Pre-crisis credit standards: monetary policy or the savings glut?

Adrian Penalver[*]

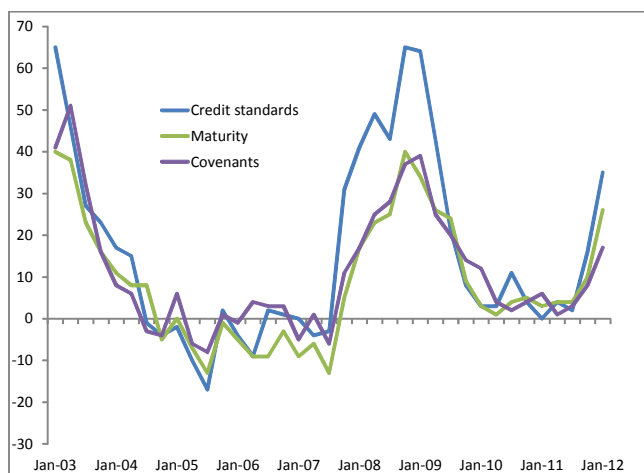14 February 2013

Abstract

   This paper presents a theoretical model of bank credit standards. It examines how a monopoly 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. The model has a recursive structure and contains heterogeneous agents who can self-select to be depositors or borrowers at any point in time. The bank loan contract considered specifies the interest rate, the monitoring intensity and a profitability covenant. Within this class of contract, he bank chooses the terms which maximise Markov stationary profits subject to the constraint that it must have as many deposits as loans. As an illustrative example, the model is used to consider whether the reduction in credit standards and credit spreads observed before the financial crisis could have been caused by low official interest rates or a positive deposit shock. The model rejects a risk-taking channel of monetary policy and endorses the savings glut hypothesis.

1. Introduction

One of the most pernicious aspects of credit booms is that they are often associated with a reduction in bank lending standards. During credit booms less

---

Figure 1: ECB Bank Lending Survey Credit Standards

creditworthy firms and households obtain loans on easier terms than they would at other points in the credit cycle. This would be of little concern - and indeed might be socially optimal - if loans were made for a single period because the contract terms can adjust in line with changing economic conditions. But loans are generally of some duration and contract terms fixed at origination with the result that changes in credit standards (taken here to be the contract terms on offer) can have persistent effects and distort the allocation of resources over time. Likewise, there is persistence in the average credit quality of the banking sector's loan portfolio. As a consequence, a sustained period of lax credit standards increases the latent credit risk in the banking system and thereby the risk of economic and financial instability[1]. Using data from the ECB Bank Lending Survey, figure 1 illustrates that the financial crisis which began in 2007/8 was preceded by several years of very low credit standards (comprising weak loan covenants and longer duration lending). A very similar picture emerges from US data. If these financial stability risks are realised, the empirical evidence shows conclusively that recessions following banking crises are deeper and longer than "normal" recessions (Jorda, Schularick and Taylor (2011), IMF (2008)). So understanding how banks

---

[1]Moore (1956) argues that lax credit standards in the 1920s contribute to the severity of the Great Depression citing the fact that loans originating in the second half of that decade defaulted at twice the rate of those originating in the first half.

set their credit standards and what macroeconomic factors might cause them to change is crucial if macroprudential policies are to be effective in counteracting these economic and financial stability risks. A particularly important subsiduary question is whether the stance of monetary policy influences credit standards - a component of the "risking-taking" channel of monetary policy (Borio and Zhu (2008)). If there is a link, then there is a potential conflict between the objectives of monetary stability and financial stability and thus an operational tension between agencies or groups responsible for individual objectives.

Popular commentary on the lead up to the recent financial crisis focuses on sociological and psychological phenomena, blaming the herd behaviour of myopic and bonus-hungry bankers who are driven towards ever more reckless lending (see Kindleburger (1978) for a catalogue of historical examples). There is likely more than a grain of truth in this description. But it is also likely to be an incomplete explanation as banks, like everyone else, respond to the economic incentives they face (see also Stein (2013)). Such economic incentives are also those most directly influenced by public policy.

The existing theoretical literature on bank credit standards mostly focuses on the role of competition between banks in the loan market.[2] Credit standards are a strategy choice in the asymmetric information games analysed by Ruckes (2004), Dell'Arricia and Marquez (2006) and Gorton and He (2008). In symmetric equilbrium, banks co-ordinate on high or low (usually no) credit standards depending on the state of the economy and beliefs about the strategies of the other banks.[3] An intuitive common element on these models is that a bank will not want to set low credit standards if it believes others will set high standards because of a form of "winner's curse". A bank with low credit standards is more likely to end up with borrowers that a high standard bank knows are bad. But if credit risk is low because times are good, then competing banks can co-ordinate on low standards. From a different perspective, Rajan (1994) describes a model of forebearance in which bank managers are more likely to roll over loans to failing firms to avoid losses and thus preserve their reputation for competence if they believe that other banks will report high earnings.

Post-crisis papers tend, unsurprisingly, to focus on agency problems. In the model of Acharya and Naqvi (2011), bank managers are rewarded for the volume

---

[2] An exception is Weinberg (1995).

[3] Anecdotal support for the competition channel is the now infamous quote from Chuck Prince "as long as the music is playing, you've got to get up and dance".

of loans they originate in order to induce effort but to control risk-taking, managers can be audited (if the bank suffers a liquidty crisis) and punished if lending is found to have been on excessively generous terms. This contract cannot completely align incentives because there is a cap on the punishment bank owners can impose. The key result is that the incentive for managers to take risks to increase lending volume rises when banks are flush with liquidity because this reduces the likelihood, *ceteris paribus*, that the manager will be audited. In Dell'Ariccia, Laeven and Marquez (2010), limited liability makes bank equity an option on loan performance. The effect of monetary policy on monitoring depends on the sensitivity of loan rates, the extent to which depositors require a risk-premium for weakly capitalised banks and the degree of leverage. When banks have the flexibility to change their capital structure, a cut in monetary policy leads to greater leverage and less monitoring because the agency problem is weakened.

The model in this paper differs from this literature in a number of important respects. Firstly, it ignores competition by considering a monopoly bank. This suggests that interbank competition is not a necessary condition for the existence of credit standards in equilibrium (although one could believe that competition amplifies the effects demonstrated here). Secondly, the model has a general equilibrium constraint because the bank must attract sufficient deposits to fund its lending. The non-bank agents in this economy are heterogeneous at any point in time and self-select to be depositors or borrowers depending on their idiosyncratic state and the parameters of the loan and the deposit contract offered by the bank.[4] As a consequence, the behaviour of non-bank agents is at least as important in explaining credit standards in equilibrium as the behaviour of the bank. Thirdly, the set-up is recursive in contrast to the two-period models common in the wider banking literature. This recursive structure makes it possible to analyse credit standards when loans are of long duration but deposits are at call. Finally, this is not an optimal contract model. The profit-maximising contract which determines equilibrium is chosen with respect to the parameters of an assumed contractural form rather than over the form of the contract itself.

This model also differs from more recent papers describing the risk-taking channel of monetary policy because banks are not the underlying source of friction but exist as an imperfect solution to frictions originating elsewhere. For example banks are not fundamentally less risk averse than society as in Agur and Demertzis (2010) or value projects in a different way from other investors as in Adrian and Shin (2008). Indeed, banks are socially useful since they supply

---

[4]The model draws heavily on the heterogeneous agent framework of Hopenhayn (1992).

maturity transformation and credit risk management. Credit risk arises out of the behaviour of borrowers and exists in equilibrium because it is too costly for banks to eliminate it. The only distortion caused by the bank is the assumption of monopoly power which will result in the bank choosing a socially sub-optimal level of credit standards.

The paper is organised as follows. Section 2 sets up the model by describing the state space, the heterogeneous agents and the banking contract. Section 3 explains the equilibrium behaviour of the agents for a given loan contract. To assist in understanding the subsequent choice of the bank, an illustrative numerical simulation is presented. Section 4 explains equilibrium bank behaviour and the profit-maximising choice of credit standards. Section 5 describes the general equilibrium properties of the model. Section 6 conducts some comparative static experiments - a monetary policy easing and an exogenous increase in bank deposits - to consider whether the global savings glut or the risk-taking channel of monetary policy are possible culprits for explaining the loosening of credit standards in the lead up to the crisis. The results show that in this model only the savings glut hypothesis is consistent with lower credit standards and falling credit spreads. Section 7 concludes.

2. Model

2.1. Space

The model takes place in an economy of infinitely small, *ex ante* identical and infinitely living risk-neutral agents of measure 1. Time is discrete and in each period every agent can choose whether or not to run a project subject to stochastic returns. Existing or potential projects are individual specific and are indexed by an idiosyncratic "profitability" state, a, drawn stochastically every period from the compact set $\{a \in A \in \mathbb{R} : 0 \leq a \leq 1\}$. This profitability state defines the individual circumstances of each agent (and thus *ex post* heterogeneity) and can be thought of as encompassing technical productivity, consumer preferences, degree of market power and managerial talent. Profitability levels are private information and costly for outsiders to verify.

2.2. Agents and production

At any point in time the agents in this economy are partitioned endogenously into two situations.

- One group of agents are waiting to enter production. Each period these "inventors" receive an "idea" for a project with a profitability state drawn from the continuous i.i.d distribution $G(a)$. Inventors chose between paying a start-up cost S and entering production next period based on this profitability state or waiting for another idea the following period. Inventors are of measure W which in equilibrium will equal $\frac{1}{2}$ and while they wait, they deposit their capital at the bank on which they receive an exogenous deposit rate r_d.

