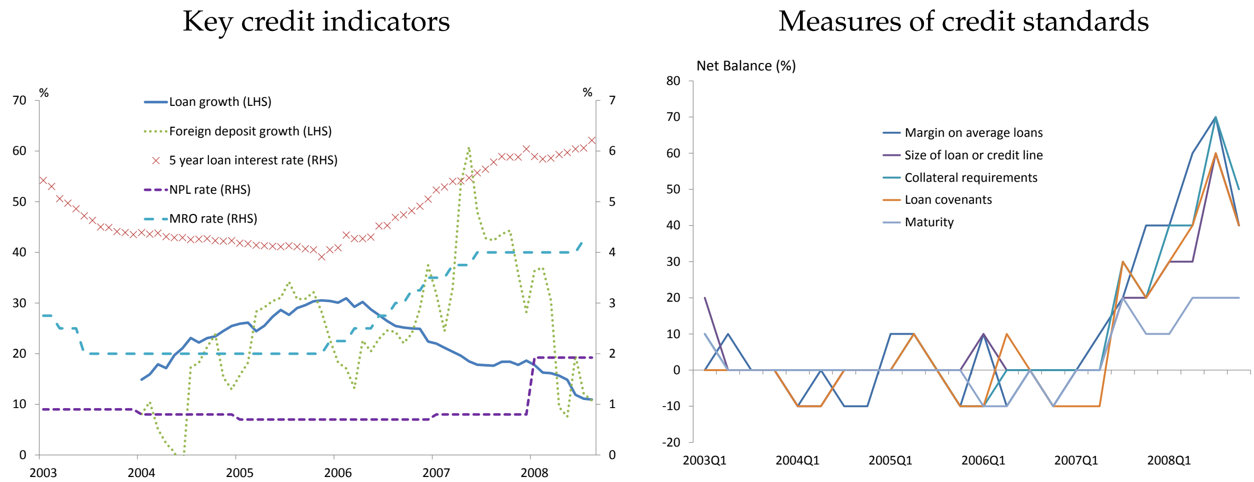


FIGURE 6. Irish credit conditions 2003-2008

slightly falling. Yet all commentators agree that the seeds of the subsequent banking crisis were sown here. In the meantime, the ECB kept its main refinancing operations rate at 2% from mid-2003 to the end of 2005.[20]

The savings glut hypothesis is introduced as an exogenous increase in the supply of deposits by 25% from outside the economy. This alters the balance sheet constraint in equation (11). We look for the endogenous equilibrium responses of the deposit rate, the loan interest rate and the monitoring intensity, all other things equal.

The results presented below are purely illustrative of the *qualitative* effects of the shock. (Although the model is quite complex, it is not anywhere near rich enough to try to provide a quantitative analysis of the lead up to the Irish banking crisis.) The parameter values were chosen to give plausible rather than targeted values of the endogenous variables that at the same time provide distinct results under the different shocks. Crucially, the qualitative results do not depend on the parameters chosen provided they deliver an interior solution. The details of the calibration of the model are in the appendix.

The first column of table I reports for the results for the main endogenous variables of the model for a baseline case. The bank randomly monitors 41% of its continuing loan portfolio on which it charges a loan interest rate of 6.9%. 2.4% of its loans default each year on which it suffers a loss of 2.5% of the loan portfolio inclusive of the cost of seizing the assets. The gross margin the bank earns over the expected losses on its portfolio is 4.6%.

The impact of the savings glut shock on the steady state is reported in the second column of table I. The monitoring rate falls from 41% to 31% and the loan interest rate

[20] In earlier versions of this paper, it was assumed that the central bank set the deposit interest rate δ - see Penalver (2014).

TABLE I. COMPARATIVE STEADY STATES

	Baseline	Savings Glut
δ	4.8%	4.4%
ρ	6.9%	6.7%
φ	0.41	0.33
Default rate	2.4%	2.7%
LGD	1.8%	2.0%

falls from 6.9% to 6.7%. The deposit rate falls from 4.8% to 4.4%. With a lower monitoring rate, the credit risk distribution worsens (as in Figure 4) and the default rate rises to 2.7%. Loss given default rises from 1.8% to 2.0%. By construction, the volume of credit is also increased. The savings glut shock thus replicates key features of the credit market prior to the crisis.

