

Banks and the credit cycle: a framework for analysing legacy assets

Adrian Penalver

Paris School of Economics and Banque de France*

September 11, 2014

Abstract

Legacy assets following a banking crisis are loans retained on the balance sheet that are impaired or underperforming that a bank would not retain if it could dispose of them. This type of behaviour can only arise if loans are made for multiple periods and the bank has limited rights to reoptimise its loan portfolio after shocks. The paper builds a model of credit risk management for a bank making long term loans to fund projects subject to idiosyncratic and aggregate shocks. The objective of the bank is to maximise profits, minimise variance in profits whilst trying to keep the endogenous volume of loans as close as possible to the endogenous supply of deposits. The solution to the dynamic programming problem faced by the bank is a set of policy functions for the loan interest rate, the deposit interest rate and a monitoring rate. Loan and deposit interest rates are pro-cyclical but do not fully absorb the aggregate shock. Default risk is thus counter-cyclical which results in counter-cyclical monitoring intensity and covenant threshold. The model is solved under two cases: fixed aggregate savings and pro-cyclical savings. Only in the second case does a legacy assets problem arise.

*This paper has benefited from comments received at the Paris School of Economics and the Bank of England. The views expressed here are not necessarily those of the Banque de France.

1. Introduction

During banking crises, policy makers often refer to ‘legacy assets’ on the balance sheets of banks.¹ These are loan commitments made in the past which are now impaired or underperforming and impeding the current performance of the bank. When a troubled bank is divided into a “good bank” and a “bad bank” the aim is sever this link between the history of lending and the issuance of new credit. If a banking crisis is systemic, the problem of legacy assets can have macroeconomic effects as well because they constrain the supply of new credit to the economy as a whole. Lack of new credit and the misallocation of existing credit is one possible explanation for why recessions following banking crises are deeper and longer than “normal” recessions (IMF WEO (2009)). Potentially profitable new firms are not started or new investments undertaken whilst capital is retained in weak or failing legacy assets.

The idea that output (and presumably welfare) increasing capital reallocations do not take place seems a significant market failure. And commentary on the behaviour of lenders before and after banking crises is sometimes expressed in terms of departures from rational forward looking decision-making (myopia, rational exuberance, search for yield) often citing Kindleberger (1978) and Minsky (1992). An increasing question, however, is whether these legacy assets might arise out of equilibrium behaviour. Could legacy assets be the observable downside risk of rational forward looking banks acting under uncertainty? The purpose of this paper is to develop a framework for analysing the origins and consequences of legacy assets to try to answer this question.

Legacy assets occur because of three key features of bank lending.

First, loans are made for multiple periods. It is obvious that if loans are only made for one period at a time, a bank can never be stuck with loans it doesn’t wish to hold given the alternative options. With single period contracts, all the consequences of issuing a loan are realised for both parties at the end of each period. There might be effects of firm or bank losses on their respective balance sheets and therefore on the volume of subsequent credit supply (see Chen (2001), Meh and Moran (2008), Gertler and Kiyotaki (2010)) but there is no lasting effect on the composition of the loan portfolio. Only with multi-period lending can the history of lending matter for the ongoing composition of the loan portfolio.

The second (and related) key feature is that banks must have limited ability

¹See for example US Treasury (2009) or Honahan (2013).

to recall loans. Multi-period lending by itself is not sufficient to generate legacy assets if a bank has the right to recall credit from loans it does not wish to retain or has other means to induce repayment. Legacy assets can only arise if there is a constraint on re-optimisation each period.

The third feature is that economic circumstances must change through time. If banks face exactly the same decision problem every period and make the same choices, then even if there is loan persistence, the consequences will be the same every period and thus an uninteresting problem. By contrast, it is a very non-trivial problem when banks make long term loans subject to aggregate shocks because how they respond to the current shock determines the balance sheet they inherit in subsequent periods which in turn constrains the choices they can make at that point. The loan portfolio is a state variable.

This paper sets out and then solves numerically the dynamic programming problem involved in managing the balance sheet of a single monopoly bank that repeatedly makes multi-period loan commitments with limited loan recall rights in the presence of aggregate shocks. There is a general equilibrium setting and a key constraint for the bank is that it must fund its loans with deposits. The agents in this economy always interact with the bank but have full freedom to choose which side of the balance sheet they wish to be on. So meeting the balance sheet constraint through time requires inducing the right volume of agents to be depositors and borrowers given the aggregate shock and the inherited loan balance sheet.

Of the three key features generating legacy assets, it is reasonable to assume that the aggregate shock process is exogenous. It will be assumed that aggregate shocks follow a time-homogeneous Markov process with positive persistence. It is not reasonable, however, to assume outright that loans are multi-period or that there are restrictions on the ability to recall loans. If these features are to drive the results, they need to be properly micro-founded. This paper builds on the framework of Penalver (2014) which wraps a bank around the outside of a firm dynamics process (following Hopenhayn (1992)). Firms enter and exit endogenously in response to persistent idiosyncratic shocks and thus exist for stochastic lengths of time. But to produce, firms need to borrow capital from a bank which, in turn, receives capital from depositors. Firms are constrained to be of identical size and borrow an identical amount, so the distribution of firms by profitability is isomorphic to the distribution of credit risk for the bank.

The relationship between the bank and its borrowers in that model is shaped by four important assumptions: (1) firms face entry and exit costs (following

Dixit and Pindyck (1994)); (2) it is costly for banks to verify the profitability state of individual loans; (3) entrepreneurs (who run firms or "projects") face limited liability; and (4) it is more costly for banks to liquidate projects than for entrepreneurs.

If there were no entry and exit frictions, then agents would be purely opportunistic in running projects: enter when they are expected to be profitable, and quit when they are not. Regardless of the behaviour of the bank, opportunistic entry and exit always delivers the most productive distribution possible (given the dynamics of firm profitability) and the lowest possible default risk because the indirect competition for capital drives out the least productive firms. With entry and exit frictions, this is no longer the case because some firms will want to continue in circumstances in which a firm would not enter. As a result of these private costs, the distribution of firms is not necessarily the constrained optimal and the bank may have an interest in changing the entry and exit rate of firms through its lending terms.

The limited liability assumption distorts the continuation decision because downside risk is capped. The level of profitability (or in this case losses) at which a borrower will wish to continue is lower in the presence of limited liability than it would be otherwise and this increases the default risk faced by the bank. However, since borrowers compete for capital, the value of the option to default under limited liability simply pushes up the market clearing loan interest rate until it is fully priced in. If both sides assign the same value to the default option, then the bank will be fully compensated for the associated credit risk and both sides would be indifferent to the equilibrium default rate.

The assumption that it is more costly for the bank to liquidate a project than for an entrepreneur is thus crucially important. The additional costs faced by banks in resolving failed projects means that the cost to bank of the default option is higher than the value of the option to the borrower. The consequence of this is that the equilibrium loan interest rate will not fully compensate the bank for the credit risk and this motivates the bank to try to influence the distribution of firms (and thus, loans).

Penalver (2014) assumes that the bank can use two instruments to manage credit risk: loan monitoring and a profitability threshold. The loan terms are assumed to stipulate that if the loan is monitored and found to be below the profitability threshold, then the bank has the right to demand immediate repayment of the loan. The bank, therefore, has three choice variables in the loan contract it offers: the loan interest rate, the monitoring intensity and the level of the prof-

itability threshold. The objective of the bank is to maximise steady state profits subject to the constraint that it must have the same volume of loans as deposits. If the profit-maximising offer has full monitoring and the profit threshold is the same as the entry threshold, then the loans would be observationally equivalent to single period loans. The bank would only retain loans to firms to which it would grant new loans. By contrast, if the profit-maximising monitoring rate is less than 1 or the profitability threshold is below the entry threshold, then existing loans are different from new loans. Penalver (2014) shows that with the assumption of costly state verification, the profit-maximising monitoring intensity can be less than 1 and the profitability threshold below the entry threshold. In other words, that model generates equilibrium behaviour in which banks voluntarily make long term loan commitments with constraints on their ability to recall loans, two of the basic ingredients of the legacy loans problem.

The purpose of this paper is to extend that model by introducing aggregate shocks. In the original model, agents face dynamic problems but the bank faces an invariant one. By a law of large numbers assumption, the individual risk faced by the agents is completely diversified away at the level of the bank's loan portfolio. In the extended model, agents face a common aggregate shock as well as their idiosyncratic shocks. The common shock introduces a correlation in the behaviour of the agents which therefore affects the balance sheet of the bank. The bank faces a complex dynamic programming problem. Its objective is to try to maximise a function of the mean and variance of its profits subject to the balance sheet constraint that loans and deposits are equal through time. Aggregate shocks affect loans and deposits in opposite directions and thus destabilising the balance sheet. But in trying to re-equilibrate the balance sheet, the bank is constrained by its inherited balance sheet and must bear in mind the consequences of its actions on the balance sheet it will take into next period given the probability distribution of shocks it will face then. The bank will be searching for Markov policy functions for loan and deposit interest rates, the monitoring intensity and the minimum profitability threshold.

The model is close in some respects to Arellano, Bai and Zhang (2012), Khan and Thomas (2013) and Khan, Senga and Thomas (2014) who also analyse how financial contracts influence firm dynamics and the resulting misallocation of capital. Those models give borrowers a rich set of choices during their life including how and when to invest and the amount of dividends to issue. These models, however, assume that lending is single-period and loan contracts can be freely and fully updated to reflect changes in the state of borrowers. For tractability,

they also require exogenous exit shocks to constrain the economy from outgrowing credit restrictions. By contrast, in this model entry and exit is entirely endogenous. Moreover, one of the key features of the model is that bankruptcy is only one of the two means of endogenous exit and default arises out of a choice between bankruptcy and voluntary liquidation.

The structure of the paper is as follows. Section 2 sets out the general structure of the model and summarises the solution of the steady-state problem in which there is only idiosyncratic risk. Section 3 explains the solution to the dynamic programming problem in the presence of aggregate risk. Section 4 analyses the credit cycle derived from an illustrative calibration of the model by describing the dynamics of loan interest rates, deposit interest rates, loan covenants, monitoring intensity and the default rate. Section 5 concludes.

2. Model set-up

This section starts by setting out the dynamic model in the most general terms and then summarises the equilibrium in steady-state.

The economy, following Penalver (2014), is populated by a measure 1 of infinitely small *ex ante* homogeneous agents. Agents are risk neutral, live forever and discount the future at rate β . Time is discrete.

Economic activity takes place through "projects" which can last indefinitely and deliver a gross stochastic idiosyncratic return per period $q(a)$ where $a \in A$ is a profitability index over the compact support $[0, 1]$. Some projects are more profitable than others at any point in time reflecting, for example, time varying differences in market power, productivity, managerial competence etc, the deeper sources of which are unmodeled. A project's profitability index evolves according to a time homogeneous Markov process represented by the cumulative distribution function $F(a', a)$. This process is assumed to have the following properties:

- A** (i) $F(a', a)$ is continuous; (ii) profitability shocks are persistent, 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 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 level in a finite number of periods.²

²The numerical example uses an AR(1) process with Gaussian shocks approximated over a

Since there are exit thresholds, this assumption implies that all projects will almost surely close at some future point.