- The other group of agents, of measure $1 - W$ are currently in production for which they will have borrowed funds at a loan rate r. These agents are called "entrepreneurs" and their per period payoffs are a function of their current idiosyncratic profitability states. Individual profitability states evolve according to a first-order Markov process $F(a, a')$ and per period payoffs to entrepreneurs are denoted $\pi_t(a_t, r)$. The following assumptions on the transition process and profit function are made:

  A (i) $F(a, a')$ is continuous in a and a'; (ii) $F(a, a')$ is strictly decreasing in a. This gives persistence in shocks to a. (iii) $F^n(a, \epsilon) > 0 \; \forall \epsilon$ for some n where $F^n(a, .)$ is the conditional probability distribution of profitability level a_{t+n} given $a_t = a$.[5] So from any given level of profitability, it is possible to transit to any other profitability level in a finite number of periods. Since there are exit thresholds, this assumption implies that all projects will almost surely close at some future point.

  B (i) π_t is continuous in all arguments; (ii) π_t is bounded above and below (based on the boundedness of A); (iii) $\pi_t(0, r) \ll 0 \ll \pi_t(1, r) \; \forall r > 0$; (iv) $\pi'_a(a, r) > 0$, $\pi''_a(a, r) < 0$; and (v) $\pi'_r(a, r) = -1 < 0$.[6]

In each period, entrepreneurs decide whether to continue in production next period or to exit and switch to being an inventor. If they do not wish to continue, there are two exit options. "Orderly" exit occurs if the entrepreneur absorbs current period losses, $\pi(a, r)$ - which includes repayment of loan interest - 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

---

[5] The notation will follow the convention that periods and states are denoted with subscripts and iterations or powers by superscripts.

[6] Assumption (v) arises from a fixed loan size to be explained below.

or non-pecuniary such as lost human capital and reputation. (For notational convenience, it is assumed that L is paid in the following period.) "Default" occurs if the entrepreneur files for bankruptcy protection in which case current period losses are excused (including repayment of loan interest) but the agent pays an exogenous bankruptcy cost B (for convenience also assumed to be paid in the following period). (Default is costly to the bank and will be explained in more detail below.) 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_δ (exit and default, respectively) and the value of B is calibrated for the sensible ordering $a_X > a_\delta$. Although a is continuous, the model is in discrete time, so the profitability index can jump to $a < a_\delta$ without passing through the interval $[a_\delta, a_X]$. In other words, projects can jump straight to default by passing over the orderly exit region. By assumption A(iii), project profitability will almost surely pass below one of these thresholds and the project is terminated on the first occasion.

To keep the model as simple as possible, all decisions take place at the extensive margin. Aggregate variables, which are used for market clearing, are thus the sums over measures of agents entering, exiting or continuing. To turn off the intensive margin various additional assumptions are required.

- It will be assumed that all agents have an inalienable, unaugmentable and indestructible unit of capital. There is thus no capacity to use *capital* to smoothe consumption through time or escape from the requirement to borrow for production. (The timing of consumption is irrelevant in the model so it could be assumed that agents can store consumption goods.) The only decision that agents are making is how to allocate their unit of capital between bank deposits and production.

- As a simplifying normalisation, 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. (This implied leverage ratio could be set to any value with an adjustment to Assumption B (v) and the market clearing condition.)

The purpose of these assumptions is to ensure that borrowing and lending take place and that the loan size remains fixed over the lifetime of the project. Since entrepreneurs make profits during the lifetime of the project, it might be

natural to think that they would want to pay down some of this debt so it will be convenient to assume that there is some tax advantage in retaining the loan. In principle, there is nothing to stop projects being continuously scalable along other dimensions (such as labour or other inputs) subject to the upper and lower bound on profits and thereby allowing the project to expand or contract over time along these dimensions depending on the state of profitability. It is also possible to consider that agents are organisms created from multiple atomic agents and thus capable of running more than one project or being a borrower and lender simultaneously. None of these ideas is pursued any further.

It will be assumed that all agents have an exogenous endowment of income every period regardless of their circumstances sufficient to cover any expenses or losses. This exogenous endowment plays no role in the model beyond giving borrowers financial flexibility to continue with the project if they wish. So all decisions in the model are based on incentives not inability to pay.

The agents are risk neutral and value expected future flows at a discount rate $\beta < 1$ per period. As a result, all choices are based on the present discounted expected sum of future flows. The timing of consumption is not relevant and no attempt is made to reconcile aggregate consumption and production. General equilibrium is restricted to equilibrating both sides of the balance sheet of the bank.

2.3. Banking

Capital in this model is intermediated between inventors and entrepreneurs by a monopoly bank. As a consequence of the assumption of infinitely small agents and the law of large numbers, the *ex ante* probability distributions over its loan portfolio are identical to the distribution of *ex post* outcomes. The bank offers the following deposit and loan contracts.

- Deposits earn an interest rate r_d set exogenously by the monetary authority. Deposits can be withdrawn at the end of any period.

- The loan contract specifies an interest rate r, 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 $\psi = \{r, \varphi, \xi\}$. The loan is provided for the duration of the project but both parties have an option to terminate it each period. The borrower has the option to repay the loan if they decide 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, a_T, at which the bank exercises its right to terminate the loan.

Comments on several aspects of this contract are in order.

Recursive models in economics almost always assume that loan contracts are a period-by-period proposition. This possibility is nested within the model presented here. The borrower has the option to repay in any period so this is trivially equivalent to a single-period loan. On the lender's side, if the loan is monitored every period ($\varphi = 1$) and the covenant condition is set equal to the entry threshold (denoted a_E), then the bank only rolls over loans to entrepreneurs to whom it would be willing to offer a new loan. Thus there exists a special case in which the bank, too, acts as if loans were for a single period. φ (or ξ) is an endogenous variable so if the bank chooses $\varphi < 1$ it will be doing so voluntarily and to maximise profits. The model will be parameterised so that an interior solution for φ is chosen which will imply that the expected duration of a loan, $\frac{1}{\varphi}$, is greater than 1.

A second issue is why does the bank need a termination option at all? If the interests of the bank and the borrower were always completely aligned, then the bank could rely on the choices of the borrower to protect itself and therefore would not need an option to act. Here their interests are partially misaligned because downside losses for the borrower are capped by the option to declare bankruptcy. In the very worst outcomes of a, the borrower can transfer losses to the bank. Thus there is some moral hazard when the borrower is choosing whether to continue in production or not.[7] Since it is the voluntary actions of the borrower that increase default probabilities, it is in the interests of the bank to monitor and to exercise its right to recall the loan.

In order to exercise the option, the bank has to know the state of the project which is private information to the borrower. If monitoring were free, the bank would always know the profitability state and would exercise its option to demand repayment of the loan whenever a fell below the covenant threshold, a_T. In this case projects would always be terminated by the decision of the bank (thereby making the option for the borrower to repay redundant). But if monitoring is

---

[7]Their interests are not totally antithetical either because there is a region $[a_\delta, a_X]$ in which the borrower bears increasing losses.

costly at an increasing rate and/or there are diminishing returns to monitoring activity, then monitoring will be incomplete. Obviously if a loan is not monitored, then the state is unknown to the bank and the loan cannot be recalled. If there is less than complete monitoring, then in steady-state some proportion of projects will be terminated by the bank and the rest terminated by the borrower repaying the loan or defaulting.

If the bank feels it needs to monitor to protect its interests because default risk is non-negligible, then it follows that the covenant profitability threshold must apply at $a_T > a_X$ otherwise it is redundant. This condition will also be assumed to be the case.

Finally, to keep the analysis tractable it is assumed that r and φ are not state-dependent. In theory and in practice we would not assume this is the case, particularly in the presence of credit risk. One might naturally assume that r and φ are decreasing in a. The bank would likely also increase the probability of monitoring a loan the longer it had been since the previous inspection. This would, of course, be a much more complex programming problem than considered here. But as hopefully will become clear, the main insight of the model and its driving mechanism would remain intact, albeit in weaker form, if it were possible to solve such an extended model. Why is this the case? Provided there is less than complete monitoring in all states, $\varphi < 1$, then not every contract could be updated every period. Therefore at least some contracts which were mutually satisfactory for a state a in one period would still be in place when the project is in state a' next period and this information is known to the borrower but not the bank. These are the circumstances in which the borrower can make decisions which are not completely in alignment with the interests of the bank. Since, as will be seen, borrowers are willing to pay a higher interest rate for contracts with $\varphi < 1$, then it may be profit-maximising for banks to offer such contracts. Intuitively, the friction in the model is that monitoring and contracts cannot keep pace with the evolution of states which are private information. Since in reality, states can change rapidly but bank monitoring and contracts do not, the mechanism working in this model would be present even if contracts, when they are updated, were state contingent.

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 situation (inventor or entrepreneur) and then draw their idiosyncratic shocks. The inventors get an idea from $G(a)$ and entrepreneurs get an update of their profitability according to $F(a, a')$.

2. Profits are realised and entrepreneurs with $a < a_\delta$ default. Other entrepreneurs pay loan interest and inventors receive deposit interest.

3. Inventors decide whether to enter production next period based on their initial profitability draw or to wait for a better idea in the future. Entrepreneurs decide whether to continue with production next period or to exit voluntarily and become an inventor. Entrepreneurs intending to pay back their loans inform the bank of their decision.

4. The bank monitors ongoing loans at the stochastic rate φ and recalls the loans of all entrepreneurs found below the covenant profitability threshold state a_T.

5. The bank receives deposits from ongoing and new inventors and makes additional loans to entering producers at a loan rate r.