What is the intuition for these results? The increase in the supply of savings requires the bank to expand lending in order to meet the balance sheet condition. (The bank does not have access to an alternative investment opportunity in this model.) The bank needs to relax the loan terms or lower the deposit facility rate in order to induce an expansion of lending which in turn requires some combination of a lower entry threshold to entice new borrowers and lower exit thresholds to retain more incumbent borrowers. The lowering of any of these thresholds necessarily implies a less profitable and thus riskier distribution of borrowers.

The bank could satisfy the new balance sheet condition by reducing the deposit rate, δ, and leave the credit conditions unchanged but this widens the net interest margin even taking into account a deterioration in credit risk. So this option isn't compatible with the zero profit condition. One the other hand, keeping the deposit rate unchanged and only relaxing the loan terms cannot meet the zero profit condition either. For a given deposit rate, any combination of a lower loan interest rate or lower monitoring rate necessarily leads to a lower risk-adjusted loan rate and thus a reduction in net profit. Satisfying both the zero profit condition and the new balance sheet constraint therefore requires some combination of a lower deposit rate and weaker loan terms.

The bank could weaker its loan terms by decreasing the loan interest rate and leaving the monitoring rate unchanged. But instead the bank chooses to cut the monitoring rate and reduce the loan interest rate by less. In other words, the bank trades off some additional rise in the riskiness of its loan portfolio to protect its lending margin. This occurs because the return to monitoring is lower but the cost of monitoring is unchanged. The lower return to monitoring reflects two features:

- for any *given* distribution of firms, a lower loan interest rate reduces default risk.
- the lower the profit margin on each loan - due to the lower loan interest rate - the less the bank has to lose from default.

This relaxation in monitoring combined with the necessary deterioration in the profitability distribution of firms as the balance sheet of the bank expands leads to an increase in the steady state default risk. The higher default rate naturally leads to a higher loss given default.

It is very clear that a permanent savings glut shock leads to a worse distribution of credit risk in steady state with higher default. But Figure 6 shows that default risk initially went down rather than up.

Figure 7 illustrates that this feature, too, can be captured by the model. This figure presents a ″naive″ transition path for the default rate between the two equilibria. The economy begins in the baseline steady state. All incumbent firms retain their credit terms (loan interest rate and monitoring rate) and exit according to the baseline dynamics. New firms, however, enter with the credit terms of the savings glut equilibria and exit with the dynamics of that scenario. In this way the portfolio of loans gradually shifts from the baseline to the savings glut distribution and the balance sheet gradually expands.[21] Figure 7 shows that the default rate initially drops before gradually rising to the new equilibrium rate.

This occurs because even though the expected life-span of infra-marginal firms is shorter and eventually more of them will default, they still enter well above the default threshold. There is therefore a lag before these new firms start to default at a higher rate and in the interim there is a higher survival rate. This feature is ultimately due to the presence of entry and exit costs in the model. If these were switched off, as previously discussed, the distribution of incumbent firms is always lower truncated at the common entry and exit threshold. As a consequence, a lowering of this threshold to allow the balance sheet to expand immediately would result in an immediate rise in the risk of default.

7. Conclusion

The model developed in this paper provides a theory of how banks manage credit risk when they make multi-period loans to entrepreneurs who have private information on

[21] The dynamics are ″naive″ because incumbent entrepreneurs should use the expected pay-off under the savings glut terms to assess their exit decision. A fully rational expectations transition path is unnecessarily complicated for the simple point being made here.