At any point in time, some of the agents are "inventors". Similar to labour search models such as McCall (1970), inventors receive one idea per period with profitability a drawn from an independent and identical distribution, $G(a)$. If an inventor decides to commence a project, she pays a start up cost S and enters the following period with gross idiosyncratic return $q(a')$ according to $F(a', a)$.³ Projects are assumed to require 2 units of capital. Entrepreneurs cannot get credit directly from inventors (perhaps for the reasons described by Diamond (1984)) but can borrow up to 1 unit from the bank (which is assumed to be binding).⁴ If an inventor decides not to enter, she remains an inventor next period and in the absence of any better alternative, deposits her savings at the bank on which she receives a deposit rate τ .

The other group of agents are running projects and are called "entrepreneurs". Entrepreneurs make net profits per period $q(a) + z - \rho(\cdot)$ where $\rho(\cdot) \in \mathbb{R}^+$ is the loan interest rate and z is the aggregate shock process to be described below. Only entrepreneurs can costlessly observe the profitability state of the project and the bank will need to pay m per loan to get a perfectly accurate report. The gross idiosyncratic profitability function $q(a)$ is assumed to be continuous and strictly increasing in a .

Bearing in mind their current and expected future profits, entrepreneurs decide whether to continue in production next period or exit and switch to being an inventor. Entrepreneurs can exit in two ways:

- "Orderly" exit occurs if an entrepreneur absorbs current period payoffs, $q(a) + z - \rho(\cdot)$, (in this situation, losses) 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.
- "Default" occurs if an entrepreneur files for bankruptcy protection in which case current period losses are excused (including repayment of loan interest)

finite dimension grid using the method of Tauchen (86).

³Caballero and Hammour (1994) estimate entry costs at around half a year's operating costs.

⁴This is not a particularly satisfying assumption. But one could assume that there is a collateral constraint in one direction and in the other direction agents prefer to borrow rather than invest more of their own funds to diversify risk.

but the agent pays an exogenous bankruptcy cost B .⁵ Naturally, L and B are calibrated so that bankruptcy is preferred over orderly exit only in extreme circumstances.

There is no capital accumulation in this model. Each agent has one unit of capital throughout their lives and the only choice that they are making is whether to use it to run a project or to store it at the bank and thus indirectly allow other agents to use it. Capital allocation decisions are inseparable from the choice whether to be an entrepreneur or an inventor and the latter choice is naturally influenced by the rates of return on capital available in the two cases. By ruling out capital accumulation, the model closes down the intensive margin over the supply and demand for bank funds, leaving the extensive margin to do all the work. The volume of deposits is equal to the measure of inventors and the volume of loans equal to the measure of entrepreneurs. Meeting the balance sheet constraint for the bank naturally requires, therefore, that there is an equal measure of inventors as entrepreneurs and the bank will have to set its policy rules so that this is satisfied.

Finally, 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 does influence any of the incentives and does not feature in any of the value functions used to calculate agent choices.

Capital/savings are 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, idiosyncratic risks have no aggregate stochastic effects. The bank offers the following deposit and loan contracts.

- Deposits earn the current deposit interest rate $\tau(\cdot)$. Deposits can be withdrawn at the end of any period.
- The loan contract specifies the terms and conditions which apply in the current period. The terms and conditions are the bank policy functions for the loan interest rate $\rho(\cdot)$, a monitoring intensity $\varphi(\cdot)$ (where $0 < \varphi < 1$) and a covenant specifying a minimum net profit level ζ . The loan is provided for the duration of the indefinite project but both parties have an option to terminate it each period. Each borrower has the option to repay the loan

⁵To simplify subsequent calculations, it is assumed that entry, exit and bankruptcy costs are paid in the following period.

if he decides to exit production and the bank can demand full repayment if it discovers that the covenant condition has been breached. This demand is enforceable and by the endowment assumption, an entrepreneur has the ability to repay.

The model is solved to find the optimal parameters for this class of loan contract. The contract terms will vary over time in response to the aggregate state variables but are applied uniformly across all borrowers regardless of their privately known state or the age of the project. Optimal dynamic contracting models with costly state verification, such as Monnet and Quintin (2005) and Clementi and Hopenhayn (2006), ask agents to report project payoffs and use complex message-dependent repayment schedules and auditing probabilities. In principle, such contracts are superior to the simpler structure considered here, although Boyd and Smith (1994) show that the welfare losses from being constrained to offer standard debt contracts is very small.⁶ A simpler debt contract structure is also much more tractable in a recursive framework with a portfolio heterogeneous loans.

Since the bank uses the covenant to protect its interests, it follows that ζ must in all cases imply a trigger value of a at least as high as that at which entrepreneurs voluntarily exit or else the covenant would be redundant. Therefore the entrepreneur faces a utility loss from having his loan recalled and he cannot be expected to reveal the profitability state of the loan voluntarily. So to discover the state and enforce the covenant, the bank must monitor its continuing loan portfolio. The monitoring intensity $\varphi(\cdot)$ gives the probability that any entrepreneur who chooses to remain in production is inspected.

The model in Penalver (2014) contains only idiosyncratic shocks and the problem for the bank is to find a loan interest rate, monitoring intensity and covenant threshold triplet (ρ, φ, ζ) which maximises profits subject to the constraint that the bank must have the same measure of deposits as loans. The equilibrium of that model is an invariant distribution with constant values for all the endogenous variables, which is why it is referred to as the steady state version of the more general model. So with constant values it does not matter in a sense whether the equilibrium loan terms $(\bar{\rho}, \bar{\varphi}, \bar{\zeta})$ are stipulated in the contract or are a rational expectation. In the model in this paper, there is an aggregate shock process so constant interest rates will no longer satisfy the balance sheet constraint.

⁶Boyd and Smith (1994) is a single period model.

There are 7 elements in the aggregate shock set, Z . They are symmetrically spaced with time homogeneous Markov transition probabilities $Z(z', z)$ which approximate an AR(1) process with normal errors using a Tauchen matrix (Tauchen (1986)). The aggregate shock is a common scalar shift in the gross return of all projects in that period. Gross output for an individual firm is thus $q(a) + z$.

To simplify notation, define $\Omega \equiv \{z, \theta_{-1}\}$ where θ_{-1} is an endogenous state variable at the end of the previous period. Exactly what is in θ is delayed until Section 3 when more of the structure of the model is in place. It is assumed that the endogenous state variable is accurately observed by the bank and the agents but only after they have taken the decisions that determine it. By contrast, the bank and the agents are assumed to be able to observe the aggregate shock before taking decisions. Ω is thus the information set on which everyone conditions their choices each period and gives the model a Markovian structure. How the bank adjusts the loan terms and how expectations are formed are fundamentally important for the equilibrium.

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) with knowledge of θ_{-1} . The aggregate shock z is revealed. The inventors get an idea from $G(a)$ and entrepreneurs get a private update of their idiosyncratic profitability state according to $F(a', a)$.
2. The bank updates the loan interest rate, the monitoring intensity and the loan covenant threshold which in this recursive setting are functions of the information set. The loan interest rate is updated according to $\rho(\Omega)$, the deposit interest rate is updated according to $\tau(\Omega)$ and the monitoring intensity according to $\varphi(\Omega)$.
3. 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 inform the bank that they will repay their loan. Inventors receive deposit interest and decide whether to enter production next period based on their profitability draw.
4. The bank monitors ongoing loans at the stochastic rate $\varphi(\Omega)$ and recalls the loans of all entrepreneurs found below the covenant profitability threshold ζ .

5. The bank receives deposits and makes additional loans to entering entrepreneurs. θ is revealed.

Before diving into the details, which become quite complex in appearance very quickly, it is worth at this stage giving an overview of the solution.

The first step is determine what is optimal for individual entrepreneurs and inventors to do given their profitability draws and what they know about the two aggregate states, z and θ_{-1} . Since these entry and exit decisions are forward looking, they will depend on the expectations entrepreneurs and inventors have about how their individual circumstances and the aggregate states will evolve. The exogenous aggregate state is discrete and the endogenous aggregate state is approximated at discrete points, so the aggregate state (and thus the information set, Ω) is a 2-dimensional grid. The economy will pass stochastically through these grid points. The agents, therefore, are working out the probabilities of passing from the current grid point to another.

The second step is to consider the problem faced by the bank. For any given shock to aggregate profitability, the effect on the loan portfolio will depend on the distribution of firms (and loans) over the idiosyncratic state. But this distribution depends on the history of entry and exit decisions and thus the history of aggregate shocks. These portfolio dynamics will affect the size of the loan portfolio and its profitability over time through default risk and loss given default. The objective of the bank is to maximise average profits, minimise the variance of profits whilst keeping the volume of loans and deposits as close to each other as possible.⁷ The dynamic programming problem for the bank is to set the deposit interest rate, the loan interest rate, the monitoring intensity and the covenant threshold to bring the demand for loans into line with the volume of deposits but bearing in mind that these loan terms and the resulting distribution of entrepreneurs will persist and affect the problem faced in future periods. Since the model has a Markovian structure, the solution will be state-contingent policy rules. The initial conditions, the agents actions, the bank's policy functions and the aggregate shock process will determine the evolution of the economy through time. Finally, since individual

⁷These three objectives cannot, of course, be separately maximised. They are jointly summarised in one continuous objective function with implicit trade-offs between the three elements. Nothing of any importance depends on the precise form of the objective function.

agents are assumed to be making decisions based on their state-contingent rational expectations, their beliefs will have to be fulfilled on average.

The third step of the solution is to recognise that agents' choices and expectations and the bank's policy rules are an enormous fixed point problem. Solving the model computationally involves an algorithm that finds agents' choices and expectations that are mutually consistent with a given set of bank policy rules and then searches over different bank rules until the objective function is maximised.

The following subsections explain the optimal behaviour of the entrepreneurs and inventors and the resulting portfolio dynamics for any specified bank policy rules.

2.1. Entrepreneurs

Depending on the idiosyncratic profitability state, a , an entrepreneur chooses between default, orderly exit and continuation. If he decides to default to escape losses, he pays B next period and switches to being an inventor. The discounted value of defaulting given information set Ω is thus

$$V_B(\Omega) = \beta\{E[V_I(a', \Omega'; \cdot) | \Omega] - B\} \quad (1)$$

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

If he chooses orderly exit from production, he absorbs current losses, pays liquidation costs L next period and also enters next period as an inventor. The value of orderly exit in state a given Ω and the bank interest rate rules is

$$V_X(a, \Omega) = q(a) + z - \rho(\Omega) + \beta\{E[V_I(a', \Omega'; \cdot) | \Omega] - L\} \quad (2)$$

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

$$V_C(a, \Omega) = q(a) + z - \rho(\Omega) + \beta\{E[V_E(a', \Omega') | a, \Omega]\} \quad (3)$$

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

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

$$V_E(a, \Omega; V_I) = \max\{V_B(\Omega), V_X(a, \Omega), V_C(a, \Omega)\} \quad (4)$$

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. It is straightforward to see from equations (1) and (2) that entrepreneurs will default for all values of $a < a_\delta(\Omega)$ where

$$q(a_\delta(\Omega)) + z - \rho(\Omega) = \beta(L - B) \quad (5)$$

thus defines the state-contingent default threshold.