The following two sections formalise the analysis of the choices of the productive and inventors. In this formal description a will always refer to profitability states for entrepreneurs and inventors in the *current* period and $a' \in A$ will always refer to profitability states in the *following* period. It should be understood that a' has all the features of a.

3.1. Entrepreneurs

In every period an entrepreneur has three options.

1. At stage 2 after the realisation of the shock, an entrepreneur can file for bankruptcy protection. If the agent chooses to declare bankruptcy then it is assumed that they pay the bankruptcy cost βB and in return are protected from current period losses including interest on their debts. Defaulting entrepreneurs reappear as inventors in the following period and at that point

are indistinguishable from other inventors.[8] The discounted value of default-
ing is thus

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

where the value function of an inventor is denoted $V_I(a; V_E)$.

2. At stage 3 an entrepreneur can choose an orderly exit from production. In
this case, the entrepreneur absorbs current losses, pays interest due on the
loan, pays liquidation costs βL to close the project and enters next period
as an inventor. The value of orderly exit in state a is

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

The expected value of being an inventor next period, $E\left[V_I(a';.)\right] = \int_A V_I(a';.)G(da')$,
that features in the two exiting options will be a known constant in equi-
librium. The value of being an inventor is described in more detail in the
next section but a key feature is that the ideas drawn by an inventor are
independent through time and therefore the profitability state for an entre-
preneur in the current period has no effect on the expected value of being
an inventor in the next period.

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

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

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 each period 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

---

[8]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.

form of exit when facing a very bad profitability state. It is straightforward to see from equations (1) and (2) that entrepreneurs will default for all values of $a < a_\delta$ where

$$\pi(a_\delta, r) = \beta(L - B) \tag{5}$$

The threshold for orderly exit, a_X, results from the comparison of $V_X(a)$ and $V_C(a)$. The only tricky aspect of this problem is the conditional expected value of being an entrepreneur next period: $E[V_E(a'; .) \mid a]$. Consider first an entrepreneur with profitability above the loan covenant threshold, $a \geq a_T$. In this case the entrepreneur faces no risk if the bank chooses to monitor, so we can safely ignore the role of the bank and

$$E[V_E(a'; .) \mid a \geq a_T] = \int_A V_E(a'; .)F(a, da')$$

The calculation is more complex for an entrepreneur with $a_\delta < a < a_T$. 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 $a_\delta \leq a < a_T$

$$E[V_E(a'; .) \mid a_\delta \leq a < a_T] = (1 - \varphi)\int_A V_E(a'; .)F(a, da') + \varphi\left(\int_A V_I(a'; .)G(da') - L\right)$$

The voluntary exit threshold a_X is the value of current period profitability at which an entrepreneur is indifferent between continuing or exiting voluntarily. Some simple cancelling gives

$$\int_A V_E(a'; .)F(a_X, da') = \int_A V_I(a'; .)G(da') - L \tag{6}$$

(Note that a' refers to productivity draws next period on both sides but these are fundamentally different because they are drawn under different circumstances and from different distributions.)

3.2. Inventors

We turn now to the decision by inventors whether or not to enter. 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 project, S, and enter production next period or keep their capital on deposit at rate r_d for another period. Profitability next period will be subject to an idiosyncratic shock according to the same function F as existing projects. So the expected value of entering production net of S is equal to the expected value of being an existing entrepreneur at the same level of profitability. An inventor this period receives the interest on their 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 an initial draw of a is therefore:

$$V_I(a; V_E) = r_d + \beta \max \left\{ \int_A V_E(a'; ,) F(a, da\prime) - S, \int_A V_I(a'; ,) G(da') \right\} \quad (7)$$

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 defined as:

$$\int_A V_E(a'; ,) F(a_E, da\prime) - S = \int_A V_I(a'; ,) G(da') \quad (8)$$

As already mentioned, 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 proposition:

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. 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 r.*

Proof. The proofs of all propositions are contained in the Appendix. ∎

Proposition 1 states that there are unique values of the entry and exit thresholds, a_E, a_δ 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 r_d. So from here on we can dispense with the value functions.

3.3. Steady state

Define $H_t([0, a); \psi)$ as the measure of entrepreneurs at the *end* of period t with profitability levels in the interval $[0, a)$ given loan contract ψ. With the behavioural assumptions of the model and recalling that W is the measure of inventors, a law of motion for the distribution of entrepreneurs can be defined by:

$$H_{t+1}([0, a'); \psi) = W \int_{a_E}^{a'} G(a) + \int_{a_X}^{1} F(a, a') H_t(da; \psi) - \varphi \int_{a_X}^{a_T} F(a, a') H_t(da; \psi)$$
(9)

The first term enumerates how many agents enter at profitability levels at or below a'. The middle term describes how many continuing entrepreneurs evolve into profitability subset $[0, a')$ from the measure of entrepreneurs above the voluntary exit threshold, a_X. The third term eliminates those entrepreneurs closed down by the bank. Defaulting entrepreneurs are implicitly removed by the lower truncation of the distribution at a_X. An invariant distribution occurs if
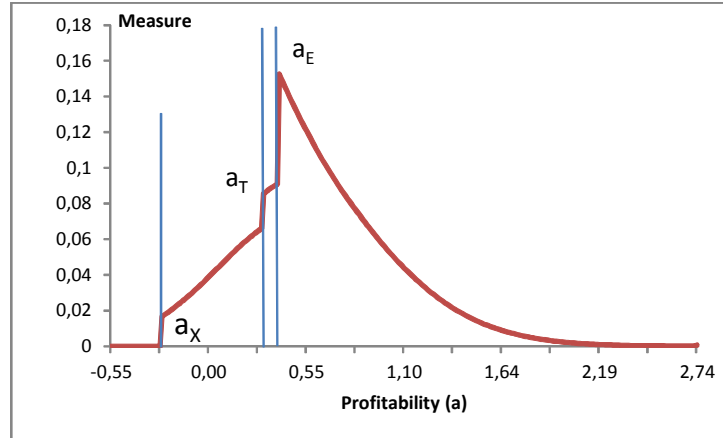
$$H_{t+1}([0, a); \psi) = H_t([0, a); \psi) = H([0, a); \psi) \quad \forall t \text{ and } \quad a \in A$$

Proposition 2. *For each ψ there is a unique invariant distribution, $H([0, a); \psi)$.*

3.4. Illustrative numerical simulation

Before turning to the decision of the bank, it is useful to illustrate an example of an invariant distribution and decompose the transition equation (9). The parameterisation in figure 2 is entirely illustrative but was chosen to give roughly sensible credit spreads and entry, exit and default rates.

Figure 2: Steady-state distribution of firms on the bank's balance sheet

It is easy to see the influence of the three behavioural thresholds on the distribution. Below a_X, there are no entrepreneurs in the distribution at the end of each period because they have either defaulted or exited voluntarily. Between a_X and a_T we have entrepreneurs that want to continue but are in breach of the loan covenant and thus at risk of having their loan recalled. Entrepreneurs in this region only survive if the bank does not monitored them. a_E marks the threshold at which it is just preferable to enter rather than wait another period. There is a concentration of entrepreneurs just above this level.

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. Those at the far right are well beyond the initial profitability draws so only exist because of the persistence of shocks and the luck of drawing positive shocks.

The model structure and figure 2 are largely consistent with the empirical evidence. Studies using US data, for example Bartelsman and Doms (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. Fariñas 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
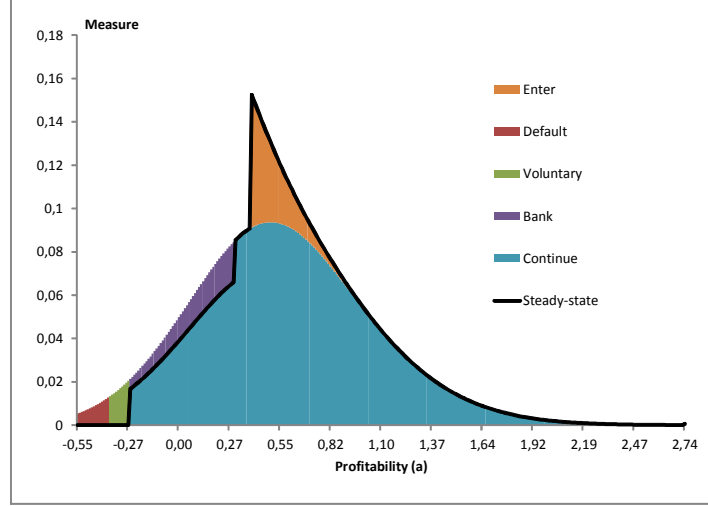
dominated by continuing firms.

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) then we obtain

$$\int_A V_E(a\prime, r, V_I) F(a_X, da\prime) = \int_A V_E(a', r, V_I) F(a_E, da\prime) - S - L \qquad (10)$$

from which it can be easily seen that if $S = L = 0$, then $a_E = a_X$. Since a covenant threshold is only relevant in the interval $[a_E, a_X]$, if we have $S = L = 0$ and $a_E = a_X$, then a_T is redundant and so is bank monitoring. In this case, even in the presence of default risk - the possibility of a stochastic fall in profitability below a_δ - 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 the bank would be willing to offer a new loan under full information. In other words, the 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 and asymmetric information because banks can no longer assume that a borrower has profitability at least as high as a new entrant. With large entry and exit costs, profitability of some existing borrowers will be considerably below the profitability of entering entrepreneurs and this creates the incentive for the bank to monitor.