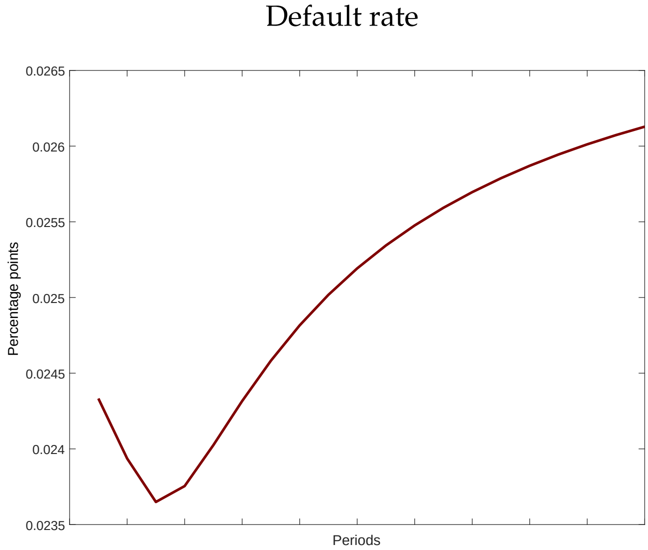


FIGURE 7. Default rate transition

firm profits and future prospects. If entry and exit were frictionless, then the individually rational choices of entrepreneurs and inventors would select the distribution of firms with the lowest credit risk and the asymmetry of information would have no bite. This is not the case with entry and exit costs since there will now be a segment of firms who choose to continue in circumstances in which they would not decide to start. Such firms are at the highest risk of defaulting in the near future and the bank has an interest in trying to reduce this portion of the distribution through its credit standards. The bank monitors continuing loans stochastically to discover breaches of loan covenants. But credit controls are costly, directly due to the cost of monitoring and indirectly through the interest rate the bank can charge on loans. Credit standards are a form of control right over the decision to continue a firm - the tighter the standards, the less control exercised by the borrower. Borrowers, therefore, are willing to pay an interest rate premium for greater control rights. So in deciding how to set its credit standards, a bank needs to take into consideration the cost of monitoring and enforcing its covenants, the effect of credit standards on default risk and the loan interest rate it can charge for different contract terms. The model shows how these competing considerations can be equilibrated whilst ensuring that the bank has sufficient deposits to fund its lending.

In the final section, the paper examined the effects of a deposit shock on equilibrium credit standards and credit spreads. An exogenous increase in deposits reduced credit standards and credit spreads, consistent with what was observed prior to the financial crisis. Macroprudential regulators should be particularly wary of lending practices when there are large capital inflows into the economy.

Acknowledgements *This model has benefitted from comments received at numerous seminars over recent years. I would like to specifically thank Vasco Carvalho, Gian Luca Clementi, Alexander Guembel, Yannick Kalantzis, Julien Matheron, Michael Reiter, Jean-Charles Rochet and Pierre-Olivier Weill. I am extremely grateful to John Stachurski for a substantial simplification and generalisation of the proof of Proposition 1. The views expressed herein are those of the authors and should under no circumstances be interpreted as reflecting those of the Banque de France. The research was begun whilst a PhD student at the Paris School of Economics.*

REFERENCES

ARELLANO, C., Y. BAI, AND J. ZHANG (2012): "Firm dynamics and financial development," *Journal of Monetary Economics*, 59(6), 533–549.

BAILY, M. N., E. J. BARTELSMAN, AND J. HALTIWANGER (2001): "Labor Productivity: Structural Change And Cyclical Dynamics," *The Review of Economics and Statistics*, 83(3), 420–433.

BERNANKE, B. (2005): "The global saving glut and the U.S. current account deficit," speech, remarks at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia.

——— (2010): "Monetary policy and the housing bubble," speech, remarks at the Annual Meeting of the American Economic Association, Atlanta, Georgia.

BONFIM, D., AND C. SOARES (2014): "The risk-taking channel of monetary policy exploring all avenues," Discussion paper.

BORIO, C., AND H. ZHU (2012): "Capital regulation, risk-taking and monetary policy: A missing link in the transmission mechanism?," *Journal of Financial Stability*, 8(4), 236–251.

BROECKER, T. (1990): "Credit-worthiness tests and interbank competition," *Econometrica*, 58(2), 429–452.

DELL'ARICCIA, G., R. MARQUEZ, AND L. LAEVEN (2010): "Monetary Policy, Leverage, and Bank Risk-Taking," IMF Working Papers 10/276, International Monetary Fund.

DIAMOND, D. W. (1984): "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51(3), 393–414.

DIXIT, A., AND R. PINDYCK (1994): *Investment under uncertainty*. Princeton University Press, Princeton, New Jersey.

DJANKOV, S., R. L. PORTA, F. LOPEZ-DE-SILANES, AND A. SHLEIFER (2002): "The Regulation Of Entry," *The Quarterly Journal of Economics*, 117(1), 1–37.