The threshold for orderly exit, $a_X(\Omega)$, which also depends on the information set, results from a comparison of V_X and V_C . The only tricky aspect of this problem is the conditional expected value of being an entrepreneur next period, $E[V_E(a', \Omega'; V_I') | a, \Omega]$. Consider first an entrepreneur with profitability above the state-contingent loan covenant threshold, $a \geq \zeta$. In this case the entrepreneur faces no risk if the bank randomly chooses to monitor his loan, so we can ignore the role of the bank and

$$E[V_E(a', \Omega'; \cdot) | a \geq \zeta, \Omega] = \int_A \int_{\Omega'} V_E(a', \Omega'; \cdot) J(d\Omega', \Omega) F(da', a)$$

where $J(\Omega', \Omega)$ describes the evolution of the state variables and subsumes the stochastic process of z and the endogenous transition of θ . The calculation is more complex for an entrepreneur with $a_\delta(\Omega) < a < \zeta$. In this case, if the entrepreneur decides to continue and escapes monitoring (with probability $(1 - \varphi(\Omega))$), then he 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 he involuntarily reverts to being an inventor. Therefore for $a_\delta(\Omega) < a < \zeta$

$$\begin{aligned} E[V_E(a', \Omega'; \cdot) | a_\delta(\Omega) < a < \zeta] = \\ (1 - \varphi(\Omega)) \int_A \int_{\Omega'} V_E(a', \Omega'; \cdot) J(d\Omega', \Omega) F(da', a) \\ + \varphi(\Omega) (E[V_I(a', \Omega'; \cdot) | \Omega] - L) \end{aligned}$$

The voluntary exit threshold $a_X(\Omega)$ is the value of current period profitability at which an entrepreneur is indifferent between continuing or exiting voluntarily. Some simple cancelling defines $a_X(\Omega)$ as

$$\int_A \int_{\Omega'} V_E(a', \Omega'; \cdot) J(d\Omega', \Omega) F(da', a_X(\Omega)) = \int_A \int_{\Omega'} V_I(a', \rho(\Omega'); \cdot) J(d\Omega', \Omega) G(da') - L \quad (6)$$

These equations look highly complex but are actually intuitively quite simple. Each period the agent has three options - continue, exit in an orderly fashion or default - and the bank has the option of demanding repayment if the project is known to breach the covenant. These four options partition the profitability space A into four regions based on three thresholds - the default threshold $a_\delta(\Omega)$, the voluntary exit threshold $a_X(\Omega)$ and the covenant threshold ζ . The thresholds are determined by the points at which a rational, forward-looking agent is indifferent between two naturally adjacent options. The thresholds vary with the information set because this conditions the bank's policy rules. Note that the policy rules affect the value functions of inventors as well as entrepreneurs because each is a function of the other.⁸ So changes in Ω have quite subtle effects on the various thresholds because the inside and outside options are both changing at once.

2.2. Inventors

Each period, an inventor receives interest on her deposit $\tau(\Omega)$ and an idea with profitability index a . The inventor chooses between paying S and entering production next period with profitability index a' given $F(da', a)$ and the loan terms prevailing at that point or waiting for another draw from $G(a)$ next period. The value of being an inventor is conditioned on the deposit rate and the expected loan terms available on subsequent entry, which given the assumed policy rule, are functions of Ω . The value function of an inventor is consequently

$$V_I(a, \Omega; V_E) = \tau(\Omega) + \beta \max \left\{ \begin{array}{l} \int_A \int_{\Omega'} V_E(a', \Omega'; \cdot) J(d\Omega', \Omega) F(da', a) - S, \\ \int_A \int_{\Omega'} V_I(a', \rho(\Omega'); \cdot) J(d\Omega', \Omega) G(da') \end{array} \right\}$$

⁸Given that there is switching between being an entrepreneur and an inventor and vice versa, the value function of each is a function of the other. This interdependency is recorded in the initial definitions but left implicit in all subsequent notation. The proof of uniquely consistent value functions is in the appendix of Penalver (2014).

giving an entry threshold $a_E(\Omega)$ defined by

$$\int_A \int_{\Omega'} V_E(a', \Omega'; \cdot) J(d\Omega', \Omega) F(da', a_E(\Omega)) - S = \int_A \int_{\Omega'} V_I(a', \rho(\Omega'); \cdot) J(d\Omega', \Omega) G(da')$$

2.3. Equilibrium

Define $H([0, a])$ as the measure of entrepreneurs at the *end* of each period with profitability index in the interval $[0, a)$. And to simplify notation define $FH(a') = \int_A F(a', a) H(da; \cdot)$, which is the measure of firms entering the interval $[0, a')$ given the distribution over a in the previous period.

With the behavioural assumptions of the model and denoting D as the measure of inventors with their funds on deposit at the start of each period, a law of motion for the distribution of entrepreneurs can be defined by:

$$H'([0, a']; \Omega, \cdot) = D \int_{a_E(\Omega)}^{a'} G(a) + \int_{a_X(\Omega)}^{a'} FH(da') - \varphi(\Omega) \int_{a_X(\Omega)}^{a' < \zeta} FH(da') \quad (7)$$

The first term measures how many agents enter at profitability levels below a' given state Ω . The middle term measures how many entrepreneurs evolve into a profitability state above the voluntary continuation threshold. The third term eliminates those entrepreneurs closed down by the bank because they are monitored and have their loan recalled. Defaulting entrepreneurs are implicitly removed because they are already below the lower truncation of the distribution at $a_X(\Omega)$. Since each entrepreneur borrows one unit of capital, $H(A; \cdot)$ is the measure of the volume of loans outstanding at the end of each period.

The following expression describes the bank's profits each period:

$$\Pi(\Omega) = \rho(\Omega) \int_{a_\delta(\Omega)}^1 FH(da') - \int_0^{a_\delta(\Omega)} \lambda(a', \Omega) FH(da') - \varphi(\Omega) m \int_{a_X(\Omega)}^1 FH(da') - \tau(\Omega) S(z)$$

The first term is the loan interest paid by non-defaulting borrowers. This depends on how many firms evolve into a profitability state above the default threshold and the loan interest rate $\rho(\Omega)$. The second term deducts state-contingent loss given default assuming that the bank closing the project is inefficient (determined by

the function $\lambda(a', \Omega)$). The third term subtracts the costs incurred in monitoring the continuing firms at a rate $\varphi(\Omega)$. The final term deducts the deposit interest paid and thus depends on the deposit rate $\tau(\Omega)$.

The invariant model in Penalver (2014) is nested within the current framework by setting the exogenous aggregate shock to zero and assuming the deposit rate is fixed exogenously at $\bar{\tau}$. An invariant, or steady-state, distribution is naturally one in which

$$H'(A, \bar{\rho}) = H(A, \bar{\rho}) = \bar{H}(A, \bar{\rho})$$

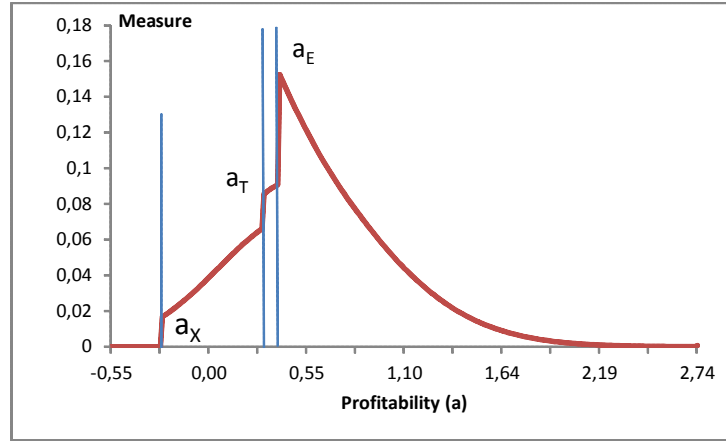
Propositions 1 and 2 of Penalver (2014) state that for a given monitoring intensity $\bar{\varphi}$ and covenant threshold $\bar{\zeta}$, there is a unique loan interest rate $\bar{\rho}$ that delivers an invariant distribution satisfying the balance sheet constraint

$$\bar{H}(A, \bar{\rho}) = D = \frac{1}{2}$$

Since there is a measure 1 of agents each with a unit of capital, the balance sheet constraint requires half the agents to be depositors and half the agents to be entrepreneurs at the end of each period. In the invariant version of the model, there is no need to condition on state variables and the entry and exit thresholds are also constant. An example of an invariant distribution of projects along the profitability index a is illustrated in Figure 1. 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 there are 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 monitor 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.

Figure 2 illustrates the one period transition of the distribution in Figure 1 with the invariant distribution overlaid. (This illustrates the expression $FH(a)$.) 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 - the distribution melts to

Figure 1: Invariant distribution of firms on the bank's balance sheet



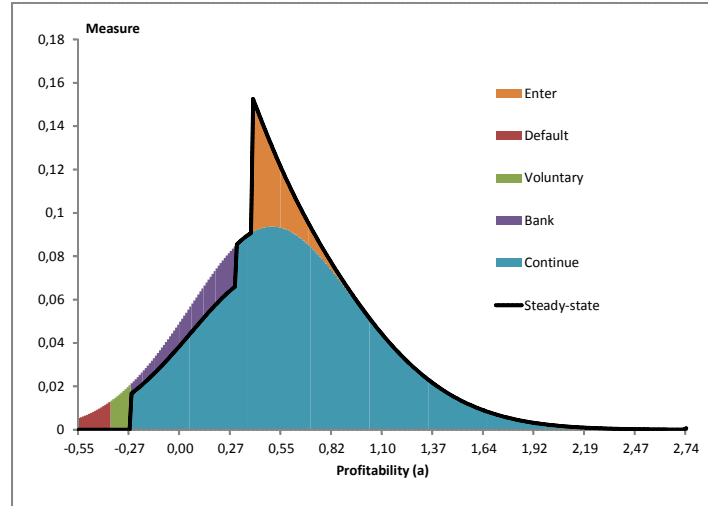
the left. The distribution is refreshed by the entry of new entrepreneurs clustered above the entry threshold. Moving further to the left, a number of entrepreneurs fall below the threshold a_T but above a_X . These are the entrepreneurs that want to continue but are at risk of having their loans recalled 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.

3. Solving the bank's dynamic problem

The solution to the bank's problem with only idiosyncratic risk was a fixed vector of loan terms. From the perspective of the bank, everything about that problem was deterministic and hence why it is the steady state of the model with aggregate shocks. With unchanging loan terms, the only thing that the agents had to forecast was their future individual state. Everything changes with an aggregate shock. The bank's loan portfolio is now stochastic and history dependent; the bank's behaviour is contingent on the aggregate states; and everyone will have to forecast what the aggregate states will be as well.

The exogenous shock process z is straightforward to forecast because it is a

Figure 2: One period transition of bank balance sheet



known AR(1) process with normal errors. The distribution of $z'|z$ in a discrete state space is just the row of a matrix.

The endogenous aggregate state variable θ is much harder. Equation (7) shows that the transition of the balance sheet is a complex function of the distribution over idiosyncratic profitability states. Part of this transition is under the control of the bank through its current choices of loan terms but the distribution is also affected by the history of past decisions. So in principle, θ should correspond to H . However, since A is a compact set of real numbers, H is an infinite dimensional object. As in Krusell and Smith (1998), to get traction with this problem, it is necessary to approximate H with a suitable summary statistic which in this case will simply be $H(A)$, ie the volume of loans outstanding. The balance sheet condition that the bank have the same measure of loans as deposits requires $\tilde{H}(A) = 1$, so $H(A)$ will also be a measure of balance sheet disequilibrium. For this reason $\theta \equiv H(A)$ will often be referred to as "excess loans".

It was noted earlier that the information set Ω contains z and θ_{-1} so agents know the current exogenous aggregate state but only know the level of excess loans from the period before. In other words, the agents and the bank do not know the aggregate endogenous state that will arise out of the decisions that they

are currently taking. This is not because agents are not fully rational. Indeed, they will be making accurate forecasts of what the endogenous state will be given the information that they have. In principle the bank ought to be able to rely on the monitoring process from the previous period and the same law of large numbers assumption plus knowledge of the entry process to estimate precisely the full distribution. But the spirit of the exercise is that agents and the bank do not have complete real time cross-sectional information.⁹ Monitoring is explicitly costly and it is not difficult to assume that information processing costs are high too.

