# Systemic Banking Crises and Business Cycles

F. Boissay, F. Collard, and F. Smets<sup>\*</sup>

May 2012

# 1 Introduction

Why do systemic banking crises tend to break out in the midst of credit intensive booms every fourth business cycle (Jordà, Schularick, Taylor, 2011)? Why are the recessions coupled with banking crises much deeper and more protracted than other recessions (Claessens, Kose, Terrones, 2011)? To answer these questions we introduce an imperfect financial sector and the possibility of systemic banking crises into a textbook neoclassical growth model, and calibrate the model using Jordà et al. (2011)'s annual historical data set. Our model is able to account for the joint dynamics of regular business fluctuations and rare systemic banking crises observed in developed countries since the end of the nineteenth century and, in particular, for the following three stylized facts (we document these facts next section).

- 1. Recessions occur on average every ten years and systemic banking crises every forty years.
- 2. The recessions coupled with a banking crisis are deeper and last longer than other recessions.
- 3. Systemic banking crises do not break out at random within business cycles but, rather, in the midst of credit intensive booms.

In our model systemic banking crises take the form of sudden financial market freezes and break out endogenously, primarily after a long sequence of positive, transitory, technology shocks. Long spells of high productivity stimulate savings, credit growth, and bank leveraging, and result in the banking sector growing disproportionately compared with the real sector. The unwinding of these imbalances triggers a deleveraging process in the banking sector, a credit crunch, and a deep recession. The fall in output is then all the deeper because output falls from high above trend. The fall also lasts relatively long because productivity tends to go down back to its trend. Not all credit

<sup>\*</sup>Boissay and Smets are with the European Central Bank; Collard is with the University of Bern. We benefitted from discussions with P. Weil, M. Trabandt, P. Hartmann, K. Nikolov, P. Karadi, G. Lombardo, I. Jaccard.

booms lead to systemic banking crises, though: in our simulations, most of them are inherent to regular business fluctuations. Our model is able to tell apart the bad, supply-driven, credit booms from the good credit booms. Its internal dynamics rests on three important features of the banking sector.

First, banks are heterogeneous and market funding plays a prominent role. Banks are heterogeneous with respect to their financial intermediation skills and may engage in core and non-core activities. Core activities consist in collecting savings from households and lending to entrepreneurs. Under the assumption that some banks face higher intermediation costs (e.g. monitoring or debt collection costs) than others, banks also lend to each others, so as to re-allocate their assets among themselves. Following Shin (2008), we refer to this re-allocation process as banks' non-core activities. The latter improve the efficiency of the banking sector, but they require the creation of inter-bank claims and bank leveraging, that leave banks exposed to funding liquidity risk. Second, as outside option banks have the possibility to finance low productivity activities. These activities do not involve intermediation costs, but they have a lower return than inter-bank and retail loans and, as a result, are normally not financed. However, the inter-bank market is assumed to be subject to frictions -moral hazard and asymmetric information- that undermine trust among banks. This is the third important feature of the banking sector. When some specific conditions are met these frictions lead to a run on banks' non-core liabilities, making it impossible for the banks to issue interbank claims and re-allocate assets. Inter-bank market freezes are particularly inefficient because in autarky unskilled banks too lend to entrepreneurs, raising the cost of financial intermediation and the spread between the corporate loan rate and the deposit rate. More importantly the least efficient banks may even prefer to invest into the low productivity activity rather than lend to the entrepreneurs, which then leads to a fall in the supply of corporate loans and to a credit crunch. Credit crunches are the reason why, in our model, financial recessions are much deeper than other recessions.

Although the banking crises modelled here may resemble some aspects of the recent crisis, like banks' funding liquidity problems (e.g. the demise of Northern Rock) and the collapse of some market segments (e.g. that of the asset backed commercial paper in the US), our first intention is not to explain the events of 2007-2008. Our model is too stylized for this. One of the reasons why we focus on the inter-bank market frictions and circumscribe inefficiencies within the banking sector is that this allows us to delineate clearly the effects of banks' inefficiencies on the rest of the economy. Another reason is that, as we show later, the possibility of market runs generates non-linearities and a powerful internal dynamics. For instance, the mere (rational) expectation of an imminent run may generate the imbalances (e.g. disproportionate bank balance sheets) that will ultimately precipitate the run. Finally, we think that our modelling of the transmission channels between the wholesale loan market, the corporate loan market, and the real economy, fills a gap in the financial macroeconomics literature.