DOMS, M., AND E. J. BARTELSMAN (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

FARINAS, J. C., AND S. RUANO (2005): "Firm productivity, heterogeneity, sunk costs and market selection," *International Journal of Industrial Organization*, 23(7-8), 505–534.

FOSTER, L., J. HALTIWANGER, AND C. J. KRIZAN (2006): "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," *The Review of Economics and Statistics*, 88(4), 748–758.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

GALE, D., AND M. HELLWIG (1985): "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, 52(4), 647–63.

GERL, A., P. JAKUBIK, D. KOWALCZYK, S. ONGENA, AND J.-L. PEYDRO (2015): "Monetary Conditions and Banks Behaviour in the Czech Republic," *Open Economies Review*, 26(3), 407–445.

GERTLER, M. (1992): "Financial Capacity and Output Fluctuations in an Economy with Multi-period Financial Relationships," *Review of Economic Studies*, 59(3), 455–72.

HAUBRICH, J. G. (1989): "Financial intermediation : Delegated monitoring and long-term relationships," *Journal of Banking & Finance*, 13(1), 9–20.

HOLMSTROM, B., AND J. TIROLE (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *The Quarterly Journal of Economics*, 112(3), 663–91.

HOPENHAYN, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–50.

IOANNIDOU, V., S. ONGENA, AND J.-L. PEYDRO (2015): "Monetary Policy, Risk-Taking, and Pricing: Evidence from a Quasi-Natural Experiment," *Review of Finance*, 19(1), 95–144.

JIMENEZ, G., S. ONGENA, J. PEYDRO, AND J. SAURINA (2014): "Hazardous Times for Monetary Policy: What Do TwentyThree Million Bank Loans Say About the Effects of Monetary Policy on Credit RiskTaking?," *Econometrica*, 82(2), 463–505.

KING, M. (2010): "Monetary Policy Developments," speech, speech at the University of Exeter.

KOLMOGOROV, A., AND S. FOMIN (1970): *Introductory Real Analysis*. Dover Publications, New York.

MCCALL, J. J. (1970): "Economics of Information and Job Search," *The Quarterly Journal of Economics*, 84(1), 113–26.

MIAN, A., AND A. SUFI (2014): *House of Debt*. Chicago UNiversity Press, Chicago.

MINNIS, M., AND A. SUTHERLAND (2017): "Financial statements as monitoring mechanisms: evidence from small commercial loans," *Journal of Accounting Research*, 55(1), 197–233.

PALIGOROVA, T., AND J. SIERRA (2012): "Monetary Policy and the Risk-Taking Channel: Insights from the Lending Behaviour of Banks," *Bank of Canada Review*, 2012(Autumn), 23–30.

PENALVER, A. (2014): "Pre-crisis credit standards: monetary policy or the savings glut?," Working Paper 519, Banque de France.

——— (2016): "Optimal monitoring of long-term loan contracts," Discussion paper.

RAMEY, V. A., AND M. D. SHAPIRO (2001): "Displaced Capital: A Study of Aerospace Plant Closings," *Journal of Political Economy*, 109(5), 958–992.

RUCKES, M. (2004): "Bank competition and credit standards," *Review of Financial Studies*, 17(4), 1073–1102.

TAUCHEN, G. (1986): "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics Letters*, 20(2), 177–181.

WILLIAMSON, S. D. (1986): "Costly monitoring, financial intermediation, and equilibrium credit rationing," *Journal of Monetary Economics*, 18(2), 159–179.

Appendix A. Proofs of Propositions

Proposition 1 *Given Assumptions A and B and a banking contract ψ, unique, bounded and mutually consistent functions $V(E,a,\psi,V(I))$ and $V(I,a,\psi,V(E))$ exist.*