Agents are assumed to make forecasts for the aggregate state vector as a function of current observables,

$$E[\Omega' | \Omega] = \sigma(\Omega)$$

It will be further assumed that the forecasting function for the endogenous aggregate state and the bank policy rules are all linear. As a result, the expected loan interest rate next period, for example, will be a linear function of the information set:

$$E[\rho' | \Omega] = E[\rho(\Omega') | \Omega] = \rho(\sigma(\Omega))$$

with a similar structure for all the other bank policy rules. All agents have the same information set and all are rationally forward-looking, so all agents will have the same forecasts of future states and thus the same expectations about the path of loan interest rates and deposit rates. As a result, all agents will have a common view on the expected value of being an inventor and an entrepreneur.

For any given set of bank rules and forecasting function (whether accurate or not), we can solve for the behaviour of the agents which is completely summarised by the entry and exit thresholds. And for any given distribution of firms (and loans) from the previous period, we can solve the transition equation for the

⁹One could consider the branch network of a commercial bank. The branch has the detailed information about borrowers and this is only passed to the centre in summary form. The branch also has a relatively stochastic demand for loans and cannot extrapolate from local conditions to the overall demand for loans across the network. The central treasury of the bank is the location at which all funding needs of the branch network are consolidated and only after the branch network has committed to loans can the central treasury reset loan and deposit rates in response to balance sheet disequilibrium.

distribution. Thus for a given starting distribution and a sequence of aggregate shocks $\{z_t\}$, there is a unique path for the evolution of the bank's balance sheet and its profits. Clearly, for arbitrary bank policy functions one side or the other of the implied balance sheet could grow until it absorbed the full measure of capital and this would not be a feasible equilibrium. So the only acceptable rules are those that deliver a dynamically balanced measure of loans and deposits (at least to satisfactory precision). But there may be more than one (indeed infinite) set of rules that satisfy this dynamic balance sheet constraint. Different rules, though, imply different paths for profits with associated means and variances. It will be assumed that the bank seeks to maximise mean profits, minimise the variance of profits subject to minimising the deviation from the balance sheet constraint.¹⁰ These multiple objectives are captured in a single objective function.

So the ultimate solution to the model is a fixed point in which:

- the bank follows linear policy rules which, given the behaviour of depositors and borrowers, and the characteristics of the common shock process, maximises the objective function;
- agents' expectations are model consistent; and
- depositors and borrowers act optimally given their heterogeneous circumstances, their expectations about the future paths of their private states and the aggregate state vector, and the policy rules of the bank.

The solution is found using numerical methods and the technical details are relegated to the appendix. The qualitative properties of the solution are general and are not parameter dependent. A sketch of the solution algorithm is as follows:

- Step 1: Guess parameters for the linear bank policy rules as functions of the information set, Ω . The rules are parameterised by the steady state loan rate and deviations in the state variables from their steady state levels. The rules for the loan interest rate, and the monitoring rate are respectively:

$$\rho = \bar{\rho} + \rho_1 z + \rho_2 (\theta_{-1} - \bar{\theta})$$

¹⁰Since balance sheet constraints are binding, one could assume that the bank holds a small buffer of liquid assets such as government bonds which pay the same as the deposit rate. The balance sheet constraint is met exactly by buying or selling this buffer stock. So the objective function of the bank is to keep this buffer stock as small as possible.

$$\tau = \bar{\tau} + \tau_1 z + \tau_2 (\theta_{-1} - \bar{\theta})$$

$$\varphi = \bar{\varphi} + \varphi_1 z + \varphi_2 (\theta_{-1} - \bar{\theta})$$

- Step 2: Guess parameters for the linear forecasting rule for the agents for the endogenous variable θ as a function of Ω .

$$E[\theta] = \sigma_1 z + \sigma_2 (\theta_{-1} - \bar{\theta})$$

- Step 3: Form a grid of a , z and θ . Calculate the value functions for inventors and entrepreneurs conditional on a and Ω and the forecast rule from Step 2 and the bank policy rules from Step 1. On the basis of these value functions, find the entry and exit thresholds for each grid point of Ω for each cohort.
- Step 4: Simulate the economy for 10,000 periods based on these rules, throw away the first 1,000 observations and estimate a new forecasting rule using ordinary least squares.
- Step 5: Repeat Steps 2 to 4 until the revision to the forecasting rule is reasonably small.
- Step 6: Calculate the loss function over the estimation range and repeat Steps 1 to 5 with a different bank policy rules using a search algorithm until the objective function is maximised.

Once this sequence of loops is completed, the solution is the profit-maximising fixed point of the problem. Given the number of steps in the algorithm, the two-dimensional aggregate state vector, the need to have a very granular idiosyncratic space so that changes in the rules result in changes in agents' behaviour (or else there are large flat regions over the minimisation surface) and the need to keep the run time to a reasonable length, many of the approximations are crude and the grid for the endogenous state variable is small (15 nodes). After all the number crunching, the optimal fixed interest rate rule for a simulation of the model is

$$\rho = 0.070 + 0.41z + 0.11(\theta_{-1} - \bar{\theta}) \tag{8}$$

the deposit rule is

$$\tau = 0.035 + 0.65z + 0.09(\theta_{-1} - \bar{\theta}) \tag{9}$$

and the monitoring intensity rule is

$$\varphi = 0.26 - 0.28z - 0.21(\theta_{-1} - \bar{\theta}) \quad (10)$$

In each case the first term is the value in the steady-state model.

The third term in each case captures the optimal response to excess loans the previous period. In each policy rule this is a positive value as might be expected. Since the loan distribution is slow moving because of its history dependence, excess loans the previous period will likely imply excess loans in the current period without any bank response. Since part of the objective is to try to keep excess loans to a minimum, it is sensible to lean against it with the instruments available. That all co-efficients on the third terms are positive indicates that the optimal response is to use all instruments together. *Ceteris paribus*, raising the loan interest rate, raising the deposit rate, and increasing the monitoring intensity will all dampen the demand for loans which, since those not borrowing are lending, automatically increases the supply of deposits. But there is also a relationship between the instruments. Increasing the loan interest rate will also increase the default rate for any given z , so it is also prudent for the bank to monitor more intensely as a precaution.

The second term in each case captures the optimal response to the aggregate shock. Naturally the loan interest rate increases when the aggregate shock is positive in anticipation of increased loan demand. But notice that it does not increase one-for-one which would neutralise the effect of the aggregate shock on profits. In other words, the bank does not fully insure firms against the aggregate shock, which in turn implies that firm profits are pro-cyclical. As a direct consequence, default risk is also counter-cyclical, *ceteris paribus*, and thus the bank can afford to reduce the monitoring rate and loosen the covenant threshold. This relaxation in credit standards is optimal partly because it economises on costly monitoring when this is relatively unimportant but also because it allows the bank to increase the loan interest rate by more than it otherwise would be able to do. Positive times are a good opportunity to trade off credit standards for credit spread. The reason why the bank increases loan interest rates less than one-for-one is because the deposit rate also moves. In fact, it moves more than the loan interest rate, so credit spreads are counter-cyclical.¹¹

¹¹This aggressive use of the deposit rate occurs entirely because of the objective of the bank to reduce the *variance* in profits. If this variance objective is turned off, so that the bank only cares about balance sheet constrained profit maximisation, then the optimal policy is to *reduce* the

The third term indicates that excess loans are expected to be smaller but the same sign as in the previous period, so they stabilise gradually. The second term indicates that excess loans are *negatively* related to the aggregate shock. This term is the net effect of two opposing forces. On the one hand, a positive aggregate shock is on average positive news for entrepreneurs and increases loan demand. On the other hand, savings are assumed to be pro-cyclical. The negative co-efficient indicates that the second effect dominates the former.

The expectations function is

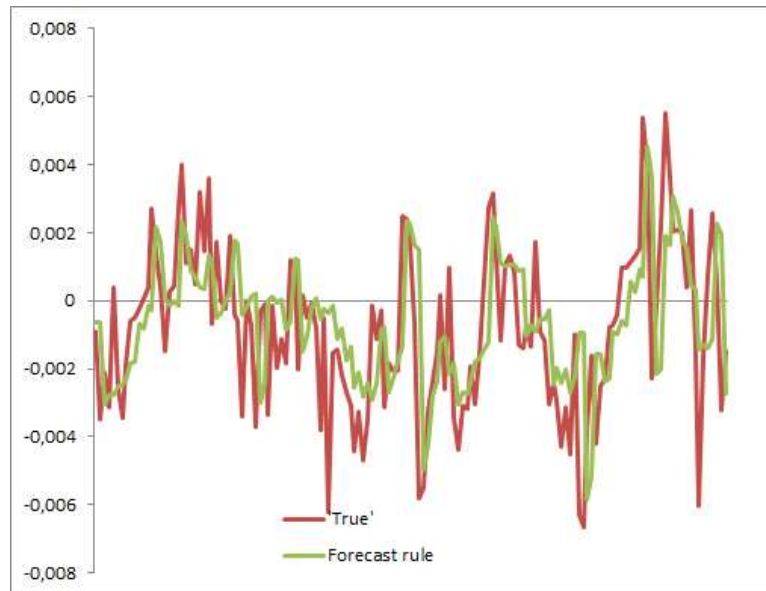
$$E\theta_t = \bar{\theta} - 0.06z + 0.81(\theta_{-1} - \bar{\theta}) \quad (11)$$

The third term indicates that excess loans are expected to be smaller but the same sign as in the previous period, so they stabilise gradually. The second term indicates that excess loans are *negatively* related to the aggregate shock. This term is the net effect of two opposing forces. On the one hand, a positive aggregate shock is on average positive news for entrepreneurs and increases loan demand. On the other hand, savings are assumed to be pro-cyclical. The negative co-efficient indicates that the second effect dominates the former.

The objective of the bank in this model is to maximise profits whilst minimising deviations between the volume of loans and the volume of deposits. An important question, therefore, is how successful is the bank in achieving its balance sheet objective. It is useful to consider as a benchmark what happens if the bank does *not* vary the loan interest rate, the deposit interest rate, the monitoring intensity or the loan covenant threshold over time but keeps them at their steady state values. In this case, the relative size of the two sides of the balance sheet moves around over time but, importantly, it does *not* exhibit any systematic drift. In other words, the balance sheet condition is met on average even with completely unresponsive rules. The variance, however, is quite wide at around 4.7%. So if it is optimal to have active policy rules, it must be because it reduces this variance. And this is indeed the case, with state-contingent rules lowering the variance to around 0.2%.

deposit rate during positive shocks and increase the loan interest rate by more than one-for-one. The logic behind this effect is that loss given default is procyclical, so a profit maximising bank would attempt to orchestrate a pro-cyclical default rate. Since this is clearly not a plausible outcome - deposit rates and default rates are procyclical - and profit smoothing is a reasonable objective, this is the equilibrium presented.