Figure 3 illustrates the one period transition of the distribution in Figure 2 with the invariant distribution overlaid. 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 existing entrepreneurs - they slide to the left. 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 that want to continue but are at risk of having their loans recalled if the bank monitors them because they are in breach of the loan covenant. φ proportion of these entrepreneurs are monitored and exit and $1 - \varphi$ are able to continue. Moving further to the left, there are entrepreneurs that fall below a_X but above a_δ and exit voluntarily. Finally, there is a portion of the distribution that falls below a_δ and defaults.

Figure 3: One period transition of bank balance sheet

4. Equilibrium bank behaviour

We can now turn to the bank's choices of parameters in the loan contract $\psi = \{r, \varphi, \xi\}$. In this framework, the effects of φ and ξ are isomorphic. The bank can protect itself against default risk by raising the monitoring rate or tightening the covenant. If one were to assume that they are substitutes with a non-linear elasticity, one would find that they move together which is exactly what is illustrated in figure 1. But it is simpler to present the equilibrium by fixing the value of one parameter and making the other an endogenous choice variable. In what follows, the covenant value is fixed, making the monitoring rate endogenous. Given this, it is trivial to show that the covenant condition ξ will be satisfied for only one value of a and this defines a_T. This leaves pairs of r and φ. Proposition 2 asserted that there is a unique invariant distribution for any loan contract and thus (r, φ) pair. The bank, however, is constrained in its choice of loan contract by the need to finance its loans by deposits. 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 banks is:

$$H(A; r, \varphi) = \frac{1}{2} \tag{11}$$

Although choosing an optimal (r, φ) pair 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 use the monitoring rate to maximise profits.
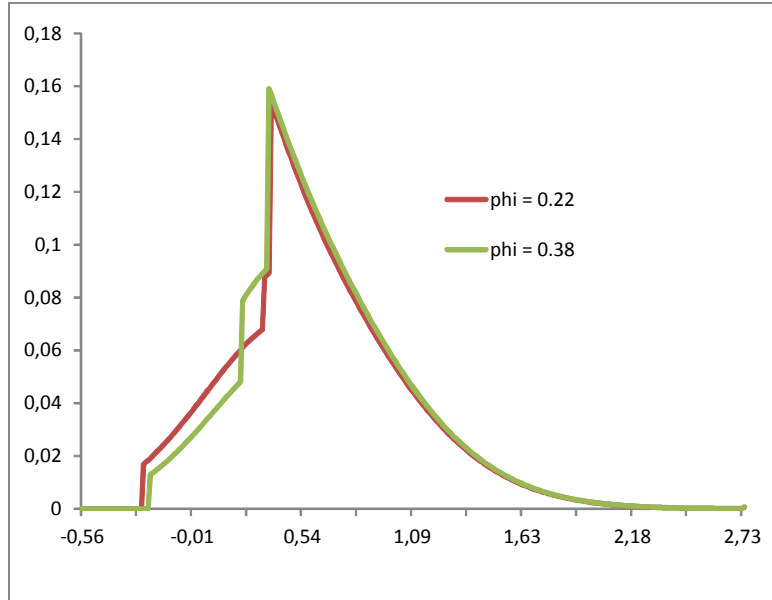
Proposition 3. *There is a unique value $\tilde{r}$ that ensures that the balance sheet of the bank is equal on both sides for given values of φ and ξ.*

This is a very intuitive proposition. If the bank faces an excess demand for loans, then raising the borrowing rate simultaneously reduces the demand for new loans (by increasing a_E) and increases the incentive for existing borrowers to repay and exit production voluntarily (an increase in a_X).[9] 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. With r a function of φ, we have a set of fully-funded invariant balance sheets described by $\tilde{H}(A, r(\varphi), \varphi)$.

Intuitively, of course, not all invariant balance sheets are the same. To illustrate what is at stake, Figure 4 compares the distributions arising from two different pairs of r and φ which satisfy the balance sheet constraint. When the monitoring rate is lower ($\varphi = 0.22$), it is intuitive that there are more projects in the left tail of the distribution because more borrowers are able to continue in breach of the covenant than otherwise. Since there is a higher potential survival rate, this also increases the interval over which it is rational for the borrower to gamble for resurrection - a_δ is lower. But more lenient contract terms are more favourable for the borrower so satisfying the bank balance sheet constraint requires a higher loan rate, r. The higher loan rate explains the other differences in the distribution. With a higher loan rate, the covenant threshold bites at a higher value of a (a_T is further to the right for $\varphi = 0.22$). A higher loan rate is also a disincentive to enter, so the entry threshold a_E is also higher, with the knock-on effect that there are marginally fewer projects with high profitability. Overall, the invariant profitability distribution for a higher monitoring rate stochastically dominates a distribution with a lower monitoring rate and has a lower default rate. In this sense, a lower monitoring rate results in a loan portfolio of lower "credit quality".

---

[9]It also increases the default rate by increasing a_δ. Since those going bankrupt are assumed to reappear as depositors in the next period, within the logic of the model, this also reduces the excess demand for loans.

Figure 4: Different Steady-state Distributions

Another way to think about the effect of an increase in φ is to consider how this affects the way entrepreneurs exit. Given the process for $F(a, a')$, entrepreneurs almost surely find themselves in the interval $[a_X, a_T]$ at some point and thus vulnerable to having their loan recalled by the bank. Since this is above the voluntary threshold, there is a utility loss for entrepreneurs ejected in this way. From the perspective of the individual agent, the recall of the loan results in premature and inefficient liquidation of the project. The parameter φ is thus effectively a distribution of control rights over the decision to exit production - the lower is φ, the higher the control rights allocated to the entrepreneur. Anticipating this, an entrepreneur will be willing to pay more for a loan with a lower monitoring rate because it gives the entrepreneur higher control rights over the exit decision and reduces this risk of premature liquidation. If earnings are volatile, this is potentially a very important consideration for a borrower. An entrepreneur required to reveal their state every period to the bank will not have its loan rolled over and be forced to pay the liquidation costs as soon as it fails to meet the covenant condition. A less monitored loan is thus a form of insurance against profit volatility and premature liquidation which an entrepreneur is willing to pay through a

higher average interest rate. From the bank's point of view, monitoring has two effects in this context. The direct effect is through discovering more entrepreneurs in the interval $[a_X, a_T]$ and thus reducing the measure of entrepreneurs that are at risk of default. But there is also an indirect effect because a higher monitoring rate reduces the expected gains from continuing, ceteris paribus.[10]

Having clarified these issues about the influence of r and φ, we can now state the decision problem of the bank more formally. The bank's objective is to:

$$
\max_{\{r,\varphi\}} \Pi = r \int_{a_\delta}^{1} F(a, da\prime) \tilde{H}(A; r, \varphi) - r_d \tilde{H}(A; r, \varphi) \tag{12}
$$

$$
- \int_{0}^{a_\delta} \lambda(a) F(a, da\prime) \tilde{H}(A; r, \varphi) - \varphi m(\varphi) \int_{a_X}^{1} F(a, da\prime) \tilde{H}(A; r, \varphi)
$$

where

$$
\tilde{H}(A; r, \varphi) = \frac{1}{2}
$$

, $m(\varphi)$ is the per unit cost of monitoring and $\lambda(a)$ is a parameter measuring loss given default which is decreasing in a. Equation (12) measures stationary bank profits. Recall that $\tilde{H}(A; r, \varphi)$ is defined at the *end* of each period. Since the bank is assumed to be sufficiently large that a law of large numbers applies, the variables in this problem are completely deterministic. The first term in equation (12) measures the interest income received on non-defaulting loans. Therefore the term takes into account the transition of the distribution from the end of the previous period. The second term deducts the payment of interest on all deposits. These deposits are contracted at the end of the previous period. The third term measures expected default losses and thus depends on the transition from the distribution at the end of the previous period. The final term measures the cost of monitoring those entrepreneurs that choose to continue. This term not only has the one period transition but is also truncated above a_X because of the voluntary exit of firms.

Figure 6 illustrates an interior solution to the model based on a constant per unit cost of monitoring m and loss given default $\lambda(a) = 2(\pi(a, r) + r) + L$. Discussion of the importance of the latter assumption is deferred to the following section on the general equilibrium properties of the model. The salient point is

---

[10]In equilibrium, as already noted, this ceteris paribus condition does not hold because of the endogenous effect on the loan rate.
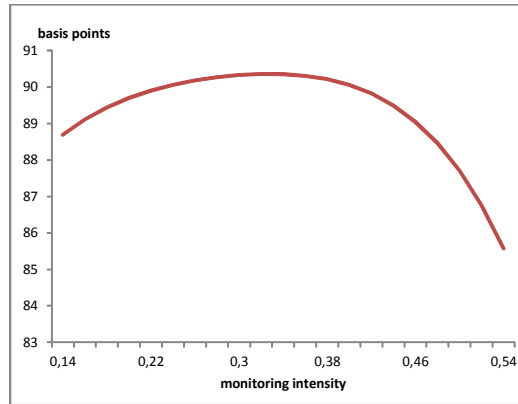
Figure 5: Profit per loan

Figure 6:

that there are additional losses for the project if it has to be liquidated by the bank in the event of default. As monitoring increases, average credit quality improves but at a decreasing rate. This decreasing marginal effectiveness of monitoring occurs because as the credit quality of the portfolio improves, there are fewer and fewer loans at risk of jumping to default. Monitoring is influencing a smaller proportion of borrowers. But monitoring is costly and the loan interest rate is falling because more highly monitored loans are less attractive to borrowers. For an interior solution, the reduction in loss given default dominates initially but is eventually outweighed by the reduction in effectiveness, the higher monitoring costs and the lower credit spread.