Related literature. As macroeconomic framework we use the textbook neoclassical growth model, bearing in mind that the insights derived from the non-trivial banking sector would survive a richer, e.g. New Keynesian DSGE, setup. Our paper belongs to the growing literature that focuses on market runs as inherent features of modern banking, like Shin (2008), Hahm, Shin and Shin (2011). Our approach differs from theirs along several important dimensions, though. First, their model is static while ours is dynamic. Second, they assume banks follow a value-at-risk rule, whereas in our case banks' behavior is derived from a profit maximization problem under constraints that arise endogenously from market imperfections. Finally, in their model a banking crisis corresponds to an amplified exogenous, adverse wealth shock to the banking sector; i.e. a crisis is a big shock. In contrast, we identify a banking crisis more clearly as a sudden freezing of the wholesale interbank market and the impossibility of trade between banks. As in Bianchi (2009), Bianchi and Mendoza (2010), Korinek (2010), Korinek and Jeanne (2010), our model exhibits non-linearities and pecuniary externalities. But unlike them we consider a general equilibrium environment where interest rates are endogenous, a crucial feature that allows us to capture the feedback effects between the real and the financial sectors. Another important difference is that our financial frictions affect banks, not firms. Accordingly, the pecuniary externalities lie within the banking sector and lead to excess credit supply when in Bianchi (2009), for instance, they lead to excess credit demand ("over-borrowing"). Finally, our model is also related to the growing number of DSGE models, in which complex financial frictions are summarized by the interest rate spread between corporate loans and bank deposits. In this literature, the spread has been assumed to be a linear increasing function of lending (Curdia and Woodford, 2009) or the output gap (Canzoneri, Collard, Dellas and Diba, 2011). Such reduced forms are useful because they keep macro-models tractable while still insightful. However, they lack micro-foundations and are subject to the Lucas critique. This is not the case here because in our model the spread is micro-founded and all parameters are structural. Given the financial frictions we consider, the derivation of the spread from first principles delivers important specific results. The spread is found to be a non-linear, discontinuous function of the corporate loan rate. This discontinuity reflects the existence of two regimes that correspond to tranquil and crisis times, and that switch endogenously over the business cycle.

*Outline of the paper.* The paper proceeds as follows. In section 2 we briefly describe the dynamics of systemic banking crises and document the three stylized facts we want to account for. Section 3 lays out the model, its micro-foundations and its dynamics. Section 4 presents the results of a number of simulation experiments. Section 5 concludes.

# 2 Three Stylized Facts on Banking Crises

We use the historical data set recently assembled by Jordà et al. (2011). This data set is annual, and includes a sample of fourteen developed countries over the period 1870-2006, as well as variables like real GDP per capita, total domestic currency loans of banks and banking institutions to nonfinancial companies and households, the dates of business cycle peaks (as defined by the NBER), and the dates of banking crises. As in Laeven and Valencia (2008), a banking crisis is an event during which the financial sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or the forced merger of major financial institutions. For the purpose of the present paper, we further define as *systemic* those of the banking crises that are concomitant with a recession, i.e. that break out between the peak and the trough of a given business cycle. In order to calculate the magnitude and duration statistics, we excluded war times and only kept complete business cycles (i.e. from peak to peak). Ultimately, our sample covers 176 full-length business cycles. Following earlier work (e.g. Reinhart and Rogoff, 2009, Claessens, Kose, Terrones, 2011, Jordà, Schularick, Taylor, 2011) we document three main stylized facts that emerge from the data.

#### 1. Recessions occur on average every ten years and systemic banking crises every forty years.

Banking crises are rare events. The countries in the sample experienced 78 banking crises, which makes one crisis every 22 years on average. Among those, about half were systemic (see table 1). In contrast, recessions are much more frequent and occur every ten years or so.

Tuble 1. Recessions and banking crises. Summary Statistics										
	N. obs	Ν	Frequency (%)	Magnitude (%)	Duration (years)					
				from peak to trough						
All banking crises	1,736	78	4.49	-	-					
Systemic Banking Crises (SBC)	1,736	42	2.42	-	-					
All recessions	1,736	176	10.20	4.86	1.85					
Recessions with SBC (A)	1,736	42	23.86	6.74	2.59					
Recessions $w/o$ SBC (B)	1,736	134	76.13	4.27	1.61					
Test A $\neq$ B, p-value (%)	-	-	-	2.61	0.00					

Table 1: Recessions and banking crises: Summary Statistics

#### 2. Financial recessions are deeper and last longer than other recessions.

While only one fourth of the recessions involve a banking crisis, these "financial" recessions are on average significantly deeper than other recessions, with real GDP per capita falling by 2.47 pp more, from peak to trough (see table 1). On average, they also last one year longer. The dynamics of these recessions is different too, in that they tend to be preceded by a faster increase in GDP and credit compared with other recessions, as Figure 1 shows. Claessens et al. (2011) report similar patterns based on a shorter data set that includes emerging countries.



Figure 1: Financial versus normal recessions

#### 3. Systemic banking crises break out in the midst of credit intensive booms.

As Gorton (1988) already showed, systemic banking crises do not occur at random. To illustrate this point, we reported in Figure 2 the distributions of GDP (left panel) and credit (right panel) gaps, as measured by the percentage deviations from Hoddrick-Prescott trend, in the year that precedes a typical systemic banking crisis (bars). The red line corresponds to the distribution in the full sample, which we use as benchmark. The figures show that, before a systemic banking crisis both the GDP and credit gaps are above their trends (with average deviations of 1.8% and 3.8%), which suggests that crises break out at a particular point in the business cycle, in the midst of credit booms. A general pattern, which Reinhart and Rogoff (2009, p. 157) also pointed out.

# 3 Model

We now turn to the model. We consider an economy populated with one representative risk averse household, one representative risk neutral competitive entrepreneur/firm, and a mass one of heterogenous, risk neutral, and competitive banks. The household lives infinitely and owns the firm and the banks, which live one period and are renewed every period.