Figure 3: Forecast accuracy plot



Before turning to a discussion of the dynamics, it is important to consider briefly the accuracy of the model. As is well known, a draw-back of heterogeneous agent models solved using numerical methods is that it is impossible to know (or bound) the true solution and therefore impossible to gauge the accuracy of the approximation. Den Haan (2010) argues that using the R^2 of the forecast equation, one of the options proposed by Krusell and Smith (1998), is neither a necessary nor sufficient statistic for measuring accuracy, primarily because the true value is used as the updated lagged dependent variable each period. A high R^2 therefore does not give any indication whether long range forecasts are accurate. Den Haan (2010) instead suggests comparing iterations of the original forecast equation with the model output using a new draw of aggregate shocks. A plot of this iteration against the true value can not only tell whether the forecast is accurate on average but also whether there are systematic errors. Figure 3 illustrates a representative segment of the iteration of the forecast equation and the model output ('True').¹² The True value is highly volatile so it is unsurprising that the forecast rule cannot

¹²The forecast equation actually uses deviations from the steady-state value as the dependent variable.

completely track it. Indeed, there is evidently a lag between the forecast rule and the true value. But overall, this is a highly encouraging fit.

4. Aggregate dynamics - fixed capital case

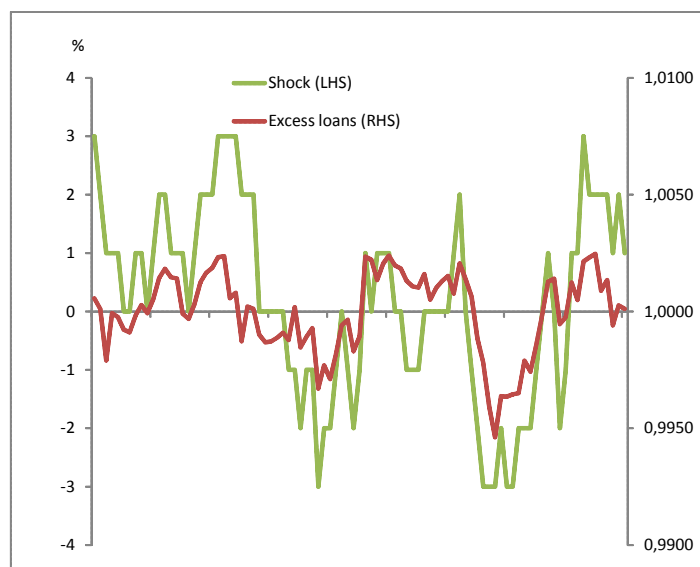
This section describes the behaviour of the key credit cycle variables over a (common) representative sample of the simulation.¹³ The paths of the control variables are, of course, simply a function of the aggregate state variables from the policy functions above. In the actual simulation, the endogenous variable is only represented at grid points and so there are only a fixed number of values for each of the control variables. But the simulation always keeps track of the actual value of θ . The Figures below record the path of the control variables implied by the policy rules and the actual value of θ .

Figure 4 illustrates the path of the two aggregate state variables. The endogenous state is not smooth, reflecting granularity in the state space. The range is also quite narrow, between plus and minus 1.5% of balance, reflecting the success of the policy rules in stabilising the balance sheet. The apparent pro-cyclicality, though, is an illusion. The aggregate shock process is mean reverting so the optimal response incorporates an average expectation of a convergence back to zero. A repeat of the same level of a positive shock (or a higher one) thus amounts to a positive shock relative to expectations. A simulated path looks like a cycle ‘ex post’ but in ‘real time’ is a sequence of surprises. The errors in meeting the balance sheet constraint are a reflection of the surprises, leading to an apparent pro-cyclicality in the endogenous aggregate state variable. It is also worth observing at this point that the endogenous state variable exhibits very low-frequency variation as well depending on whether the recent history of aggregate shocks has been on average above or below the mean. It can be seen from the chart that excess loans is on average negative over this particular segment of the simulation.

Having established the paths of the aggregate state variables, we can now analyse the behaviour of the control variables. Figure 5 plots the path of the loan interest rate over the cycle which is clearly procyclical and also moves by less than

¹³All data in this section have been smoothed using Henderson trend weights to remove high frequency noise in the time series due to various approximations in the numerical simulation. Nothing of substance is hidden by this smoothing.

Figure 4: Simulated aggregate state variables



one-for-one with the exogenous shock.¹⁴ This, of course, was already clear from the parameters of the policy rules.

Figure 6 decomposes the variation of loan interest rates from their steady-state value into the contribution of the two aggregate state variables. Given the positive co-efficients on both variables in the bank loan interest rate setting rule, equation (8) and that the two aggregate state variables move in the same direction, both state variables contribute to the pro-cyclicality of the loan interest rate. But the contribution from the endogenous state variable is negligible, a pattern that occurs with all the other control variables.

Figure 7 plots all of the control variables over the simulation range. Loan interest rates and deposit interest rates are procyclical but monitoring intensity is counter-cyclical.

Finally, Figure 8 illustrates the default rate. The default rate is clearly counter-cyclical as one would expect. Since the loan interest rate moves by less than one-

¹⁴All the remaining charts present data that have been smoothed using a weighted moving average filter.

Figure 5: Loan interest rates and the exogenous shock

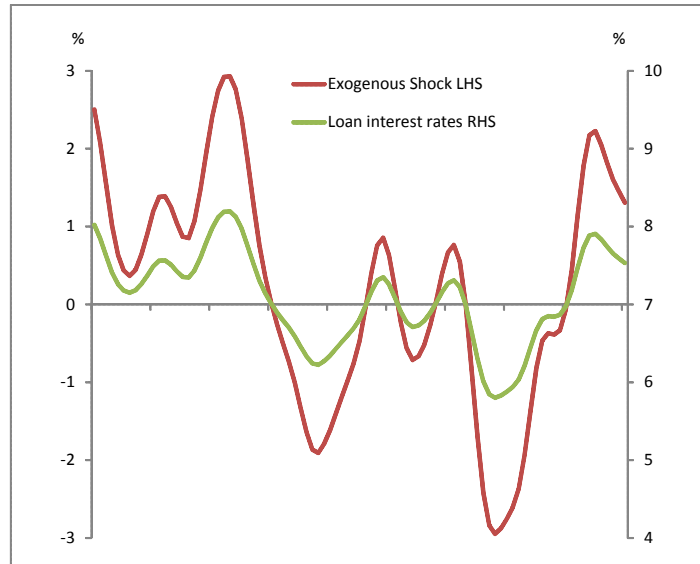


Figure 6: Contribution of aggregate state variables to the loan interest rate

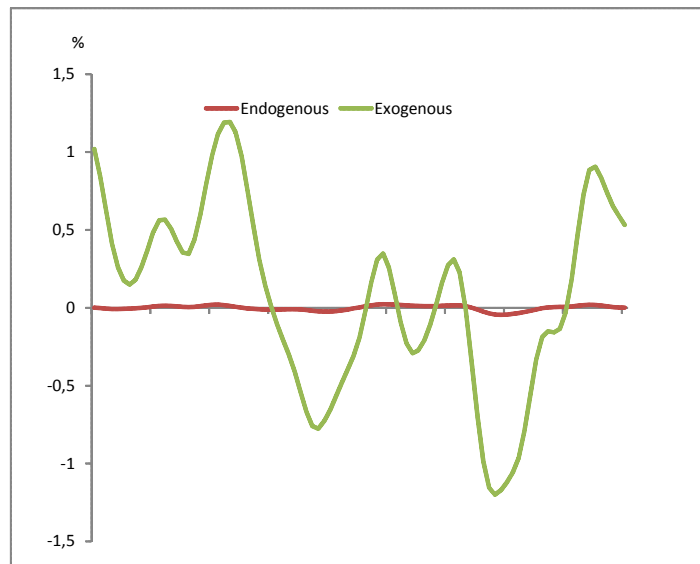


Figure 7: Control variables

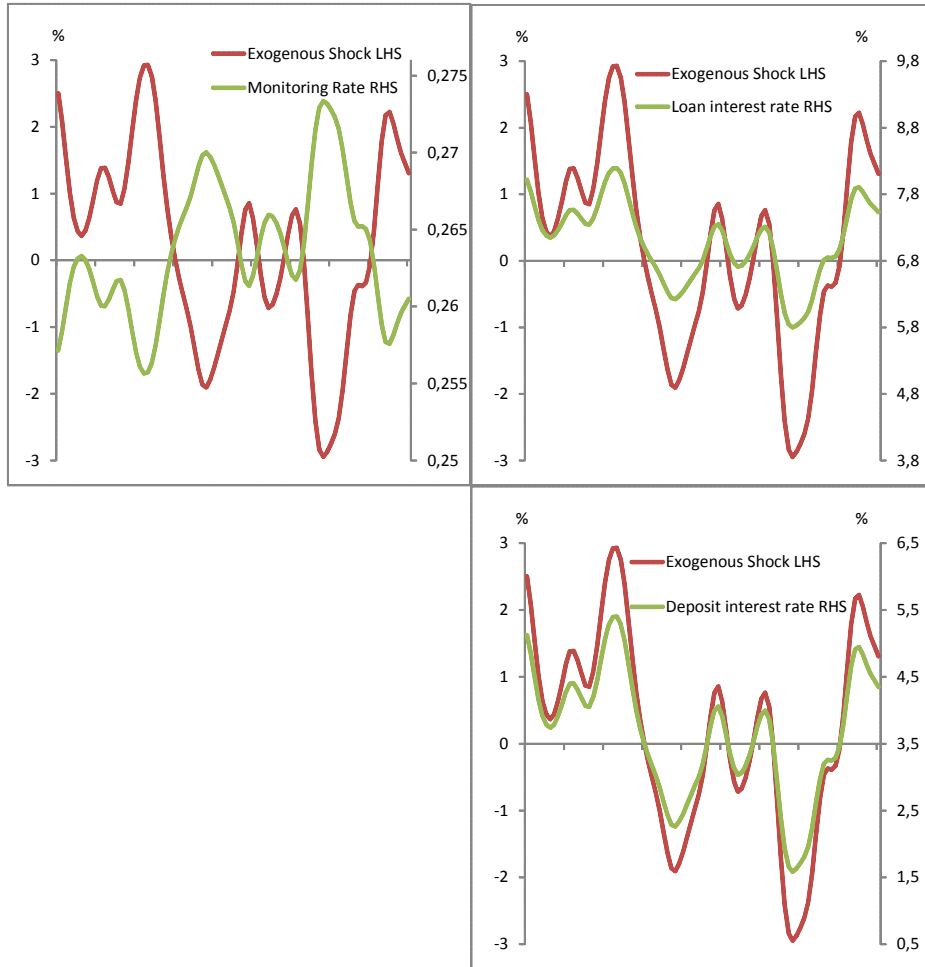
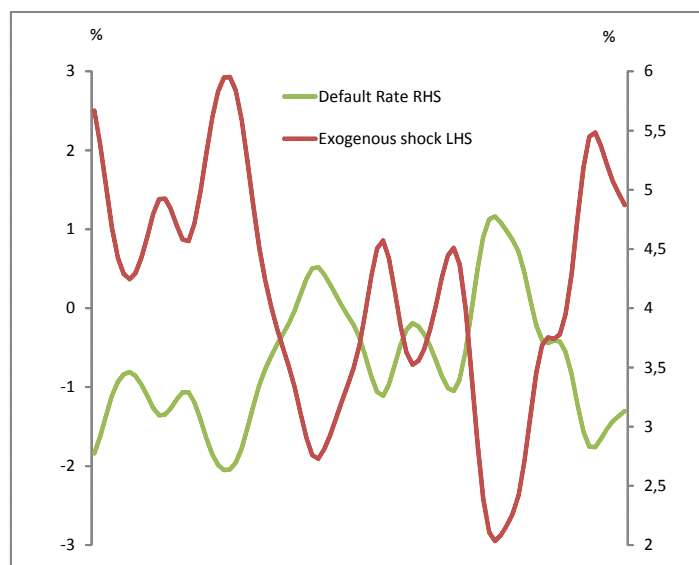


Figure 8: Simulated default rate



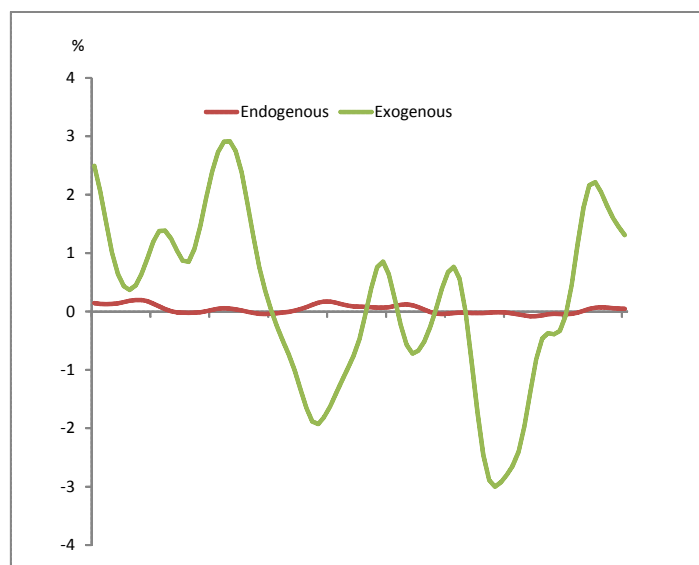
for-one with the aggregate shock, net profits for the entrepreneurs are procyclical. Since the bankruptcy cost is not assumed to change over time, a larger fraction of entrepreneurs find themselves in the region in which it is better to walk away than liquidate in an orderly manner during negative shocks.