Proof. It greatly simplifies the presentation of the proof to ignore the presence of the default option and focus on the choices of voluntary entry, exit and continuation. (The default option simply puts a floor on the value function of the entrepreneur. By assumption the default option is never the alternative to continuation.) We work in $\mathcal{B}(A)$, the set of all bounded functions on the set A with the sup norm. For any $V(E) \in \mathcal{B}(A)$, the operator associated with $V(I)$

$$T_1\left(V(E),V(I)\right)(a) = \max\left\{\begin{array}{c} \delta + \beta E\left[V(I)\right], \\ \int_0^1 \left[q(a') - \rho + \beta V(E,a')\right] F(da'|a) - C \end{array}\right\}$$

satisfies Blackwell's sufficient conditions for a contraction on $\mathcal{B}(A)$ with modulus β and a unique $V(I)$ exists for a given $V(E)$ with

$$V(I) = T_1\left(V(E),V(I)\right) \tag{14}$$

Likewise, for any $V(I) \in \mathcal{B}(A)$, the operator

$$T_2\left(V(E),V(I)\right)(a) = \max\left\{\begin{array}{c} -L + \delta + \beta E\left[V(I)\right], \\ \int_0^1 \left\{\begin{array}{c} q(a') - \rho \\ +\beta\left\{(1-\varphi)\, V(E,a')\right\} \\ +\beta\left\{\varphi\left(1-\mathbf{1}(a)\right) V(E,a')\right\} \\ +\beta\left\{(1-\varphi\mathbf{1}(a))\left[E\left[V(I)\right]-L\right]\right\} \end{array}\right\} F(da',a) \end{array}\right\}$$

is a contraction with modulus β and therefore a unique $V(E)$ exists for a given $V(I)$ with

$$V(E) = T_2\left(V(E),V(I)\right) \tag{15}$$

($\mathbf{1}(a)$ is an indicator function taking 0 below a_T and 1 above.) The existence of each value function individually does not, however, imply the existence or uniqueness of any pair of functions $(V(E),V(I))$ satisfying both conditions (14) and (15) simultaneously. For any $V(I), V(I)' \in \mathcal{B}(A)$ and$V(E), V(E)' \in \mathcal{B}(A)$

$$\left|\int_A \left[V(I)(a',.) - V(I)'(a',.)\right] G(da')\right| \leq \left\|V(I) - V(I)'\right\|$$

$$\left|\int_A \left[V(E)(a\prime,.) - V(E)'(a\prime,.)\right] F(da\prime|a)\right| \leq \left\|V(E) - V(E)'\right\|$$

and

$$\left| \int_0^1 \left\{ \begin{array}{c} (1-\varphi)\left[V(E,a') - V'(E,a')\right] \\ +\varphi\left(1-\mathbf{1}(a)\right)\left[V(E,a') - V'(E,a')\right] \\ +(1-\varphi\mathbf{1}(a))\left[E\left[V(I)\right] - E\left[V'(I)\right]\right] \end{array} \right\} F(da'|a) \right| \leq \max \left(\begin{array}{c} \|V(I) - V(I)'\|, \\ \|V(E) - V(E)'\| \end{array} \right)$$

where $\|v\| = \sup_a |v(a)|$ is the usual sup norm. Let $\mathcal{M}$ be the set of ordered pairs $(V(E), V(I))$ such that $V(E)$ is in $\mathcal{B}(A)$ and $V(I)$ is in $\mathcal{B}(A)$ and impose the following metric d on $\mathcal{M}$:

$$d((V(E), V(I)), (V'(E), V'(I))) = \max\left\{\left\|V(E) - V(E)'\right\|, \left\|V(I) - V(I)'\right\|\right\}$$

Now consider the operator $T : \mathcal{M} \to \mathcal{M}$ defined by

$$T\left(V_E, V_I\right) = \left(T_1\left(V(E), V(I)\right), T_2\left(V(E), V(I)\right)\right)$$

Fix $a \in A$ and observe that

$$\begin{aligned}
\left| \begin{array}{c} T_1\left(V(E), V(I)\right)(a) \\ -T_1\left(V'(E), V'(I)\right)(a) \end{array} \right| &= \left| \begin{array}{c} \max\left\{ \begin{array}{c} \delta + \beta E\left[V(I)\right], \\ \int_0^1 \left[q(a') - \rho + \beta V(E,a')\right] F(da'|a) - C \end{array} \right\} \\ -\max\left\{ \begin{array}{c} \delta + \beta E\left[V'(I)\right], \\ \int_0^1 \left[q(a') - \rho + \beta V'(E,a')\right] F(da'|a) - C \end{array} \right\} \end{array} \right| \\
&\leq \max\left\{ \begin{array}{c} \left| \begin{array}{c} \int_0^1 \left[q(a') - \rho + \beta V(E,a')\right] F(da'|a) - C \\ -\left[\int_0^1 \left[q(a') - \rho + \beta V'(E,a')\right] F(da'|a) - C \right] \end{array} \right|, \\ \left|\delta + \beta E\left[V(I)\right] - \delta + \beta E\left[V'(I)\right]\right| \end{array} \right\} \\
&= \beta \max\left\{ \begin{array}{c} \left| \int_0^1 \left[V(E,a') - V'(E,a')\right] F(da'|a) \right|, \\ \left|E\left[V(I)\right] - E\left[V'(I)\right]\right| \end{array} \right\} \\
&\leq \beta \max\left\{\left\|V(E) - V'(E)\right\|, \left\|V(I) - V'(I)\right\|\right\}
\end{aligned}$$