Figure 2: Distributions of GDP and credit gaps

## 3.1 The Representative Entrepreneur

The representative entrepreneur is born at the beginning of period t and dies at the end of period t. He produces  $f_t(k_t, h_t) = z_t k_t^{\alpha} h_t^{1-\alpha}$  consumption goods with capital  $k_t$  and labour  $h_t$ . Capital depreciates at rate  $\delta$  within the period, and the logarithm of the productivity level  $z_t$  follows the following AR(1) process  $\ln z_t = \xi \ln z_{t-1} + \varepsilon_t$ , where the productivity shock  $\varepsilon_t$  is normally distributed with zero mean and standard deviation  $\sigma_{\varepsilon}$ , and  $\xi \in (0, 1)$ . Variations in productivity are the only source of uncertainty and  $\varepsilon_t$  is realized at the beginning of period t, before the entrepreneur chooses  $k_t$  and  $h_t$ . The unit wage is denoted by  $w_t$  and wages are paid at the end of the period, whereas capital goods must be purchased at the beginning of the period. Since the entrepreneur is born with no resource he borrows  $k_t$  from the banks at a gross corporate loan rate  $R_t$  at the beginning of the period. The entrepreneur chooses his inputs so as to maximize the end-of-period profit,

$$\max_{k_t, h_t} \pi_t = z_t k_t^{\alpha} h_t^{1-\alpha} + (1-\delta) k_t - R_t k_t - w_t h_t.$$
(1)

The first order conditions on capital and labour yield the optimal demands for capital

$$k_t = \left(\frac{\alpha z_t}{R_t + \delta - 1}\right)^{\frac{1}{1 - \alpha}} h_t \tag{2}$$

and labour

$$h_t = \left(\frac{(1-\alpha)z_t}{w_t}\right)^{\frac{1}{\alpha}}k_t.$$
(3)

## 3.2 The Representative Household

The representative household has concave preferences, consumes  $c_t$ , supply labour  $h_t$ , and accumulate financial assets  $a_{t+1}$  so as to maximize his intertemporal expected utility:

$$\max_{\{a_{t+i+1}, c_{t+i}, h_{t+i}\}_{i=0,\dots,+\infty}} E_t \sum_{i=0}^{i=+\infty} \beta^i \left( c_{t+i} - \vartheta \frac{h_{t+i}^{1+\upsilon}}{1+\upsilon} \right)^{1-\sigma} \frac{1}{1-\sigma},\tag{4}$$

where  $\sigma$  is the coefficient of relative risk aversion,  $\beta$  is the discount factor, v is the Frish elasticity of labour,  $\vartheta$  captures the disutility of labour (and thus governs the size of labour supply), and the expectation operator  $E_t$  (.) is taken over the  $\{\varepsilon_{t+i+1}\}_{i=0,...,+\infty}$ . The household's income comes from wages, the return on assets and the dividends paid out by the representative entrepreneur. Hence his budget constraint reads

$$c_t + a_{t+1} = r_t a_t + w_t h_t + \pi_t.$$
(5)

We do not need to take a stance on whether  $a_t$  is made of bank deposits or bank equity because, in the absence of frictions between the household and the banking sector, this is immaterial (i.e. Modigliani and Miller's theorem applies). What is important, though, is that  $r_t$  be contingent on the state of the nature. For expositional purposes, and to emphasize the fact that the household owns the banks and  $r_t$  is state contingent, we will henceforth refer to  $a_t$  as bank assets/equity and to  $r_t$  as both the return on bank assets/equity. The first order conditions yield the euler equation:

$$\left(c_t - \vartheta \frac{h_t^{1+\upsilon}}{1+\upsilon}\right)^{-\sigma} = \beta E_t \left( \left(c_{t+1} - \vartheta \frac{h_{t+1}^{1+\upsilon}}{1+\upsilon}\right)^{-\sigma} r_{t+1} \right)$$
(6)

and the optimal labour supply

$$w_t = \vartheta h_t^{\upsilon}.\tag{7}$$

Importantly, there is a positive wedge between banks' gross return on corporate loans and the return on bank equity/assets  $(r_t < R_t)$  because of inefficiencies in the banking sector.

#### 3.3 The Banking Sector

The banking sector is at the core of our model and plays a non-trivial role because of two specific features. First, banks are heterogenous with respect to their intermediation technology; some banks are more efficient than others. Banks may engage in two main activities. On the one hand they run traditional banking operations, which consist in collecting deposits/equity from households and lending the funds to the entrepreneur.<sup>1</sup> On the other hand, banks also issue inter-bank claims ("non-core" assets/liabilities) so as to re-allocate assets toward the most efficient banks of among

<sup>&</sup>lt;sup>1</sup>In Shin (2008)'s language, these are "core" activities and, accordingly, bank deposits/equity are banks' core liabilities.

them. The second specific feature of the banking sector is that it is subject to both asymmetric information and moral hazard problems, which impair the functioning of the inter-bank market. As will be clear in a moment, the aggregate supply of corporate loans depends on how well the inter-bank market functions and on whether banks can issue inter-bank claims in the first place. Hence to derive the aggregate supply of credit, we first describe banks' optimal behavior and the functioning of the inter-bank market.