It is now time to discuss the reasons for including profit variance in the objective function. The key is to observe that loss given default for any given loan, $L - 2(q(a') + z)$, is pro-cyclical. So a bank who has the power to manipulate the default rate (through its loan interest rate policy) will attempt to engineer a pro-cyclical default rate. But to do this, it has to generate counter-cyclical profits for entrepreneurs through counter-cyclical loan interest rates. All this is feasible provided the deposit rate is manipulated to meet the balance sheet constraint. Apart from the fact that this contradicts all the evidence on default rates, profit rates and interest rates, it also implies very large swings in bank profits. The bank makes very large losses in downturns (in order to subsidise the profits of firms) and then makes very large profits in booms (to push firms into default). This is neither realistic nor reasonable. This can be ruled out by giving the bank the incentive to smoothe out volatility in profits, naturally done in such a way that this does not become an overriding objective.

4.1. General equilibrium

The previous section presented the ‘headline’ business cycle properties of the dynamic model. There was pro-cyclicality of loan and deposit interest rates but counter-cyclicality in credit standards and the default rate. In this section I explain how these results arise from the interaction of the actors in the model and the exogenous shock process. How do the microeconomic frictions in the model influence the dynamic equilibrium? A useful benchmark is to compare the model against one in which the loan interest rate moves one-for-one with the exogenous aggregate shock with no other feedback effects. In terms of the policy functions of the previous section, this imposes $\rho_1 = 1$ and $\rho_2 = \tau_1 = \tau_2 = \varphi_1 = \varphi_2 = 0$. Recall that the net profit function is $q(a) + z - \rho(\cdot)$ and so with this policy rule, net profits are constant across the business cycle: $q(a) - \bar{\rho}$. This loan contract, in other words, completely neutralises the effect of the aggregate shock on individual firms. As a counterpart, the bank is absorbing all the aggregate risk in the economy. This is a feasible equilibrium of the model since in effect, it collapses the dynamic model back down to the static one in Penalver (2014). Firms enter and exit on the basis of their idiosyncratic risk alone and the bank solves its balance sheet constraint by finding the steady-state value of $\bar{\rho}$ which clears the market for capital. With this rule and dynamic equilibrium, the entry threshold is constant as are all the exit thresholds. The distribution of firms through time is thus always constant too. In a sense, therefore, the only reason the dynamic model of the previous section is interesting is because the bank is not willing to absorb all the aggregate risk in the economy. Naturally, any other constellation of equilibrium policy rules will not deliver an invariant distribution. So the key question is to what extent do the rules chosen by the bank imply significant variation in the allocation of capital? The implication from Figure 9 is scarcely at all. Figure 9 illustrates the contributions of the exogenous aggregate shock and the endogenous distribution of firms to the cyclical variation in aggregate productivity. In this version of the model, the aggregate supply of capital is fixed, so productivity is the same as total output. It is evident from the figure that the endogenous component is negligible. Endogenous productivity is ever so slightly counter-cyclical, consistent with the theory of Schumpeter (1939). In the preferred rules of the bank, net profits of firms are pro-cyclical so there is a mild tendency for incumbent firms to want to continue when in the benchmark model they would quit. However, the distortion is very small precisely because preferred rules are very successful in keeping dynamic equilibrium in the balance sheet. To keep the measure of deposits almost

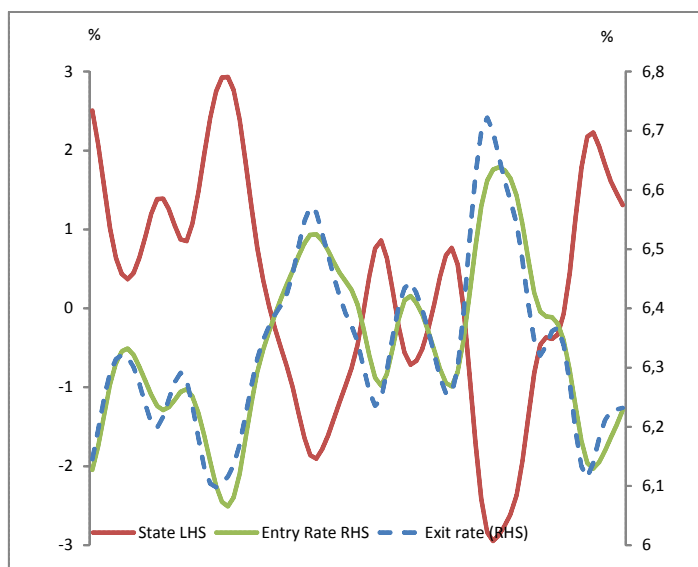
Figure 9: Contribution to productivity



identical to the measure of loans at each point in time requires in a setting of fixed loan size and fixed aggregate capital stock that the entry rate and exit rate are almost the same. Figure 10 illustrates that the entry and exit rates track each other very closely and move together in a counter-cyclical fashion. In the previous section it was observed that the default rate is counter-cyclical because of the pro-cyclicality of firm profitability. There is an equivalent effect on the voluntary exit rate as well. Therefore, if the bank is to maintain dynamic stability in its balance sheet, it needs to match the exit rate with new entrants. These parallel moves in the entry and exit rate have similar effects on allocative efficiency. The margin exiter has a level of productivity well below the margin entrant, so a higher turnover rate of firms increases productivity. But the range of the turnover rate in the illustrative calibration is relative narrow (6.0% to 6.7) and so the overall distribution changes very little over time. To put this in a slightly different way, the microeconomic frictions in this model affect the steady state distribution but do not make a great deal of difference to the dynamic adjustment of the economy.

This observation is reinforced by the experiment presented in Figure 11. This figure illustrates the default rate under two scenarios - labelled a long boom and a short boom. Both scenarios start with a neutral aggregate state ($z = 0$). Then the

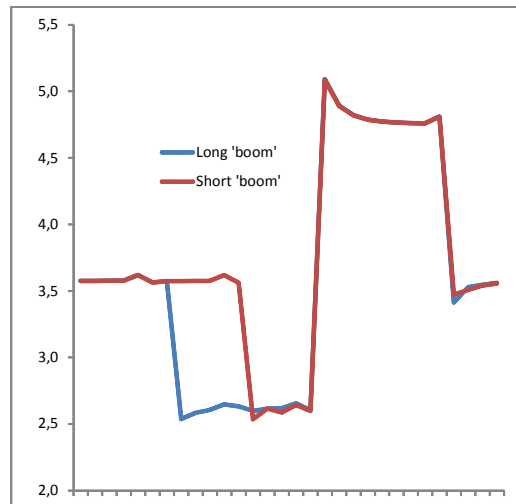
Figure 10: Entry and exit rates



aggregate state shifts to the highest value of aggregate productivity ($z = 0.03$). In the long boom, this high aggregate state lasts for 10 periods and in the short boom, this lasts for 5 periods. Both positive shocks end at the same point and the aggregate state switches to the lowest level ($z = -0.03$) for 10 periods, before returning to the neutral level. What is evident from Figure 11 is that the path of default rate after the common negative shock is identical between the two scenarios. In other words, the length of the boom makes no difference to the size of the bust. This is despite the result that positive shocks slow down the turnover rate of firms and thereby worsen the distribution of productivity. What this implies is that the economy in the model adjusts very quickly to shocks. Notice that after the positive shock, the default rate initially falls sharply but by the second period after the shock, the default rate has converged to a form of conditional steady-state.¹⁵ Remaining at the conditional steady-state for 4 or 9 periods is irrelevant. At a deeper level, this means that the apparent cyclicity of the default rate and endogenous productivity is due to the persistence of the shock

¹⁵By conditional steady state, I mean the invariant state that would arise if the aggregate state remained constant but expectations and behaviour remained consistent with the fully stochastic economy.

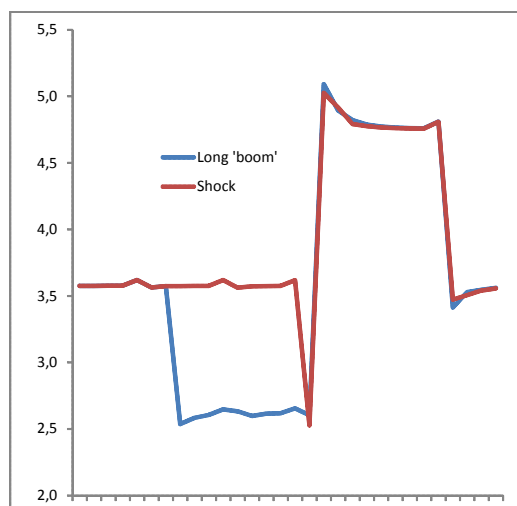
Figure 11: Long v short boom



process rather than endogenous cyclicity. This occurs because the economy faced by the agents (for equilibrium behaviour of the bank) is highly competitive and this market does a very good job of identifying the weakest firms. So, although net profits are pro-cyclical, the *relative* incentives to be an entrepreneur or an inventor are virtually cycle-free. This is because the value of the outside option, being an inventor, is also pro-cyclical. This is a crucial difference between this model and models which assume that the outside option is fixed (such as Clementi and Palazzo (2013) and Khan and Thomas (2013)). Those models find dynamic distortions in the allocation and it is not clear the extent to which this is due to lack of endogenous response of the outside option.

This analysis is confirmed by Figure 12 in which the short boom of the previous figure has been shortened to a single period positive shock. There is no difference in the default rate following the common negative shock, emphasising that the length of the positive shock is irrelevant in this case.