5. General equilibrium properties

The model presented in the previous two sections decribes a monopoly bank intermediating funds between heterogeneous agents who have the possibility to enter and exit production. Unlike in standard entry and exit models, in this framework there is an actor - namely, the bank - who has an interest in the aggregate distribution of firms and has instruments to influence the shape. By setting its constrained, profit-maximising monitoring rate, the bank determines the profile of

profitability of projects in the economy. It is the bank that ultimately determines aggregate profits, the turnover rate of firms and default risk.

It is interesting first to ask why there are any bank profits at all. Naturally, the assumptions that the bank is a monopoly and that inventors cannot fund entrepreneurs directly are crucial. The bank gets a rent from its position as an intermediary. But there are some more subtle issues too. One of these is that even in equilibrium, there is a difference in utility between borrowers and lenders. Only the marginal entrant is indifferent between borrowing and remaining a depositor.[11] The key friction is the assumption that projects have some market power and thus gross profitability (ie before loan interest) and that the outcomes are heterogeneous. Competition for the profits accessible through borrowing bids up the loan interest rate until the balance sheet condition is satisfied. An interesting consequence of these assumptions is that it is the bank that would gain all the benefit from a common shock to profitability. Consider, for example, a constant increase κ in the profitability of all firms. This does not change the relative positions of any of the agents and so the balance sheet constraint requires the bank to make the same marginal entrant indifferent which occurs if the loan rate is increased by exactly κ.[12] Banks are only forced to share the benefits of a common shock to profitability with the agents if the monetary policy authority raises the deposit rate.

A second important aspect of the general equilibrium of the model is the assumption about loss given default - $\lambda(a)$. Consider first the case in which $E\left[\lambda(a)\right] \geq \pi(a) - L$. In this situation, expected bank losses in the event of default are less than that which would be absorbed by the agent if they did not default. This could arise if banks are occasionally bailed out by taxpayers. In this case, the monitoring rate will be zero in equilibrium. This occurs because the option to default is valuable to the borrow and, as previously discussed, they are willing to pay a higher loan interest rate for a low rate of monitoring. If the bank can effectively shift some of these default losses onto the taxpayer, then the credit spread increases by more than the bank pays in expected loss given default. By setting the monitoring rate to zero, the bank maximises the implicit subsidy on default risk provided by taxpayers and minimises monitoring costs. Therefore in this case, credit standards are always as low as possible and thus unaffected by any shifts in parameters.

---

[11] At the point agents decide to be a depositor their expected utility is the same.

[12] Such a simple solution does not arise with a shock that alters the shape of the distribution of profits.

In the more likely case that $E\left[\lambda(a)\right] < \pi(a) - L$ - for example under the reasonable assumption that the bank is less good at running a project until it can be liquidated than entrepreneurs - the potential for an interior solution exists. In this case, there is a genuine trade-off between reducing default risk and the rising cost of monitoring.

A final interesting question is a comparison between the social welfare maximising level of monitoring and the one chosen by the bank. A natural social welfare benchmark is the sum of gross profitability less monitoring costs, the start-up and liquidation costs and the excess loss given default $(\lambda(a) - (\pi(a) - L))$. This is the aggregate amount of surplus which can be divided between the bank and the borrower. The comparison here is between the loan contract terms which maximise this surplus and the one chosen by the bank. As noted earlier, a higher monitoring rate results in a distribution of profitability which stochastically dominates a lower monitoring rate because the lower interest rate induces more entrants. The marginally induced entrants have higher profitability than the marginally induced exits. The higher credit quality distribution also has a lower default rate and thus lower losses given default. But set against this are the higher monitoring costs and also the costs of entry and exit. The higher the monitoring rate, the higher the turnover rate of firms and the higher then entry and exit costs. For similar reasons to the explanation of figure 6, the improvement in average quality dominates initially but the rate of improvement is decreasing and the costs are increasing until eventually the social benefits of monitoring peak and fall. As might be expected, the socially optimal rate of monitoring is higher than the bank choice. This occurs because the bank does not capture all of the benefits in the twist of the distribution - the agents gain some of the surplus as monitoring increases - but the bank pays the monitoring costs.

6. Monetary policy v the savings glut

Since the start of the recent financial crisis there has been a lively debate about whether the low credit standards observed in the years before it erupted were in part caused by accommodative monetary policy. On one side are those who charge monetary policy makers with having ignored or misunderstood the effect of a prolonged period of low official policy rate on the willingness of banks to take financial risks (Borio and Zhu (2008), Borio, C. and P. Disyatat (2011), Taylor (2009), Adrian and Shin (2008b), Gambacorta (2009)). On the other side, including major monetary policy makers, are those who argue that credit

standards deteriorated because of the strong inflows into the western banking system from excess savings in emerging market countries (Bernanke (2010), King (2010), Portes (2009)).

There is conflicting evidence on existence of the risk-taking channel. Jiménez, Ongena, Peydró and Saurina (2008) use data from the Spanish credit registry over the period 1984-2006 and find a statistically significant increase in the credit riskiness of *new* loans (measured by the duration to default of individual loans) when policy rates are lower at the time of loan origination. Moreover, if interest rates subsequently rise, then the hazard rate on these riskier loans is materially higher.[13] Ioannidou, Ongena, and Peydró (2008) report similar results using Bolivian credit register. They also report evidence that borrowers with worse credit histories are more likely to obtain loans when policy interest rates are low.[14] Adrian and Shin (2008a) and Adrian,Moench and Shin (2010) examine the interaction between monetary policy, the size of the balance sheets of leveraged financial institutions, credit risk premia, asset prices and macroeconomic activity. Although they differ in empirical strategy, a common theme is that during their pre-crisis sample period, non-bank financial intermediaries were the marginal price setter in many risky asset markets. Even commercial banks, which could really largely on deposits, would borrow on wholesale markets to add to their lending capacity and were thus affected by the market price of risk. A crucial finding of their empirical analysis is that financial sector leverage is highly sensitive to short-term official interest rates. By contrast, Lown and Morgan (2008) find no evidence in the US that credit standards are affected by the policy rate. And Dell'Ariccia et al (2012) show that the key factors determining whether a credit boom is followed by a crisis are the length of the boom and the stance of macroprudential policies. The stance of monetary policy is insignificant.
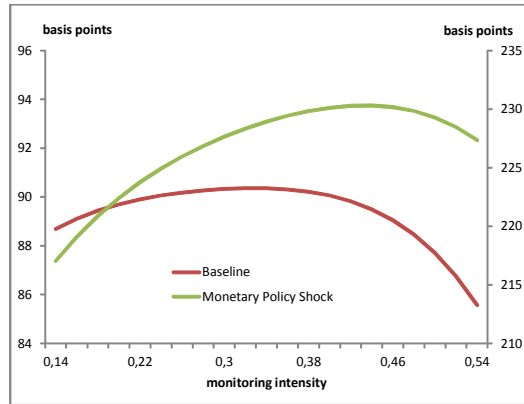
The theoretical model developed in this paper can be used to compare the effect of the two shocks on equilibrium credit standards and credit spreads. The equilibria presented are invariant distributions so the implicit assumption is that these shocks are permanent.[15] The results presented below are general, not parameter-

---

[13]If the interest rate at origination was 4.13% and the ex post interest rate was 4.09%, then the estimated annualised hazard rate is 0.56%. If the loan rate at origination was the lowest in the sample (2.16%) and the ex post interest rate the highest (9.62%), then the hazard rate would be around 6 times higher at 3.38%.

[14]See also Altunbas, Gambacorta and Marques-Ibanez (2010) and Maddaloni and Peydró (2010).

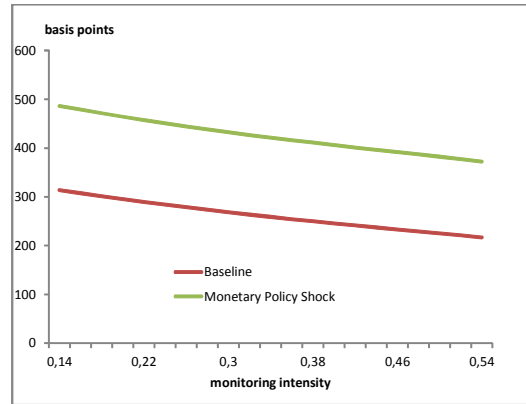[15]Dynamic versions of the model are currently under development.

Figure 7: Monetary policy shock - monitoring intensity

specific. Figure 7 shows that in contradiction of the hypothesis of a risk-taking channel the profit-maximising monitoring intensity increases in response to a cut in monetary policy. And figure 8 illustrates that credit spreads are higher too. How does such a counterintuitive result arise? The partial effects are in place for a risk-taking channel: the lower deposit rate increases the incentive for inventors to enter production and encourages existing entrepreneurs to continue rather than exit and the default rate would fall *ceteris paribus*. The reduction in the deposit rate increases the lending spread and thus the incentive to lend. The problem is the balance sheet constraint. Without an external source of funds, the only way for the bank to reduce the resulting excess demand for loans is to make borrowing less attractive and so loan rates increase. But since increasing the loan rate is profitable and monitoring is costly, why doesn't the bank rely solely on interest rates to equilibrate its balance sheet? Why does monitoring also increase? The reason is that in this model the average credit quality of the loan portfolio is decreasing in the interest rate - the portfolio is less refreshed by new borrowers. A combination of higher interest rates and higher monitoring offers a better mix of loan spread and default risk. Note that a cut in monetary policy substantially increases bank profits per loan.