#### 3.3.1 Banks

There is a continuum of banks that collect the household's savings  $a_{t+1}$  at the end of period t and lend to entrepreneurs at the beginning of period t+1, after the realization of productivity shock  $\varepsilon_{t+1}$ . At the time they collect savings banks are identical and, therefore, all have the same size as they enter period t+1. Banks are risk-neutral and competitive. They remunerate deposits/equity at rate  $r_t$ , when corporate loans yield  $R_t$ , with  $R_t \ge r_t$ . As we will show later, the difference between  $R_t$  and  $r_t$  is due to banks' intermediation costs. At the beginning of period t + 1, before they lend to the entrepreneur, banks draw at random bank-specific intermediation technologies and become heterogeneous. Denoting by p the bank with technology p, we assume that the ps are distributed over interval [0,1] with a cumulative distribution F(p), with F(0) = 0, F(1) = 1, F'(p) > 0. More precisely, bank p must pay an intermediation, deadweight, cost  $(1-p)R_t$  per unit of loan at the end of the period, so that its net return on the loan is  $pR_t$ . This cost reflects the bank's operational costs, for example, the cost of collecting corporate loans or monitoring the entrepreneur.<sup>2</sup> As outside option, banks also have the possibility to invest assets into their own project. This project does not involve any intermediation cost but yields a relatively low, constant, and exogenous payoff  $\gamma$  per unit of good invested. Such an investment is inefficient, i.e.  $\gamma < R_t$ , because of an Inada condition that  $\lim_{k_t \searrow 0} f_t(k_t, h_t) = +\infty^3$  While there are several ways to interpret this outside option, to fix ideas we will think of, and refer to, it as a storage technology. What is important here is that the funds invested in this outside option are not used to finance the entrepreneur: this is how we introduce credit crunches in the model. Bank heterogeneity gives rise to an inter-bank market, where the least efficient banks lend to the most efficient ones. Unlike

<sup>&</sup>lt;sup>2</sup>This assumption is not crucial but convenient, because in this case bank heterogeneity is immaterial to the entrepreneurs, who always pay their debt irrespective of the bank they borrow from. One could consider several alternative setups. For example, one could otherwise assume a setup à la Rochet and Tirole (1996), where banks have different monitoring skills and different incentives to monitor entrepreneurs, which affects the probability that the entrepreneur's project succeeds. Typically, the entrepreneurs that borrow from the skillful banks would be able to pay their loan in full, while those that borrow from inefficient banks would default. In this case the banks would seize the proceeds of the entrepreneur and obtain only a fraction of the loan back. In this setup deadweigh losses would take the form of a lower aggregate output, instead of financial intermediation costs. We do not use this setup because we want to confine the inefficiencies within the banking sector and, by doing so, be the closest possible to the textbook neoclassical growth model (where firms do not default).

<sup>&</sup>lt;sup>3</sup>Indeed, if  $\gamma$  were to be above  $R_t$  then banks would not finance the entrepreneur and the marginal productivity of capital would be infinite, hence a contradiction. We will also assume that  $\gamma \ge 1 - \delta$ , which means that for the bank it is more efficient to store goods than to let them depreciate. (This assumption is not critical at all.)

corporate loans, interbank loans do not involve operational costs. Let  $\rho_t$  be the gross inter-bank loan rate. Then this rate is the same for all borrowers (otherwise those that promise the lowest returns would not attract any lender), with  $\rho_t \leq R_t$  (otherwise no bank would be willing to borrow on the inter-bank market), and  $\rho_t \geq \gamma$  (otherwise no banks would be willing to lend).

Banks take the inter-bank rate  $\rho_t$  and the corporate loan rate  $R_t$  as given. Given these rates, bank p decides whether, and how much, it borrows or lends so as to maximize its profit. Since bank p gets a unit net return  $pR_t$  if it lends to the entrepreneur and a unit net return  $\rho_t$  if it lends to other banks, it is optimal to lend to other banks if  $p < \rho_t/R_t$ , and to leverage up and lend to entrepreneurs when  $p \ge \rho_t/R_t$ . We will refer to the banks that supply funds on the interbank market as "lenders" and to those that borrow as "borrowers". Let  $\phi_t$  be the amount borrowed per unit of equity/deposit by a borrower p, with  $\phi_t \ge 0$ , so that  $\phi_t a_t$  is borrower p's inter-bank loan. We will refer to  $\phi_t$  as bank leverage.<sup>4</sup> It is endogenous and publicly observable. The objective of borrower p consists in maximizing its return on assets/equity, denoted by  $r_t(p)$ , which is equal to the net revenue from corporate loans minus the payment of inter-bank loans:

$$\max_{\phi_t \ge 0} r_t(p) \equiv p R_t \left( 1 + \phi_t \right) - \rho_t \phi_t \tag{8}$$

with respect to  $\phi_t$ , under the constraint that it is optimal that bank p borrows,

$$p \geqslant \overline{p}_t \equiv \frac{\rho_t}{R_t}.$$
(9)

Constraint (9) is borrower p's participation constraint and pins down the type of the marginal bank  $\overline{p}_t$ , which is indifferent between borrowing and lending on the inter-bank market. In a frictionless world all banks p < 1 would lend to the most efficient bank p = 1, so that  $\overline{p}_t = 1$ . This bank would have an infinite leverage ( $\phi_t \to +\infty$ ) and corner all savings. The economy would then reach the First Best allocation, there would be no deadweight loss. To make things interesting we introduce two frictions on the inter-bank market that prevent the economy from reaching this First Best: moral hazard and asymmetric information.