Figure 12: Shock v boom



5. Aggregate dynamics - pro-cyclical savings case

In this sub-section I make a small adjustment to the model in the previous section by introducing an aggregate savings rule. Specifically, I will assume an aggregate marginal propensity to save of 90%. When aggregate output rises (in response to a positive aggregate shock), the agents as a whole will save 90% of this additional amount in the form of deposits. This is, of course, an *ad hoc* assumption and I will not make any attempt in this paper to microfound this behaviour. Likewise, I will not attempt to allocate this aggregate savings behaviour across the individual agents, although clearly one option would be that all agents save 90% of the difference between their current income and their life-time average. A consequence of that rationalisation would be that those currently in production will save with deposits, rather than reduce their debt. This, too, is rationalisable by appealing, for example, to tax deductability of corporate borrowing. The purpose of this assumption is to capture in the most tractable and most transparent way the fact that savings are, in general, pro-cyclical. We can then see whether time-varying supply of capital has an implication for allocative efficiency and induces any endogenous persistence.

The algorithm to find a solution is exactly the same as in the previous sec-

tion. The algorithm searches over linear rules that maximise the same objective function subject to the optimising behaviour of the agents and model consistent expectations. The reference level of the endogenous state variable, $\bar{\theta}$, has been adjusted to take account of the pro-cyclicality of savings so that the gap $(\theta_{-1} - \bar{\theta})$ continues to represent the concept of "excess loans" and thus the pressure on the balance sheet.

The optimal rules that emerge are the following: the loan interest rate rule is

$$\rho = 0.070 + 0.29z + 0.03(\theta_{-1} - \bar{\theta}) \quad (12)$$

the deposit rule is

$$\tau = 0.035 + 0.63z + 0.10(\theta_{-1} - \bar{\theta}) \quad (13)$$

and the monitoring intensity rule is

$$\varphi = 0.20 - 1.04z - 0.51(\theta_{-1} - \bar{\theta}) \quad (14)$$

There are notable differences between these rules and those in the fixed capital case. To consider them, it is worth bearing in mind the different nature of the problem faced by the bank. With pro-cyclical savings, the bank now needs to induce additional demand for loans during periods of positive shocks and constrain lending during negative shocks. Assuming that the bank continues to move the loan interest rate by less than one-for-one with the aggregate shock so that net profits are pro-cyclical, the aggregate shocks themselves are naturally going to induce pro-cyclical loan demand. However, the bank is going to still have to work out the best constellation of policies to manage this process. Comparing equations (8) and (12), one can see that loan interest rates respond less to the aggregate shock than in the fixed capital case. This is intuitive. The bank previously had to completely neutralise the effect of the aggregate shock on the demand for loans because the supply of capital was fixed. With pro-cyclical savings, the bank does not need to be so aggressive with the loan rate to constrain loan demand. The new deposit rule, equation (13) is almost identical to the fixed capital case, with a very slight decrease in the responsiveness of the deposit rate to the aggregate shock. The noticeable differences are in the monitoring rate, equation (14). As in the analysis of the fixed capital case, the bank wants to monitor more intensely when interest rates are higher because the default risk is increased. However, from the point of view of the agents and their relative willingness to be borrowers or depositors, a more monitored loan is less attractive and they will be only induced

to borrow at a lower interest rate. With lower interest rates, the default rate decreases so the bank can afford to reduce its monitoring rate in return for a slightly higher loan interest rate. This is exactly the relationship contained in the new rules for the loan interest rate and the monitoring rate. The monitoring rule moves more aggressively to reduce the sensitivity of the loan interest rate to the aggregate shock.

To make comparison of the two version of the model easier, the following figures present the dynamics of the amended model for the same range of the simulated shock process. The first major difference, illustrated in Figure 13, is that there is now an obvious lag in the relationship between the exogenous shock and the endogenous state. A positive shock leads to additional long term lending and the new entrants are some distance from the exit threshold so a smaller fraction of the portfolio is vulnerable to shocks that would induce them to exit. Therefore, there is a tendency for the portfolio to be slightly too large if there are subsequent negative shocks. (It is also noticeable that the Markov rules are less effective in stabilising the balance sheet than in the fixed rate case, although the size of the deviations are still relatively small.)

Figure 14 shows that the entry and exit rates no longer move together. The exit rate is still counter-cyclical, reflecting the cyclical incentives to exit voluntarily and default but the size of the loan book needs to move pro-cyclically in line with aggregate savings. As a result, the entry rate is pro-cyclical.

The pro-cyclicality of the entry rate reverses the relationship of endogenous productivity to the cycle. With a fixed supply of capital, endogenous productivity was counter-cyclical. Schumpertarian "gales of creative destruction" caused weakly profitable firms to close, freeing up resources to flow to new entrants with higher profitability (and productivity). Capital resources are now pro-cyclical so the entry of more productive new firms is higher during booms. Endogenous productivity is thus now pro-cyclical, although, as before, the effect is quite small in relation to the exogenous shock process.

Figure 16 illustrates that the default rate is mildly more counter-cyclical than in the fixed savings case. This partly reflects the lower sensitivity of the loan interest rate to the aggregate shock which results in a greater sensitivity of net profits to the aggregate shock.

However, it is also the case that there is some endogenous persistence in the default rate that was absent in the fixed capital case. Figure 17 illustrates the

Figure 13: Simulated aggregate state variables

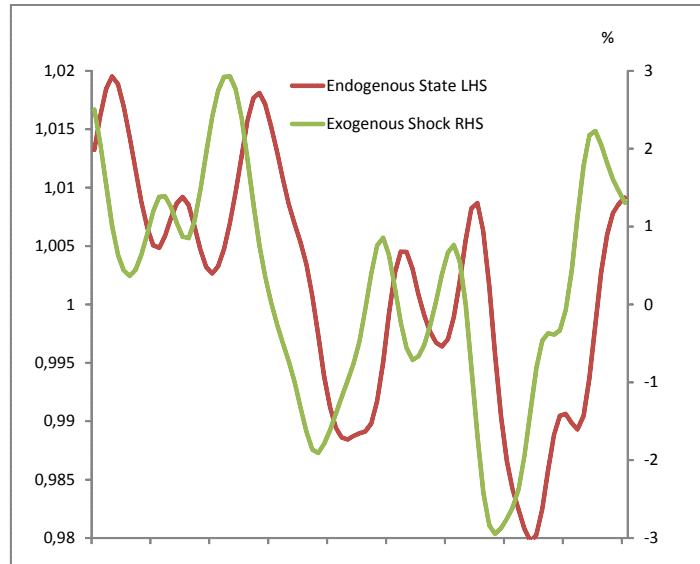


Figure 14: Entry and exit rates

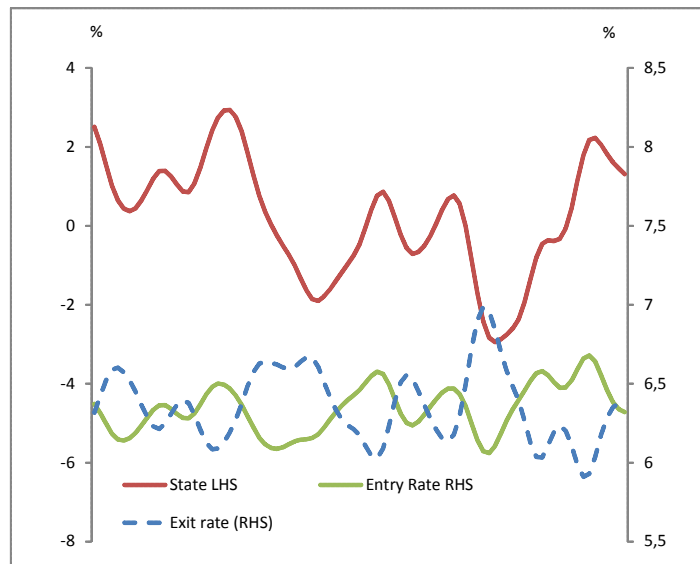


Figure 15: Contribution to productivity

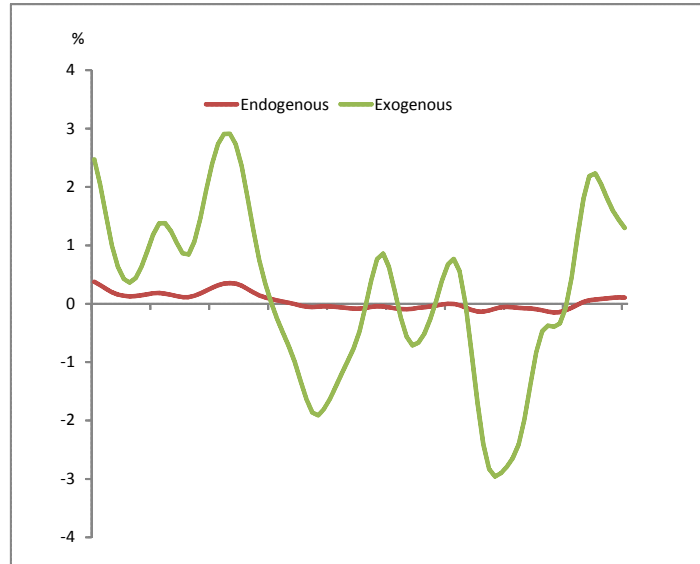


Figure 16: Default rates - fixed v pro-cyclical savings

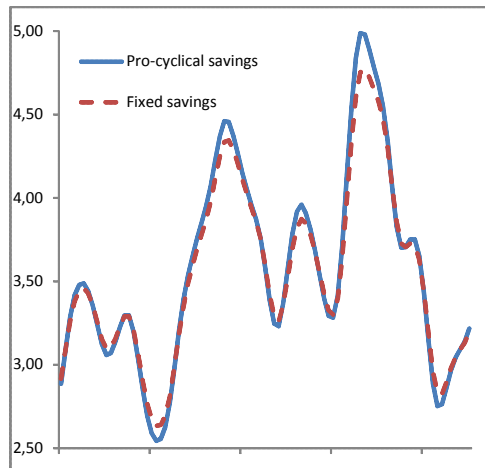
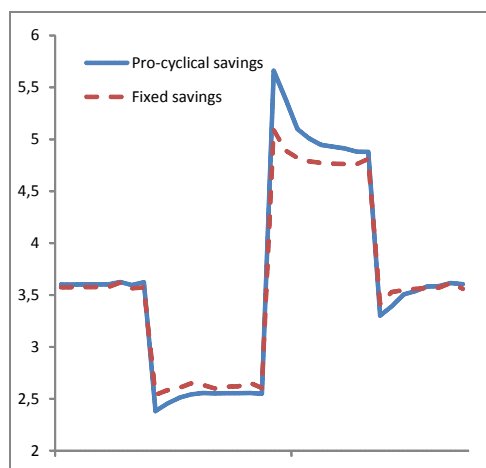


Figure 17: Impulse response - fixed v pro-cyclical savings

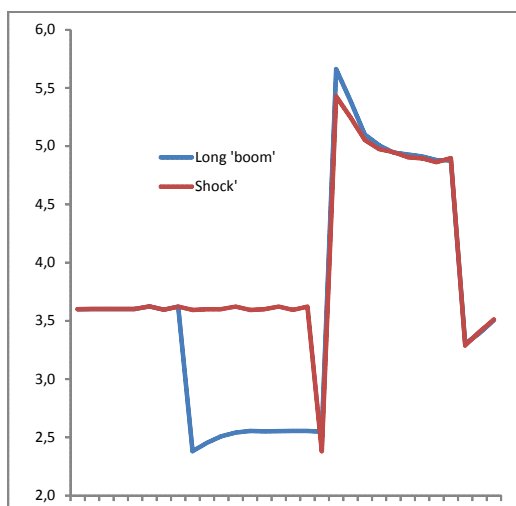


impulse responses to the same boom-bust scenarios described in the previous section. Besides confirming the greater sensitivity of the default rate to the aggregate shocks, it is evident that the default rate takes several periods to converge towards the conditional steady-state.