where the first inequality follows from the fact that $|\max(a,b) - \max(c,d)| \leq \max(|a-c|, |b-d|)$ for any $a, b, c, d \in \mathbb{R}$.Taking the supremum over both sides:

$$\left\|T_1\left(V(E), V(I)\right)(a) - T_1\left(V'(E), V'(I)\right)(a)\right\| \leq \beta \max\left\{\left\|V(E) - V'(E)\right\|, \left\|V(I) - V'(I)\right\|\right\}$$

Similar arguments give

$$\left\|T_2\left(V(E), V(I)\right)(a,\alpha) - T_2\left(V'(E), V'(I)\right)(a,\alpha)\right\| \leq \beta \max\left\{\left\|V(E) - V'(E)\right\|, \left\|V(I) - V'(I)\right\|\right\}$$

and thus

$$d(T\left(V(E), V(I)\right), T\left(V'(E), V'(I)\right)) \leq \beta d\left(\left(V(E), V(I)\right), \left(V'(E), V'(I)\right)\right)$$

Hence T is a contraction mapping on the complete metric space $(\mathcal{M}, d)$ establishing a unique fixed point exists. $V(E, a, \alpha)$ and $V(I, a)$ are unique continuous functions. $\square$

Proposition 2 *These value functions yield unique and continuous functions in ψ for the entry threshold a_E and exit thresholds $a_X(\alpha)$. $a_E(\psi)$ and $a_X(\psi, \alpha)$ are both strictly increasing in ψ.*

Proof. To show that the threshold functions are strictly increasing in ψ, consider the effect of a reduction in ρ on $a_X(\psi)$. To simplify the exposition, ignore the bank threshold and the default option, and set $\delta = C = L = 0$. These simplifications give the following value function for the agents conditional on a given ρ

This leaves us with

$$V(a) = \max\left\{\int \left(q(a') - \rho + \beta V(a')\right) F(da', a), \beta EVG\right\}$$

where $EVG = \int V_I(a)G(da)$ and there is a unique switching threshold, $\hat{a}$, defined implicitly by:

$$V(\hat{a}) = \int \left(a' - \rho + \beta V(a')\right) F(da', \hat{a}) = \beta EVG$$

Iterating on the value function, for any $a \geq \hat{a}$,

$$V(a) = \Lambda(a) - \frac{1}{1-\beta}\Upsilon(a)\rho + \beta\left\{1 - \Upsilon(a)\right\} EVG$$

where $\Upsilon(a) = \frac{1}{1-\beta}\left[1 - \beta p\,(a)\,(1) - \beta^2 p\,(a)\,(2)....\right]$ and $p\,(a)\,(i)$ the probability of exiting in i period's time given starting state a. Iterating from the entry perspective gives

$$EVG = \frac{\int_{\hat{a}}^{1} \Lambda(a)G(da) - \frac{1}{1-\beta}\rho \int_{\hat{a}}^{1} \Upsilon(a)G(da)}{(1 - \beta G(\hat{a})) - \beta \int_{\hat{a}}^{1}(1 - \Upsilon(a))G(da)}$$

Fix $\hat{a}$ and take the derivative with respect to ρ of both sides of equation X. If $V(\hat{a})$ falls by more than βEVG, then $\hat{a}$ must rise to find a new threshold.

$$\frac{\partial V(\hat{a})}{\partial \rho} \mid \hat{a} = -\frac{1}{1-\beta}\Upsilon(a) + \left\{1 - \Upsilon(a)\right\}\beta\left[\frac{\partial EVG}{\partial \rho} \mid \hat{a}\right]$$