Moral Hazard. We assume that the proceeds of storage are not traceable and cannot be seized by any creditor. This implies that interbank loan contracts are not enforceable and that banks have the possibility to walk away with the funds raised on the inter-bank market, without paying the inter-bank loans at the end of period t+1. Following current practice (e.g. Hart, 1995, Burkart and Ellingsen, 2004), we refer to such opportunistic behavior as "cash diversion".<sup>5</sup> When a bank diverts cash, the proceeds ultimately accrue to the household, who is its shareholder. The cash diverted is

<sup>&</sup>lt;sup>4</sup>More precisely,  $\phi_t$  is also the ratio of market funding (non-core liabilities) to traditional funding (core-liabilities).

<sup>&</sup>lt;sup>5</sup>One can easily recast our moral hazard model into a setup  $\dot{a}$  la Holmstöm and Tirole, whereby borrowers may misuse the funds and enjoy private benefits at the expense of their creditors.

stored until the end of the period, so that the return of diversion is  $\gamma$ .<sup>6</sup> However, a bank that diverts  $(1 + \phi_t) a_t$  faces a diversion  $\cot(1 - \theta) \phi_t a_t$  proportional to the size of the interbank loans, and can run away with  $\gamma (1 + \theta \phi_t) a_t$  only as net payoff, with  $\theta \in [0, 1]$ . Parameter  $\theta$  reflects the cost of diversion, which is null when  $\theta = 1$  and maximal when  $\theta = 0$ . In itself this particular specification is not critical, but it helps keep the model tractable. As an additional degree of freedom, parameter  $\theta$  will also turn out to be useful when we calibrate the model. What is important here is that the gain from diversion increases with  $\phi_t$  (i), and that the opportunity cost of diversion increases with bank efficiency p (ii) and with the corporate loan rate  $R_t$  (iii). Points (i) and (ii) are most standard in the corporate finance literature (e.g. Tirole, 2006). They mean that efficient banks with much "skin in the game" are less inclined to run away than highly leveraged and inefficient banks. Point (iii) is similar in spirit to point (ii) but in the "time series" dimension as opposed to the "cross-section" dimension. It means that banks are more inclined to run away in times when the return on corporate loans is low. This assumption captures recent empirical evidence that banks tend to take more credit risk in such particular times (Maddaloni and Peydro, 2011).

Asymmetric Information. There is an asymmetry of information between borrowers and lenders in the sense that the ps are privately known. Lenders do not observe borrowers' skills and therefore do not know borrowers' incentive to divert cash. In this context, the loan contracts signed on the inter-bank market are the same for all banks; neither  $\phi_t$  nor  $\rho_t$  depend on p.<sup>7</sup>

Clearly, no bank will lend if there exist other banks willing to borrow and divert the funds. Lenders want to deter borrowers from diverting. They can do so by limiting the quantity of funds that borrowers can borrow, so that even the most inefficient banks (i.e. those that should be lending) have no interest in demanding a loan and diverting it:

$$\gamma \left(1 + \theta \phi_t\right) \leqslant \rho_t. \tag{IC}$$

This incentive compatibility constraint sets a limit to  $\phi_t$ . It will be useful to think of this limit as lenders' leverage tolerance, i.e. the limit above which a bank refuses to lend or, in Tirole's language, the bank's pledgeable income. The program of borrower  $p \ge \overline{p}_t$  thus consists in maximizing its

<sup>&</sup>lt;sup>6</sup>Note that, to be coherent, the return of cash diversion cannot be higher than that of storage. Otherwise, the diversion technology would dominate storage and would therefore become the relevant outside option for the banks. This is one of the two reasons why we assume a storage technology (the other being the possibility of a credit crunch).

<sup>&</sup>lt;sup>7</sup>Consider indeed a menu of debt contracts  $\{\rho_t(\tilde{p}), \phi_t(\tilde{p})\}_{\tilde{p} \in [0,1]}$  intended for the borrowers of types  $\tilde{p}$ s, and notice that lenders' arbitrage across these contracts requires that  $\rho_t(\tilde{p}) = \rho_t \ \forall \tilde{p} \in [0,1]$ . Then it is easy to show that such a menu of contracts cannot be revealing, for any borrower p that would claim being of type  $\tilde{p}$  would then make profit  $r_t(\tilde{p} \mid p) = pR_t + (pR_t - \rho_t) \phi_t(\tilde{p})$  and pick the contract with the highest  $\phi_t(\tilde{p})$ . It is equally easy to show that there is no revealing menu of equity contracts either. Indeed, consider a menu of equity contracts  $\{\eta_t(\tilde{p}), \phi_t(\tilde{p})\}_{\tilde{p} \in [0,1]}$ , where  $\eta_t(\tilde{p})$  would be the share of earnings that accrue to the bank that raises equity. Then the net profit of a bank p would be  $\eta_t(\tilde{p}) (1 + \phi_t(\tilde{p})) pR_t$ . It is clear that this bank would pick the contract that yields the highest net return  $\eta_t(\tilde{p}) (1 + \phi_t(\tilde{p}))$ , independently of its own p.