By contrast, figures 9 and 10 show that a positive exogenous increase in deposits reduces monitoring intensity and reduces credit spreads. In this case, the bank finds itself with an excess supply of deposits and must reduce credit terms to induce more loan demand. Clearly one option is to cut loan interest rates.

Figure 8: Monetary policy shock - credit spreads

But the bank will also try to preserve some of its credit spread by weakening the monitoring intensity. It can do this partly because the lower interest rate reduces default risk *ceteris paribus*. However, once the reduction in monitoring intensity is taken into account, the default rate is higher in the deposit shock equilibrium than in the baseline.

Overall, therefore, the comparison of the two shocks shows that, at least as far as the model presented in this paper is concerned, the global savings glut is a more likely explanation for the reduction in credit standards and lending spreads observed in the run-up to the crisis than the risk-taking channel of monetary policy.

7. Conclusion

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 presented in this paper has shown how these competing considerations can be equilibrated whilst ensuring that the bank has sufficient deposits to fund its lending. The key friction in the model is the asymmetry of information between the borrower and the lender - the borrower is always better informed than the lender about the profitability of the project the loan is funding. In the context of multi-period loans, less than complete monitoring gives the borrower more
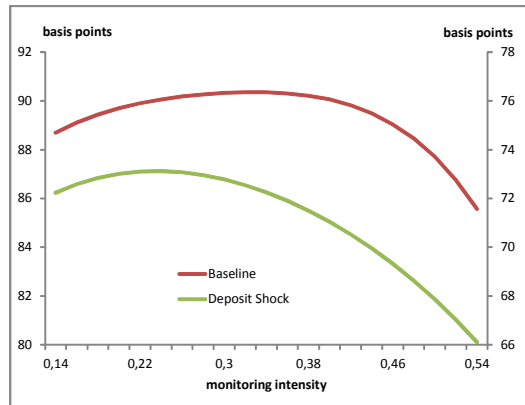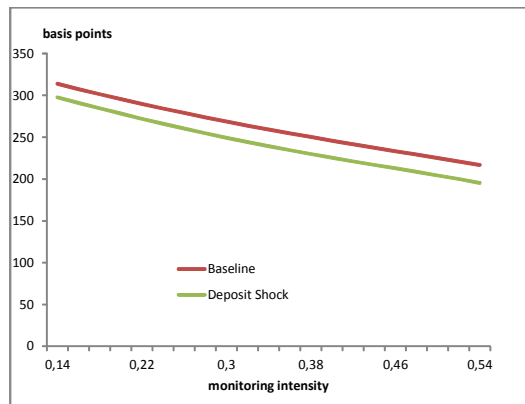
Figure 9: Deposit shock - monitoring intensity

Figure 10: Deposit shock - credit spreads

opportunity to act than the bank. The interests of the borrower and the bank are partially aligned because borrowers have an incentive to close down unprofitable projects. But the ability to obtain bankrupcty protection shifts some of the down side risk onto the bank and thereby creates a tension between the interests of the borrower and the bank over whether a project should continue. It is this which gives the incentive for the bank to monitor loans. But as the paper has shown, credit standards are essentially a control right over the continuation decision and borrowers will pay a higher interest rate for a greater say. This is likely to be particularly important for projects with highly volatile profits. Since monitoring is costly and increasingly less effective at reducing credit risk, banks are willing to offer loan contracts with less than complete monitoring.

The choice of credit standards has a profound effect on the distribution of firms and thus the allocation of resources in the economy. It also affects the rate of entry and exit of firms in equilibrium and the default rate. The profit maximising choice of lending standards, however, is likely to be below the social optimal because banks bear all the social costs of monitoring but only gain a portion of the social benefits.

In the final section, the paper examined the effects of a cut in monetary policy and a deposit shock on equilibrium credit standards and credit spreads. In a closed economy, a cut in monetary policy led to a tightening of credit standards and an increase in credit spreads in contradiction to the risk-taking channel of monetary policy. By contrast, an exogenous increase in deposits reduced credit standards and credit spreads, consistent with what was observed prior to the financial crisis. Clearly this model is not a complete description of banking behaviour and therefore one cannot entirely rule out the risk-taking channel in favour of the savings glut hypothesis but it does suggest that the ability to obtain foreign funding is a crucial element in the evolution of credit standards.

8. References

References

[1] Acharya, V. and H. Naqvi (2011), 'The seeds of a crisis: a theory of bank liquidity and risk-taking over the business cycle', *Journal of Financial Economics* forthcoming.

[2] Adrian, T., E. Moench, H.S. Shin (2010), 'Financial intermediation, asset prices, and macroeconomic dynamics' Federal Reserve Bank of New York Staff Report no 422, January 2010.

[3] Adrian, T. and H.S. Shin (2008a), 'Liquidity and leverage' Federal Reserve Bank of New York Staff Report no 328, May 2008

[4] Adrian, T. and H.S. Shin (2008b), 'Procyclical leverage and value-at-risk' *Federal Reserve Bank of New York Staff Report* no 338, July 2008.

[5] Adrian, T. and H.S. Shin. (2009), 'Financial Intermediaries and Monetary Economics' In Handbook of Monetary Economics, Vol. 3, ed. Benjamin Friedman and Michael Woodford. North-Holland.

[6] Agur, I. and M. Demertzis, (2010), 'Monetary policy and excessive bank risk taking' mimeo Dutch Central Bank.

[7] Baily, M., E. Bartelsman and J. Haltiwanger (2001), 'Labor productivity: structural change and cyclical dynamics', *The Review of Economics and Statistics*, vol. 83, 3, pp. 420-433.

[8] Bartelsman, E. and M. Doms (2000), 'Understanding productivity: lessons from longitudinal microdata', *Journal of Economic Literature*, vol. 38, pp. 569-594.

[9] Bernanke, B.S. (2010), 'Monetary policy and the housing bubble', speech at the Annual Meeting of the American Economic Association, Atlanta, Georgia.

[10] Borio, C. and P. Disyatat (2011), 'Global imbalances and the financial crisis: link or no link?' *BIS Working Papers* no 346, Bank for International Settlements.

[11] Borio, C. and H. Zhu (2008), 'Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism?' *BIS Working Papers* no 268, Bank for International Settlements.

[12] Dell'Ariccia, G., D. Igan, L. Laeven and H. Tong with B. Bakker and J. Vandenbussche (2012), 'Policies for macrofinancial stability: how to deal with credit booms' IMF Staff Discussion Note 12/06.

[13] Dell'Ariccia, G. , L. Laeven and R. Marquez (2010), 'Monetary policy, leverage and bank risk-taking' *IMF Working Paper* no 10/276.

[14] Diamond, D.W. (1984), 'Financial intermediation and delegated monitoring' *Review of Economic Studies*, vol. 51, 3, pp. 393-414.

[15] Fariñas, J. and Ruano, S. (2005) 'Firm productivity, heterogeneity, sunk costs and market selection', *International Journal of Industrial Organization*, vol. 23, pp. 505-534.

[16] Foster, L., J. Haltiwanger and C.J. Krizan (2006), 'Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s', *The Review of Economics and Statistics*, vol. 88, 4, pp. 748-758.

[17] Gambacorta, L. (2009), 'Monetary policy and the risk-taking channel', *BIS Quarterly Review*, December.

[18] Hopenhayn, H. (1992), 'Entry, exit and firm dynamics in long run equilibrium', *Econometrica*, vol. 60, 5, pp. 1127-1150.

[19] Lown, C. and D.P. Morgan (2008), 'The credit cycle and the business cycle: new findings using the loan officer opinion survey', *Journal of Money, Credit and Banking*, vol 38., no.6, pp.1575-1597.

[20] Kindleberger, C. (1978), **Manias, Panics, and Crashes: A History of Financial Crises**, New York: Basic Books.

[21] King, M.A. (2010) speech at the University of Exeter.

[22] Portes, R. (2009), 'Global imbalances' in M. Dewatripont, X. Freixas and R. Portes (eds), **Macroeconomic stability and financial regulation: key issues for the G20**, Centre for Economic Policy Research, London.

[23] Stachurski, J. (2012) mimeo.

[24] Stein, J. (2013) 'Overheating in credit markets: origins, measurement, and policy responses' speech given at the "Restoring household financial stability after the Great Recession: why household balance sheets matter" research symposium sponsored by the Federal Reserve Bank of St. Louis, St. Louis, Missouri.

[25] Taylor, J. (2009), 'The financial crisis and the policy response: an empirical analysis of what went wrong', *NBER working papers*, no. 14631.

9. Appendix

9.1. 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. These value functions yield unique and continuous functions in ψ for the entry and exit thresholds a_E and a_X respectively. $a_E(\psi)$ and $a_X(\psi)$ are both strictly increasing in r.*

Proof. It greatly simplifies the presentation of the proof (and with no loss of

generality) to ignore the presence of the default option and focus on the choices of voluntary exit or continuation.