Thus

$$\frac{\partial V(\hat{a})}{\partial \rho} \mid \hat{a} < \beta\left[\frac{\partial EVG}{\partial \rho} \mid \hat{a}\right]$$

if and only if

$$\frac{\partial \beta EVG}{\partial \rho} \mid \hat{a} > -\frac{1}{1-\beta}$$

or

$$\frac{-\frac{1}{1-\beta}\beta \int_{\hat{a}}^{1} \Upsilon(a)G(da)}{(1 - \beta G(\hat{a})) - \beta \int_{\hat{a}}^{1}(1 - \Upsilon(a))G(da)} > -\frac{1}{1-\beta}$$

which holds provided $\beta < 1$ which is true.

Therefore the entry option is more sensitive to the rise in the interest rate than the waiting option for a fixed $\hat{a}(\rho)$. As a result

$$\int \left(a' - \tilde{\rho} + \beta V(a', \tilde{\rho})\right) F(da'|\hat{a}(\rho)) < \beta EV(\tilde{\rho})$$

for $\tilde{\rho} > \rho$ and $\hat{a}(\rho)$. This requires $\hat{a}(\tilde{\rho}) > \hat{a}(\rho)$ to equilibrate the rise in ρ. This condition carries over to the more complex model with entry and exit costs, bankruptcy and the bank covenant with monitoring. □

Proposition 3 *For each ψ and given N there is a unique invariant distribution, $\bar{H}(\psi)$*

Proof. The proof mimics Hopenhayn (1992). The transition equation (16) for the cross-sectional distribution can be written as

$$H'(\psi) = P(\psi)H(\psi) + Ng(\psi) \tag{16}$$

where $P(\psi)$ is a bounded linear operator on the space of positive bounded measures defined by $PH(\Gamma, \psi) = \int P(a, \Gamma, \psi)H(da)$ for all borel sets Γ in A. $P(\psi)$ is defined in the main text. The monotone mixing assumption, A(iii), plus the presence of a positive exit rate ensure that $\|\hat{P}^n(\psi)\| < 1$ where $\hat{P}^n(\psi)$ is the n-fold composition of $\hat{P}(\psi)$ and $\|.\|$ is the operator norm. As a result, following Hopenhayn (1992) drawing on Kolmogorov and Fomin (1970) $(I - \hat{P}^n(\psi))^{-1} = \Sigma_{t=0}^{\infty}\hat{P}^{nt}(\psi)$ and since $\|\hat{P}^n(\psi)\| < 1$ is non-increasing in n, $(I - \hat{P}(\psi))^{-1}$ also exists and

$$\bar{H}(\psi) = N(I - \hat{P}(\psi))^{-1}g(\psi)$$

is the unique invariant distribution given ψ and N. □

Proposition 4 *There is a unique value $\tilde{\rho}$ that ensures that the balance sheet of the bank is equal on both sides for given values of φ and ζ.*

Proof. For any φ there is a unique invariant distribution $\bar{H}(\psi)$ for any N from Proposition 3. However, in equilibrium we require

$$\int_A \bar{H}(da; \psi) = \frac{1}{2} = N \tag{17}$$

which will not be the case for arbitrary ψ. For any $\rho' > \rho$

$$X(\Gamma, \rho, \rho') \equiv \hat{P}(a, \Gamma; \rho) - \hat{P}(a, \Gamma; \rho') = \begin{array}{ll} 0 & \text{for } \Gamma \cap [a_T(\rho'), 1] \\ \varphi \int_{\Gamma \cap [a_T(\rho), a_T(\rho')]} F(dq \mid a) & \text{for } \Gamma \cap [a_T(\rho), a_T(\rho')] \\ 0 & \text{for } \Gamma \cap [a_X(\rho'), a_T(\rho)] \\ (1-\varphi) \int_{\Gamma \cap [a_X(\rho), a_X(\rho')]} F(dq \mid a) & \text{for } \Gamma \cap [a_X(\rho), a_X(\rho')] \\ 0 & \text{otherwise} \end{array}$$