profit (8) with respect to  $\phi_t$  under constraint (IC). It is easy to see that in the optimum this constraint binds, i.e. that

$$\phi_t = \frac{\rho_t - \gamma}{\gamma \theta}.\tag{10}$$

The positive relationship with  $\rho_t$  is a crucial feature of bank leverage. When  $\rho_t$  increases the net present value of corporate loans diminishes and only the most efficient banks remain on the demand side of the inter-bank market. Since these banks have little incentive to divert the cash, lenders do not need to incentivize them as much and therefore tolerate a higher leverage ( $\phi_t$  goes up). That is, borrowers and lenders' interests are all the more aligned when  $\rho_t$  is high. Intuitively, this is due to the negative externality that the marginal borrower exerts on other borrowers when he enters the demand side of the market as, by having higher incentive to run away, he raises lenders' counterparty fears. A limit case is  $\rho_t = \gamma$ , when there is no demand for inter-bank loan because borrowers cannot commit to repay any loan. Leverage  $\phi_t$  and the type of the marginal borrower  $\overline{p}_t$ fully describe banks' optimal decisions.

#### 3.3.2 Inter-bank Market

We now turn to the functioning of the inter-bank market. The equilibrium of the inter-bank market is characterized by the gross interest rate  $\rho_t$  such that the market clears. We first look for an equilibrium where  $\rho_t > \gamma$  so that  $\phi_t > 0$  and trade can take place. In this case, a mass  $F(\bar{p}_t)$  of banks lend  $a_t$  and the aggregate supply of funds on the inter-bank market is equal to  $F(\bar{p}_t) a_t$ . And a mass  $1 - F(\bar{p}_t)$  of banks borrow  $\phi_t a_t$ , so that aggregate demand is equal to  $(1 - F(\bar{p}_t)) \phi_t a_t$ . The market thus clears when (using relations (9) and (10)):

$$\underbrace{F\left(\frac{\rho_t}{R_t}\right)}_{\text{supply}} = \underbrace{\left(1 - F\left(\frac{\rho_t}{R_t}\right)\right)}_{\text{demand}} \underbrace{\frac{\rho_t - \gamma}{\gamma\theta}}_{\text{demand}} \Leftrightarrow R_t = \frac{\rho_t}{F^{-1}\left(\frac{\rho_t - \gamma}{\rho_t - (1 - \theta)\gamma}\right)}.$$
(11)

Aggregate supply increases monotonically with  $\rho_t$ , whereas aggregate demand is driven by two opposite effects. On the one hand, a rise in  $\rho_t$  lowers aggregate demand because fewer borrowers demand funds when the cost of funds is higher; this is the "extensive margin" effect. On the other hand, it also works to augment aggregate demand because each borrower is able to leverage more; this is the "intensive margin" effect. This latter effect more than offsets the extensive margin effect when the externality affects a large mass of borrowers, i.e. when  $\rho_t$  is below a certain threshold. It is easy to see that expression  $\rho_t/F^{-1}((\rho_t - \gamma) / (\rho_t - (1 - \theta) \gamma))$  goes to infinity when  $\rho_t \searrow \gamma$ , is above  $R_t$  for  $\rho_t \nearrow R_t$ , and reaches a minimum for some value  $\rho_t = \overline{\rho} > \gamma$ . Hence there exists a threshold  $\overline{R} \equiv \overline{\rho}/F^{-1}((\overline{\rho} - \gamma) / (\overline{\rho} - (1 - \theta) \gamma))$  for  $R_t$  below which there is no equilibrium with trade possible.<sup>8</sup> This threshold is the minimum corporate loan rate that is necessary for banks to

<sup>&</sup>lt;sup>8</sup>Equivalently, one can also write the market clearing condition in terms of  $\overline{p}_t$  (since it is a multiple of  $\rho_t$ ) and then obtain condition  $\gamma (1 - (\theta - 1) F(\overline{p}_t)) / \overline{p}_t (1 - F(\overline{p}_t)) = R_t$ . It is easy to see that the left hand side expression

trust, and accept to lend to, each other. To see this point better consider Figure 3, which depicts the movements of the aggregate supply and demand curves as  $R_t$  falls from  $R_{high}$  (plain black lines) to  $R_{low}$  (dashed black lines), with  $R_{low} < \overline{R} < R_{high}$ . Following this change, the supply curve shifts to the right and the demand curve shifts to the left so that, given  $\rho_t$ , demand falls below supply. Market clearing then requires that  $\rho_t$  go down, which results in more banks demanding funds (extensive margin). But since the banks that switch from the supply to the demand side are less efficient and have a relatively higher incentive to divert cash, lenders require borrowers to de-leverage even more. By construction, this intensive margin effect is the strongest when  $R_t < R$ . It follows that, ultimately, aggregate demand decreases and excess supply goes further up, not down. The deleveraging process feeds itself and goes on until the market freezes, in point A. At this point autarky prevails. Demand and supply are both equal to zero, and the market clears because borrowers have no pledgeable income ( $\phi_t = 0$ ) and lenders are indifferent between interbank loans and storage ( $\rho_t = \gamma$ ). The marginal bank  $p = \gamma/R_t$  is indifferent between financing the entrepreneur and storage, and a fraction  $F(\gamma/R_t)$  of banks' total assets are kept within the banking sector, instead of being channelled to the entrepreneur. Moreover, this autarkic equilibrium is stable in the sense that, following any small perturbation to  $\rho_t$  away from A, a standard Walrasian tatonnement process –whereby  $\rho_t$  increases (decreases) whenever demand (supply) is in excess– would bring  $\rho_t$  back to A. In the rest of the paper, we will interpret such a situation as a banking crisis.