Let $\mathcal{B}(A)$ denote the set of all bounded functions on the set A. And for $v \in \mathcal{B}(A)$, let $\|v\| = \sup_a |v(a)|$ be the usual sup norm. Re-write equation (4) as

$$V_E(a) = \pi(a, r) + \beta \max \{\phi_1 (V_E, V_I, a), \psi_1(V_I, a)\}$$

where

$$\phi_1 (V_E, V_I, a) = \varphi I_{a_T}(a) \left(\int_A V_I(a; .)G(da) - L \right) + (1 - \varphi I_{a_T}(a)) \int_A V_E(a\prime; ,)F(a, da\prime)$$

$I_{a_T}(a)$ is an indicator function with value 1 if $a < a_T$ and zero otherwise and

$$\psi_1(V_I, a) = \int_A V_I(a; .)G(da) - L$$

Define T_1 to be the operator $(V_E, V_I) \mapsto T_1 (V_E, V_I)$:

$$T_1 (V_E, V_I) (a) = \pi(a, r) + \beta \max \{\phi_1 (V_E, V_I, a), \psi_1(V_I, a)\}$$

For any $V_I \in bA$, $V_E \mapsto T_1 (V_E, V_I)$ satisfies both of Blackwell's sufficient conditions for a contraction on $\mathcal{B}(A)$ with modulus β. Thus a unique V_E exists with

$$V_E = T_1 (V_E, V_I) \tag{13}$$

Similarly re-write equation (7) as

$$V_I(a) = r_d + \max \beta \left\{ \phi_2 \left(V_E, a \right), \psi_2(V_I, a) \right\}$$

where

$$\phi_2 \left(V_E, a \right) = \int_A V_E(a'; ,) F(a, da\prime) - S$$

and

$$\psi_2(V_I, a) = \int_A V_I(a; .) G(da)$$

Define T_2 to be the operator $(V_E, V_I) \mapsto T_2 \left(V_E, V_I \right)$:

$$T_2 \left(V_E, V_I \right)(a) = r_d + \max \beta \left\{ \phi_2 \left(V_E, a \right), \psi_2(V_I, a) \right\}$$

Again, for any $V_E \in bA$, $V_I \mapsto T_2 \left(V_E, V_I \right)$ satisfies both of Blackwell's sufficient conditions for a contraction on $\mathcal{B}(A)$ and a unique V_I exists with

$$V_I = T_2 \left(V_E, V_I \right) \tag{14}$$

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 (13) and (14) simultaneously.

As a preliminary step towards the proof of the existence of a unique pair note that for $V_I, V_I' \in \mathcal{B}(A)$

$$\left| \psi_1(V_I, a) - \psi_1(V_I', a) \right| = \left| \int_A \left[V_I(a; .) - V_I'(a; .) \right] G(da) \right| \leq \| V_I - V_I' \|$$

The absolute value of the difference in expected value between any V_I and V_I' must be less than the largest absolute difference. By similar arguments:

$$\left| \psi_2(V_I, a) - \psi_2(V_I', a) \right| = \left| \int_A \left[V_I(a; .) - V_I'(a; .) \right] G(da) \right| \leq \| V_I - V_I' \|$$

$$\left| \phi_2 \left(V_E, a \right) - \phi_2 \left(V_E', a \right) \right| = \left| \int_A \left[V_E(a\prime; ,) - V_E'(a\prime; ,) \right] F(a, da\prime) \right| \leq \| V_E - V_E' \|$$

Function $\phi_1(V_E, V_I, a)$ is a bit more tricky because of the presence of both value functions but note that for $V_E, V'_E, V_I, V'_I \in \mathcal{B}(A)$

$$
\begin{aligned}
|\phi_1(V_E, V_I, a) - \phi_1(V'_E, V'_I, a)| &= \left| \begin{array}{l} \varphi I_{a_T}(a) \left[\int_A [V_I(a; .) - V'_I(a; .)] G(da) \right] \\ + (1 - \varphi I_{a_T}(a)) \left[\int_A [V_E(a\prime; ,) - V'_E(a\prime; ,)] F(a, da\prime) \right] \end{array} \right| \\
&\leq \|V_E - V'_E\| \vee \|V_I - V'_I\| \\
\text{since } 0 &\leq \varphi I_{a_T}(a) \leq 1
\end{aligned}
$$

Now let $\mathcal{M}$ be the set of ordered pairs (V_E, V_I) such that both V_E and V_I are in $\mathcal{B}(A)$. Impose the following metric d on $\mathcal{M}$:

$$
d((V_E, V_I), (V'_E, V'_I)) = \|V_E - V'_E\| \vee \|V_I - V'_I\|
$$

where $a \vee b$ is the max of a and b. Now consider the operator $T : \mathcal{M} \to \mathcal{M}$ defined by

$$
T(V_E, V_I) = (T_1(V_E, V_I), T_2(V_E, V_I))
$$

A fixed point exists if

$$
(V_E, V_I) = (T_1(V_E, V_I), T_2(V_E, V_I))
$$

which is equivalent to

$$
V_E = T_1(V_E, V_I)
$$

and

$$
V_I = T_2(V_E, V_I)
$$

which are the separate value functions. Fix $a \in A$ and observe that

$$
\begin{aligned}
|T_1(V_E, V_I)(a) - T_1(V'_E, V'_I)(a)| &= \beta |\phi_1(V_E, V_I, a) \vee \psi_1(V_I, a) - \phi_1(V'_E, V'_I, a) \vee \psi_1(V'_I, a)| \\
&\leq \beta\{ |\phi_1(V_E, V_I, a) - \phi_1(V'_E, V'_I, a)| \vee |\psi_1(V_I, a) - \psi_1(V'_I, a)| \} \\
&\leq \beta\{ \|V_E - V'_E\| \vee \|V_I - V'_I\| \}
\end{aligned}
$$

where the first inequality is simply an example of the general property that $|\max(a, b) - \max(c, d)| \leq \max(|a - c|, |b - d|)$ for any $a, b, c, d \in \mathbb{R}$ and the

second inequality uses the properties of ϕ_1 and ψ_1 stated above. Taking the supremum over both sides:

$$\|T_1\left(V_E, V_I\right)(a) - T_1\left(V'_E, V'_I\right)(a)\| \leq \beta\{\|V_E - V'_E\| \vee \|V_I - V'_I\|\}$$

Exactly the same arguments give

$$\|T_2\left(V_E, V_I\right)(a) - T_2\left(V'_E, V'_I\right)(a)\| \leq \beta\{\|V_E - V'_E\| \vee \|V_I - V'_I\|\}$$

Therefore

$$\|T_1\left(V_E, V_I\right)(a) - T_1\left(V'_E, V'_I\right)(a)\| \vee \|T_2\left(V_E, V_I\right)(a) - T_2\left(V'_E, V'_I\right)(a)\| \leq \beta\{\|V_E - V'_E\| \vee \|V_I - V'_I\|$$

This is the same as

$$d(T\left(V_E, V_I\right), T\left(V'_E, V'_I\right)) \leq \beta d\left(\left(V_E, V_I\right), \left(V'_E, V'_I\right)\right)$$

Hence T is a contraction mapping on the complete metric space $(\mathcal{M}, d)$ establishing a unique fixed point exists.

Both $V_E(a)$ and $V_I(a)$ are continuous. For states above the default threshold, $a \geq a_\delta$, $V_E(a)$ is strictly increasing in a through the properties of $\pi(a, r)$.

Now we need to show that a_E and a_X, which are defined uniquely by $V_E(a; \psi, V_I)$ and $V_I(a; \psi, V_E)$, are both strictly increasing in r. As a preliminary, recall from the text that a_E and a_X are related by the following equation

$$\int_A V_E(a\prime; .)F(a_X, da\prime) = \int_A V_E(a'; .)F(a_E, da\prime) - S - L$$

Since S and L are constant, $V_E(a; .)$ is strictly increasing in a for $a_E \geq a_X > a_\delta$, and $F(a, a')$ decreasing in a, it follows that any increase in a_E has to be associated with an increase in a_X. So it suffices to show that a_E is strictly increasing in r to establish the case for a_X.

Note that $a_E \geq a_X > a_\delta$ so we are in the region of $V_E(a; \psi, V_I)$ that is strictly increasing in a and decreasing in r. The proof that a_E is strictly increasing in r is long and involved so again it might help the reader to have a road-map. The key to the intuition of the proof is equation (8):

$$\int_A V_E(a'; ,)F(a_E, da\prime) - S = \int_A V_I(a'; ,)G(da')$$

$V_E(a';,)$ is directly a function of r through the effect on $\pi(a,r)$. But $V_I(a';,)$ is an increasing function of $V_E(a';,)$ and so is indirectly a function of r too. So both sides of equation (8) will be affected by a change in r and in the same direction. But if an increase in r reduces the left hand side by more than the right hand side for a given a_E, then a_E must rise to restore equality since $F(a,da\prime)$ strictly increasing in a. Intuitively, it makes sense that a change in r should have more effect directly than indirectly but it takes a long time to get there. The means to do this is to find $\int_A V_I(a';,)G(da')$ and $\int_A V_E(a';,)F(a_E,da\prime)$ as functions of a_E, a_X, the distribution functions $F(a,a')$ and $G(a)$ and $\pi(a,r)$. Then we can take derivatives of both sides with respect to r keeping a_E and a_X constant and show that indeed the left is more sensitive than the right and thus a_E is increasing in r. It will be easy to see why fixed loan size facilitates this proof. To reduce notation, the effect of φ will be suppressed by setting it to 0. A very simple transformation of notation shows that the proof is identical if $\varphi = 1$. $0 < \varphi < 1$ is simply a complicated convex combination which is not reported. The default option is also ignored as it simply adds the sum of a probability-weighted discounted constant to the problem.