Part of the reason for this can be seen in Figure 18 which compares the effect on the default rate of a long period of positive shocks (a boom) with a single period shock. The Figure shows that the default rate is higher after a long boom than after a single shock, recalling that there was no difference in the fixed capital case. Why do longer booms lead to bigger busts? At least initially, this seems counter to the earlier observation that a higher entry rate improves productivity. There are two effects occurring simultaneously. The first is that after several periods, the cohort of new firms at the start of the boom have diffused into the distribution which is on-average of low profitability and higher credit risk. The positive productivity benefits of a boom gradually decay away. The second, and more important, effect is the lower monitoring during a boom. With lower monitoring, more lower credit quality firms are allowed to continue and there is thus a large mass of firms vulnerable to default following a subsequent negative shock. The longer the boom, the larger this build-up in latent vulnerability. There is a larger increase in the default rate after a sustained boom than a short shock.

It is important, however, to realise that this behaviour is not irrational or my-

Figure 18: Shock v boom



opic. Consider a bank when the aggregate shock is at its highest level. Given the persistence of the aggregate shock process, $Z(z', z)$, a boom is far more likely to gradually fade than switch discontinuously to a bust (indeed there is zero probability of this occurring in the calibration). The policy rules embed the idea that the best policy response from the bank is to gradually tighten the monitoring rate as the boom gradually unwinds. On average, this is the most rational response. Therefore, it is only with hindsight, in the rare cases in which a boom switches directly to a bust, that the ex ante behaviour looks inappropriate.

6. Conclusion

Legacy assets are a serious constraint on the functioning of banks during financial crises. Yet existing microeconomic models of banking or macroeconomic models with financial constraints are built on single period loan contracts and therefore cannot generate any history dependence of lending decisions. This paper extended a static model with multi-period loan contracts and constraints on portfolio re-optimisation to incorporate an aggregate shock. The aggregate shock induces correlation in the behaviour of the agents which destabilises the balance sheet of the bank. The bank's control variables - loan and deposit interest rates, and

the monitoring intensity - depend, in principle, on the aggregate shock and the inherited balance sheet.

An illustrative example of the model was solved under fixed or pro-cyclical savings behaviour. The bank's policy rules for loan interest rates are pro-cyclical but less than one-for-one with the aggregate shock. This implies that net profits of firms are also pro-cyclical. (Moving loan interest rates one-for-one amounts to full insurance against the aggregate shock.) Pro-cyclicality of profits results in counter-cyclical default risk. Since the aggregate shock process is persistent, there is an element of predictability about future aggregate states and thus a predictability in the default rate. Since credit standards are costly, indirectly through the equilibrium interest rate path and directly through the cost of monitoring, the bank can increase average profits through counter-cyclical credit policies. The bank also uses pro-cyclical deposit interest rates (again less than one-for-one) as a final measure to stabilise the balance sheet. These results match key features of the credit cycle.

In the fixed capital case, there is no legacy assets problem. There is no difference in the default rate following a long boom with a long period of lower credit standards than a short boom with a short period of lower credit standards. This is despite persistence in the composition of the balance sheet and time varying credit standards. This serves to show that they are necessary but not sufficient conditions for legacy assets.

By contrast, there is a legacy assets problem when there is pro-cyclical savings behaviour. In this case, a longer boom does result in a build-up in latent vulnerability in the balance sheet of the bank. But this is not irrational or myopic behaviour. The policy rules embed the idea that the best policy response from the bank is to gradually tighten the monitoring rate as the boom gradually unwinds. On average, this is the most rational response. Much of the judgement of banks' pre-crisis behaviour is with the wisdom of hindsight.

7. Bibliography

References

- [1] Arellano, C., Y. Bai and J. Zhang (2012), *Journal of Monetary Economics*, vol. 59(6) pp. 533-549

- [2] Boyd, J. and B. Smith (1994), ‘How good are standard debt contracts? Stochastic versus nonstochastic monitoring in a costly state verification environment’, *Journal of Business*, vol. 67(4), pp. 539-561.
- [3] Chen, N.-K. (2001), ‘Bank net worth, asset prices and economic activity’, *Journal of Monetary Economics*, vol 48, pp. 415-436
- [4] Clementi, G.L. and H. Hopenhayn (2006), ‘A theory of financial constraints and firm dynamics’, *The Quarterly Journal of Economics*, vol. 121, 1, pp.229-265.
- [5] Den Haan, W. (2010), ‘Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents’, *Journal of Economic Dynamics and Control* vol. 34(1), pp. 79-99.
- [6] Dixit, A and R. Pindyck, (1994), **Investment under uncertainty**, Princeton University Press, Princeton.
- [7] Gertler, M. and P. Karadi (2011), ‘A model of unconventional monetary policy’, *Journal of Monetary Economics*, vol. 58, pp. 17-34.
- [8] Gertler, M. and N. Kiyotaki (2010) ‘Financial intermediation and credit policy in business cycle analysis’, in **Handbook of Monetary Economics** edited by B. M. Friedman & M. Woodford, edition 1, volume 3, number 3, Elsevier, October 2010.
- [9] Hopenhayn, H. (1992), ‘Entry, exit and firm dynamics in long run equilibrium’, *Econometrica*, vol. 60, 5, pp. 1127-1150.
- [10] International Monetary Fund (2009), **World Economic Outlook**, October 2009
- [11] Khan, A. and J. Thomas (2013), ‘Credit shocks and aggregate fluctuations in an economy with production heterogeneity’ *Journal of Political Economy* vol. 121(6), pp.1055-1107.
- [12] Khan, A., T. Senga and J. Thomas (2014), ‘Credit shocks in an economy with heterogeneous firms and default’ Ohio State University mimeo
- [13] Kindleberger, C. (1978), **Manias, panics and crashes**, 6th edition, Palgrave Macmillan,

- [14] Krusell, P. and A. Smith (1998), ‘Income and wealth heterogeneity in the macroeconomy’ *Journal of Political Economy*, vol. 106, pp.867-896.
- [15] Lee, Y. and T. Mukoyama (2008), ‘Entry, exit and plant-level dynamics over the business cycle’ *Federal Reserve Bank of Cleveland Working Paper 07-18R*.
- [16] McCall, J. (1970), ‘Economics of information and job search’ *Quarterly Journal of Economics*, vol 84, pp. 113-126.
- [17] Meh, C. and K. Moran (2008), ‘The role of bank capital in the propagation of shocks’ *Bank of Canada working paper 2008-36*.
- [18] Minsky, H. (1992), ‘The financial instability hypothesis’ Working paper no. 74, Levy Economics Institute, Bard College.
- [19] Monnet, C. and E. Quintin (2005), ‘Optimal contracts in a dynamic costly state verification model’ *Economic Theory*, vol. 26(4) pp 867-885.
- [20] Penalver, A. (2014), ‘Bank credit standards and firm dynamics’ Paris School of Economics mimeo.
- [21] Schumpeter, J. (1939), ‘Business cycles: A theoretical, historical and statistical analysis of the capitalist process’, McGraw-Hill, New York.
- [22] Tauchen, G. (1986) ‘Finite state Markov chain approximations to univariate and vector autoregressions’, *Economic Letters*, vol. 20, 2, pp. 177-181.

8. Appendix

The model takes place in $A \times Z \times \theta$ space, respectively the idiosyncratic state, the aggregate exogenous shock and the endogenous aggregate state. This space is approximated by a 3-dimensional grid $1145 \times 7 \times 15$. The idiosyncratic grid has to be highly granular for the effect of different behavioural rules of the bank to have observable effects on the entry and exit thresholds. A coarse grid would severely limit the precision with which the rules could be calculated. The solution to the model relies on a simulation for a 10,000 periods based on a random sequence of aggregate exogenous shocks consistent with the transition probability

$Z(z', z)$. $Z(z', z)$ is a Tauchen matrix approximation for an AR(1) with zero drift, an autoregressive co-efficient of 0.9 and standard deviation of 0.75. The shocks are evenly space between ± 1.74 standard deviations. The aggregate economy passes stochastically from gridpoint to gridpoint in $Z \times \theta$ space. Since the actually evolution of θ hits a gridpoint with probability zero, the economy is randomly assigned to one of the two nearest gridpoints based on the relative distance between the true value and the gridpoints.

Entrepreneurs and inventor receive draws from A so it is necessary to solve for the value functions V_E and V_I at each point on the $A \times Z \times \theta$ grid. This is a complicated system of value functions since each is dependent on the other and the transition probabilities. The idiosyncratic state also follows an AR(1) process and approximated by a Tauchen matrix, $F(a', a)$. $F(a', a)$ has zero drift, autogressive co-efficient 0.75, standard deviation 0.3 and the outer limits are ± 4 standard deviations.

The solution to the model is a fixed point in which:

- the bank follows linear policy rules which, given the behaviour of depositors and borrowers, and the characteristics of the common shock process, maximises the objective function;
- agents' expectations are model consistent; and
- depositors and borrowers act optimally given their heterogeneous circumstances, their expectations about the future paths of their private states and the aggregate state vector, and the policy rules of the bank.

The solution is found using numerical methods using the following steps:

- Step 1: Guess parameters for the linear bank policy rules as functions of the information set, Ω .
- Step 2: Guess parameters for the linear forecasting rule for the agents for the endogenous variable θ as a function of Ω .

$$E[\theta] = \sigma_1 z + \sigma_2(\theta_{-1} - \bar{\theta})$$

- Step 3: Form a grid of a , z and θ . Use iteration to solve the system of the value functions for inventors and entrepreneurs conditional on a and Ω and

the forecast rule from Step 2 and the bank policy rules from Step 1. On the basis of these value functions, find the entry and exit thresholds for each grid point of Ω for each cohort.

- Step 4: Simulate the economy for 10,000 periods based on these rules, throw away the first 1,000 observations and estimate a new forecasting rule using ordinary least squares.
- Step 5: Repeat Steps 2 to 4 until the revision to the co-efficients on the forecasting rule is less than 1%. This is not very precise but since it is within a larger loop, I believe it to be a reasonable trade-off between accuracy and time.
- Step 6: Calculate the loss function over the estimation range and repeat Steps 1 to 5 with a different bank policy rules using the Nelder-Mead search algorithm until the objective function is maximised. The objective function is specified as

$$OBJ = 100 * \bar{\Pi}_t - (var(\Pi_t)/3)^4 - var(\theta_t)^3 - \bar{\theta}$$

where the first term is average bank profits over the truncated simulation range, the second term is variance in profits, the third term is the variance in the error on the balance sheet constraint and the fourth is the mean of the balance sheet error. This fourth term in the balance sheet constraint and is included in the objective function for tractability. There are potentially an infinite set of parameters satisfying the balance sheet constraint (to satisfactory approximation) so it would be impossible to determine this set first and the optimise over it. It is calibrated so the constraint is met very closely but not dominate the objective function. The variance in profits term is also calibrated to be relevant but not overwhelming. If set too loosely (for example ignored entirely), then the bank will try to exploit the aggregate shock process to deliver a pro-cyclical default rate with wild swings in interest rates. The bank can only achieve this by highly counter-cyclical profits. This is clearly not a reasonable outcome. By contrast, setting the variance term too tightly forces the bank to forego average profits in the pursuit of constant profitability, again an unreasonable outcome.

In Step 6, the algorithm is searching over 8 co-efficients so the outer loop needs to be repeated many times. With this many co-efficients, the algorithm takes

several days to solve even with judicious choice of starting values. How close the co-efficients are to their true values for some of the less significant parameters is of course difficult to say. Starting from different initial co-efficients yields very similar results.