Things are very different when  $R_t \ge \overline{R}$  in that borrowers may then have enough incentives to finance the entrepreneur and an inter-bank market equilibrium with trade may exist. For usual cumulative distributions  $F(p)^9$  there are two possible inter-bank market equilibria with trade, but only the one associated with the highest inter-bank loan rate, i.e.  $\rho_t \ge \overline{\rho}$ , is stable. In figure 3 this equilibrium is represented by point E, which we refer to as "normal times". We rule out the other equilibrium, represented by point U, because it is unstable. Finally, notice that the autarkic equilibrium too exists when  $R_t \ge R$ . This is because of strategic complementarities between lenders (see Cooper and John, 1988), that imply that no bank has interest in making a loan if the other banks refuse to lend. It follows that the autarkic equilibrium coexists with the equilibrium with trade whenever the latter exists. In order to rule out potential coordination failures we will assume that banks always coordinate on the equilibrium with trade (which is pareto-dominant) in this case. Hence, the inter-bank market freezes only when no equilibrium with trade exists. Based on relations (8) and (11), we can complete the description of the banking sector by deriving the sector's return on

is infinite for  $\overline{p}_t = 0, 1$  and reaches a minimum  $\overline{R}$  for some value  $\overline{p}_t = \overline{\overline{p}} \in (0, 1)$ . <sup>9</sup>This is notably the case for the family distribution  $F(p) = p^{\lambda}$  (with  $\lambda \ge 0$ ) that we will be using later in the calibration.



equity:

$$r_{t} \equiv \int_{0}^{1} r_{t}(p) dF(p) = \begin{cases} R_{t} \int_{\overline{p}_{t}}^{1} p \frac{dF(p)}{1 - F(\overline{p}_{t})} , \text{ if an equilibrium with trade exists} \\ R_{t} \left( \frac{\gamma}{R_{t}} F\left( \frac{\gamma}{R_{t}} \right) + \int_{\overline{R}_{t}}^{1} p dF(p) \right) , \text{ otherwise} \end{cases},$$
(12)

The interpretation of  $r_t$  is clear. When the inter-bank market exists, inefficient banks delegate financial intermediation to a mass  $1 - F(\bar{p}_t)$  of efficient banks, each of which therefore lending a multiple  $1 + \phi_t = 1/(1 - F(\bar{p}_t))$  of their initial assets to the entrepreneur against net return  $pR_t$ . In autarky, in contrast, a mass  $1 - F(\gamma/R_t)$  of the banks make corporate loans, whereas the remainder use the storage technology. Note that the banking sector is fully efficient when  $\gamma \to 0$ , i.e. when inter-bank loan contracts are fully enforceable, as in this case  $\overline{R} \to 0$  (the inter-bank market always exists),  $\overline{p}_t \to 1$  (only the best bank does the intermediation), and  $\phi_t \to +\infty$  (the best bank is infinitely leveraged). The same is true when  $\lim_{p \nearrow 1} F(p) = 0$ , since in this case there is a mass one of banks with p = 1 and banks are homogenous and all efficient. We can now turn to the derivation the aggregate supply of bank loans.

## 3.3.3 Aggregate Supply of Corporate Loans

From (11) it is easy to see that in normal times all bank assets  $a_t$  are channelled to the entrepreneur. When the interbank market freezes, in contrast, only the banks with  $p \ge \gamma/R_t$  lend to the entrepreneurs and the banking sector supplies only  $(1 - F(\gamma/R_t))a_t$  of corporate loans. Denoting by  $k_t^s$  banks' aggregate supply of corporate loans, one thus gets:

$$k_t^s = \begin{cases} a_t , \text{ if an equilibrium with trade exists} \\ (1 - F(\gamma/R_t)) a_t , \text{ otherwise} \end{cases}$$
(13)

Everything is now in place to derive the general equilibrium of the economy.

#### 3.4 Recursive Decentralized General Equilibrium

We solve the general equilibrium in a sequential way, starting with the labour market equilibrium, defined as the situation where the labour market clears. Using (3) and (7) one can express the equilibrium level of labour as a function of entrepreneurs' capital stock,

$$h_t = \left(\frac{(1-\alpha)z_t}{\vartheta}\right)^{\frac{1}{\upsilon+\alpha}} k_t^{\frac{\alpha}{\upsilon+\alpha}},\tag{14}$$

which, once substituted out of relation (2), yields the demand for capital as a function of the corporate loan rate only:

$$k_t = \left(\frac{1-\alpha}{\vartheta}\right)^{\frac{1}{\upsilon}} \left(\frac{\alpha}{R_t + \delta - 1}\right)^{\frac{\upsilon + \alpha}{\upsilon(1-\alpha)}} z_t^{\frac{1+\upsilon}{\upsilon(1-\alpha)}}.$$
(15)