We start by finding the limit of $\int_A V_E(a';,)F(a_E,da\prime)$ through repeated substitution (and the natural transversality condition).

$$\lim_{T\to\infty} \int_A V_E(a';.)F(a,da\prime) = \beta\Delta\,(a)\left(\lim_{T\to\infty} EV_I - L\right) + \Lambda(a) \tag{15}$$

where

$$\Delta\,(a) \equiv \lim_{T\to\infty}\left\{\sum_{t=0}^{T}\beta^t\int_0^{a_X} F(a',da'')F_{uX}^t(a,da')\right\}$$

$$\Lambda(a) \equiv \lim_{T\to\infty}\left\{\sum_{t=0}^{T}\beta^t\int_A \pi(a_{t+1};.)F_t(a',da'')F_{uX}^t(a,da')\right\}$$

and

$$F_{uX}^t(a,da') \text{ is the t-th iteration of } \int_{a_X}^{a_u}..\int_{a_X}^{a_u}\int_{a_X}^{a_u} F(a,da\prime)..F(a,da\prime)F(a,da\prime)$$

$\Delta\left(a\right)$ measures the "survival-weighted" discount rate. It is necessarily the case that $0 \leq \Delta\left(a\right) \leq 1$. *If* $a_X = 1$ so that it is optimal for all projects to exit in period 0, then $\Delta\left(a\right) = \beta^0 = 1$. On the other hand, if $a_X = 0 + \epsilon$, then $\Delta\left(a\right) \to 0$ as $\epsilon \to 0$ because the probability of exit becomes vanishingly small. Since survival rates are increasing in a because $F(a, a')$ is strictly decreasing, it follows that $\Delta\left(a_k\right) < \Delta\left(a_j\right)$ for $a_k > a_j$. Relatedly, $\Delta\left(a\right)$ is increasing in a_X since the probability of surviving to any given future period is strictly smaller. $\Lambda(a)$ records the discounted expected profits of a project starting at profitability state a. The longer an entrepreneur survives, the higher the sum of expected profits, although given discounting the weight applying to future dated profits is smaller over time. (Since profits are bounded, there is a finite limit of this summation.) $\Lambda(a)$ is increasing in a and decreasing in a_X.

Turning to EV_I, we have (again with the obvious transversality condition)

Taking the limit of both sides we can obtain

$$
\lim_{T\to\infty} EV_I = r_d - \left(1 - \int_0^{a_E} G(da')\right)S + \beta \lim_{T\to\infty} EV_I \int_0^{a_E} G(da')
$$
$$
+ \lim_{T\to\infty} \int_{a_E}^1 \left[\beta \int_A V_E(a'', r, V_I)F(a', da'') \right] G(da')
$$

and by the monotone convergence theorem the limit passes under the integral and we have

$$
\lim_{T\to\infty} EV_I = r_d - \left(1 - \int_0^{a_E} G(da')\right)S + \lim_{T\to\infty} EV_I \beta \int_0^{a_E} G(da')
$$
$$
+ \int_{a_E}^1 \left[\beta \lim_{T\to\infty} \int_A V_E(a''; .)F(a', da'') \right] G(da')
$$

Substituting in equation (15) and re-arranging gives a term solely in exit thresh-

olds and transition probabilities:

$$\lim_{T \to \infty} EV_I = \frac{r_d - (1 - \int_0^{a_E} G(da'))S + \int_{a_E}^1 \beta \left[\Lambda(a') - \beta \Delta(a') L\right] G(da')}{\left[1 - \beta \int_0^{a_E} G(da') - \beta^2 \int_{a_E}^1 \Delta(a') G(da')\right]} \quad (16)$$

(The analysis of limits present above is equivalent to the case for $\varphi = 0$. If instead we consider the case in which $\phi = 1$ (i.e. full monitoring), then the structure of the analysis is identical simply by exchanging a_T for a_X in the integrals and iterations, noting that although $V_E(a, r)$ is discontinuous at a_T, it is bounded over all $a \in A$ and thus the expected values are defined and continuous due to the continuity of F.)

Now if we take the derivative of equation (15) with respect to r we get

$$\frac{\partial \lim_{T \to \infty} \int_A V_E(a'; .) F(a_E, da\prime)}{\partial r} \mid a_E = \beta \Delta(a_E) \frac{\partial EV_I}{\partial r} + \frac{\partial \Lambda(a_E)}{\partial r}$$

The final term is

$$
\begin{aligned}
\frac{\partial \Lambda(a_E)}{\partial r} &= \lim_{T \to \infty} \left\{ \sum_{t=0}^T \beta^t \int_A \frac{\partial \pi(a_{t+1}; .)}{\partial r} F_t(a', da'') F_{uX}^t(a, da') \right\} \\
&= (-1) \lim_{T \to \infty} \left\{ \sum_{t=0}^T \beta^t \int_A F_t(a', da'') F_{uX}^t(a, da') \right\} \\
&= -\lim_{T \to \infty} \left\{ \sum_{t=0}^T \beta^t F_{uX}^t(a, da') \right\} \\
&\equiv -\Upsilon
\end{aligned}
$$

where $0 \leq \Upsilon = \lim_{T \to \infty} \left\{ \sum_{t=0}^T \beta^t F_{uX}^t(a, da') \right\} \leq \frac{1}{1-\beta}$.

From equation (16):

$$\frac{\partial \lim_{T \to \infty} EV_I}{\partial r} \Bigg|\ a_E, a_X = \frac{\int\limits_{a_E}^{1} \beta \left[\frac{\partial \Lambda(a')}{\partial r}\right] G(da')}{1 - \beta \left[G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(da')\right]}$$

$$= \frac{-\beta \Upsilon (1 - G(a_E))}{\left[1 - \beta \left[G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(da')\right]\right]}$$

Therefore

$$\frac{\partial \lim_{T \to \infty} \int\limits_{A} V_E(a'', r, V_I) F(a_E, da'')}{\partial r} = - \left(\frac{\beta (1 - G(a_E)) \beta \Delta(a_E) + 1 - \beta \left[G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(da')\right.}{\left[1 - \beta \left[G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(da')\right]\right]} \right.$$

Finally, by comparing the two derivatives,

$$\frac{\partial \lim_{T \to \infty} \int\limits_{A} V_E(a'', r, V_I) F(a_E, da'')}{\partial r} - \frac{\partial \lim_{T \to \infty} EV_I}{\partial r} \Bigg|\ a_E =$$

$$- \left(\frac{(1 - \beta) - \beta^2 \left[\int\limits_{a_E}^{1} \Delta(a') G(da') - (1 - G\right.}{\left[1 - \beta \left[G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(\right.\right]}\right.$$

Take the denominator first. $0 \le G(a_E) + \beta \int\limits_{a_E}^{1} \Delta(a') G(da') \le 1$ given the range

of $G(a)$ and $0 \le \Delta(a') \le 1$. So the denominator is strictly positive for $\beta < 1$.

The numerator depends on the sign of the term in brackets. Notice that this compares the expected value of $\Delta(a')$ conditional on entry with the probability-of-entry-weighted value of Δ at the entry threshold. It was established above that $\Delta(a')$ strictly decreasing in a, so $\int_{a_E}^{1} \Delta(a') G(da') - (1 - G(a_E)) \Delta(a_E) < 0$ and the numerator is strictly positive for $\beta < 1$. Since $0 \leq \Upsilon \leq \frac{1}{1-\beta}$, the difference in derivatives is non-negative and strictly positive for $\beta < 1$. This establishes that

$$\int_A V_E(a', r_2, V_I) F(a_E(r_1), da') - S < \int_A V_I(a', r_2) G(da') \text{ and thus } a_E(r_2) > a_E(r_1)$$

the result to be established.

Finally, it is trivial to show that the threshold a_T is increasing in r. ∎

Proposition 2 *For each ψ there is a unique invariant distribution, $H([0, a); \psi)$ $a \in A$.*

Proof. The transition equation for the end of period distribution of entrepreneurs can be re-written as an operator on probability measures:

$$(T^*H)(A, r) = W \int_{a_E}^{1} G(a) + \int_{a_X}^{1} F(a, a') H_t(da; \psi) - \varphi \int_{a_X}^{a_T} F(a, a') H_t(da; \psi) \tag{17}$$

The state-space for this function is the endogenously compact set $[a_X, 1]$ since by self-selection, agents are only productive at profitability above the exit threshold level. Since $W \leq 1$, $\varphi \leq 1$, and $G(a)$ and $F(a, a')$ are probability measures, T^* is bounded. Since $\int_{a_E}^{1} G(a)$ and $F(a, a')$ are continuous probability measures in H is continuous by assumption, T^* maps a continuous function into another continuous function and thus has the Feller property. Therefore the operator function (17) satisfies the requirements for Theorem 12.10 in Stokey and Lucas (1989) and an invariant distribution exists. T^* is also obviously monotonic in H_t since $0 \leq F(a, a') \leq 1$. Monotonicity plus the assumption of recurrence, Assumption A (iii), ensure the requirements of Theorem 12.12 in Stokey and Lucas (1989) are satisfied and the invariant distribution is unique. ∎

Proposition 3 *There is a unique value $\tilde{r}$ that ensures that the balance sheet of the bank is equal on both sides for given values of φ and ξ.*

Proof. For the bank to be able to match deposits with liabilities we require:

$$\frac{1}{2} = \tilde{H}(A, \tilde{r}) \tag{18}$$

From equation (9) one can observe that the forcing term is the entry term and is driven by $\int_{a_E}^{1} G(a')$. The second and third terms are scaling terms. To reduce $\int_{a_X}^{1} \tilde{H}(da, \tilde{r})$ we can (i) lower the entry rate by raising a_E and/or (ii) reduce survival probabilities in the second term by increasing a_X and/or (iii) increasing the effective zone of monitoring by increasing a_T. Note that since $\varphi \leq 1$, the effect of a_X on the third term cannot dominate the effect on the second. By Proposition 1 all routes point to increasing r. H is a unique and continuous function from Proposition 2, so therefore so is $\int_{a_X}^{1} \tilde{H}(da, \tilde{r})$. There is thus a unique value for r which satisfies equation (18). ∎