$X(\rho,\rho')$ is also a bounded linear operator. Thus

$$\begin{aligned}
\Sigma_{t=0}^{\infty}\hat{P}^t(\rho) &= \left[\hat{P}(\rho')+X(\rho,\rho')\right]^0+\left[\hat{P}(\rho')+X(\rho,\rho')\right]^1+\left[\hat{P}(\rho')+X(\rho,\rho')\right]^2... \\
&= \hat{P}^0(\rho')+\hat{P}^1(\rho')+\hat{P}^2(\rho')...+X^0\left(\rho,\rho'\right)+X^1\left(\rho,\rho'\right)+\hat{P}^1(\rho')X^1(\rho,\rho')+X^2(\rho,\rho')... \\
&= \hat{P}^0(\rho')+\hat{P}^1(\rho')+\hat{P}^2(\rho')...+XP\left(\rho,\rho'\right) \\
\left[I-\hat{P}(\rho)\right]^{-1} &= \left[I-\hat{P}(\rho')\right]^{-1}+XP\left(\rho,\rho'\right)
\end{aligned}$$

Define $\mu\left(\Gamma,\rho,\rho'\right)=g(\Gamma,\rho)-g(\Gamma,\rho')=G(\Gamma\cap[a_E(\rho),a_E(\rho'))$ which is a positively bounded measure. The invariant distribution for ρ and $N=\frac{1}{2}$ is

$$\begin{aligned}
\bar{H}(\rho) &= \frac{1}{2}(I-\hat{P}(\rho))^{-1}g(\rho) \\
&= \frac{1}{2}\left[\left[I-\hat{P}(\rho')\right]^{-1}+XP\left(\rho,\rho'\right)\right]\left[g(\rho')+\mu\left(\rho,\rho'\right)\right] \\
&= \frac{1}{2}\left[I-\hat{P}(\rho')\right]^{-1}g(\rho')+\frac{1}{2}\left[I-\hat{P}(\rho')\right]^{-1}\mu\left(\rho,\rho'\right)+\frac{1}{2}XP\left(\rho,\rho'\right)\left[g(\rho')+\mu\left(\rho,\rho'\right)\right] \\
&= \bar{H}(\rho')+\frac{1}{2}\left[I-\hat{P}(\rho')\right]^{-1}\mu\left(\rho,\rho'\right)+\frac{1}{2}XP\left(\rho,\rho'\right)\left[g(\rho')+\mu\left(\rho,\rho'\right)\right]
\end{aligned}$$

The second and third terms have positive measure and thus $\int_A \bar{H}(da,\rho') > \int_A \bar{H}(da,\rho)$. These terms are all continuous and monotonic in ρ' so the measure of incumbent borrowers can be continuously and monotonically scaled up or down with ρ. There is thus a unique $\tilde{\rho}$ at which $\int_A \bar{H}(da,\tilde{\rho})=\frac{1}{2}=N$. □

APPENDIX B. ALTERNATIVE SPECIFICATIONS

TABLE II. BASELINE

	Baseline	Contestable	Welfare
φ	0.57	0.48	1.0
ρ	0.071	0.069	NR
Default rate	0.030	0.031	0.024
LGD	0.025	0.026	0.020
$\rho-$ LGD	0.046	0.043	NR

APPENDIX C. MODEL PARAMETERISATION

This is a stylised model that is intended to demonstrate a mechanism rather than replicate empirical data. The parameters of the illustrative numerical example are a mixture of calibration from direct model counterparts and trial and error for the rest to deliver plausible values for the loan interest rate, entry and exit rate of firms, default rate and loss given default. Yet at the same time, the model needs relatively modest "curvature" so that the mechanism is clearly visible in the illustrations. The example uses a normal distribution for $G(a)$ and an AR(1) process for $F(a, a')$ using a Tauchen matrix approximation (see Tauchen (1986)). The values of the idiosyncratic shock process comes from Foster, Haltiwanger, and Syverson (2008) (FHS).

TABLE III. PARAMETER VALUES

Parameter	Description	Value
β	discount rate	0.975
δ	deposit rate	3.5%
ρ	idiosyncratic shock persistence	0.9
σ	idiosyncratic shock st dev	0.3
C	entry cost	0.3
L	exit cost	0.3
B	bankruptcy cost	0.45
m	monitoring cost	0.006
K	bank excess loss	0.6