Using (2), (13) and (15), we can write the corporate loan market clearing condition as

$$\left(\frac{1-\alpha}{\vartheta}\right)^{\frac{1}{v}} \left(\frac{\alpha}{R_t+\delta-1}\right)^{\frac{v+\alpha}{v(1-\alpha)}} z_t^{\frac{1+v}{v(1-\alpha)}} = \begin{cases} a_t , \text{ if an equilibrium with trade exists} & (a) \\ a_t - F(\gamma/R_t) a_t , \text{ otherwise} & (b) \end{cases}, \quad (16)$$

which yields the equilibrium  $R_t$  as a function of the two predetermined, state variables of the model,  $a_t$  and  $z_t$ . Relation (16) points to the two-way relationship that exists between the interbank loan market and the retail corporate loan market. We have shown earlier that the way the interbank functions depends on whether or not  $R_t \ge \overline{R}$ . At the same time whether or not the interbank market functions has an impact on the supply of corporate loans and, ultimately, on  $R_t$ . To solve for the general equilibrium we need to take into account these feedback effects between the wholesale and the retail loan markets. We proceed in two steps. First, we solve (16a) for  $R_t$  under the conjecture that the interbank market equilibrium with trade exists, and then check a posteriori whether indeed  $R_t \ge \overline{R}$ . If  $R_t \ge \overline{R}$  then the initial conjecture is correct and  $R_t$  is the equilibrium corporate loan rate. If in contrast  $R_t < \overline{R}$  then the initial conjecture is false: the interbank market equilibrium with trade exists for the initial conjecture is false. In this case the equilibrium corporate loan rate is the  $R_t$  that solves (16b). Proposition 1 follows.

**Proposition 1 (Inter-bank loan market freeze)**: The inter-bank loan market functions well if and only if  $a_t \leq \overline{a}_t$ , with  $\overline{a}_t \equiv ((1 - \alpha) / \vartheta)^{\frac{1}{\upsilon}} (\alpha / (\overline{R} + \delta - 1))^{\frac{\upsilon + \alpha}{\upsilon(1 - \alpha)}} z_t^{\frac{1 + \upsilon}{\upsilon(1 - \alpha)}}$ , and freezes otherwise.

**Proof:** It is easy to see that the solution  $R_t$  to (16a) decreases monotonically with  $a_t$  and falls below  $\overline{R}$  once  $a_t > \overline{a}_t$ .

The threshold  $\overline{a}_t$  is the maximum quantity of assets that the banking sector can reallocate efficiently. Above this threshold counterparty fears on the inter-bank market are so large that mistrust prevails and the inter-bank market freezes. In the rest of the paper we will refer to  $\overline{a}_t$  as the absorptive capacity of the banking sector. Importantly, proposition 1 suggests that the capacity of the banking sector to re-allocate assets internally ultimately depends on the level of productivity in the real sector,  $z_t$ . The more productive the real sector, the more efficient the banking sector ( $\partial \overline{a}_t/\partial z_t > 0$ ). The intuition and mechanics are clear. An increase in total factor productivity raises the demand for capital and the equilibrium corporate loan rate. By raising banks' opportunity cost of storage and cash diversion, the increase in  $R_t$  also increases trust within the banking sector, making it less likely that the inter-bank loan market freezes. Overall, our model captures the notion that banks' core liabilities (equity/deposits  $a_t$ ), which are predetermined, are a stable source of funding whereas non-core liabilities are unstable funding because they are subject to market runs. Proposition 2 below shows that the disruptions in the wholesale financial market may spill over the retail loan market and trigger a credit crunch.

Proposition 2 (Credit crunch): Inter-bank market freezes trigger credit crunches. A credit

crunch is characterized (i) by the sudden increase in the corporate loan rate  $R_t$  ( $\lim_{a_t \searrow \overline{a}_t} R_t$  >  $\lim_{a_t \nearrow \overline{a}_t} R_t$ ) and (ii) by the sudden increase in the interest rate spread  $R_t/r_t$  ( $\lim_{a_t \searrow \overline{a}_t} R_t/r_t$  >  $\lim_{a_t \nearrow \overline{a}_t} R_t/r_t$ ).

**Proof:** Point (i): From (16) and proposition 1 it is easy to see that  $\lim_{a_t \nearrow \overline{a}_t} R_t + \delta - 1 = \lim_{a_t \searrow \overline{a}_t} (1 - F(\gamma/R_t))^{\frac{-\nu(1-\alpha)}{\alpha+\nu}} (R_t + \delta - 1) > \lim_{a_t \searrow \overline{a}_t} R_t + \delta - 1$ . Point (ii): From (12) we know that  $\lim_{a_t \nearrow \overline{a}_t} r_t/R_t = \int_{\overline{p}_t}^1 p \frac{dF(p)}{1 - F(\overline{p}_t)}$ , which conditional expectation monotonically increases in  $\overline{p}_t$ . Moreover, we know from (11) and point (i) that  $\lim_{a_t \nearrow \overline{a}_t} \overline{p}_t > \lim_{a_t \nearrow \overline{a}_t} \gamma/R_t > \lim_{a_t \searrow \overline{a}_t} \gamma/R_t$ . Hence,  $\lim_{a_t \nearrow \overline{a}_t} r_t/R_t > \lim_{a_t \searrow \overline{a}_t} \int_{\gamma/R_t}^1 p \frac{dF(p)}{1 - F(\gamma/R_t)}$ . Finally, we know from (12) that  $\lim_{a_t \searrow \overline{a}_t} r_t/R_t = \lim_{a_t \searrow \overline{a}_t} \frac{\gamma}{R_t} F\left(\frac{\gamma}{R_t}\right) + \int_{\gamma/R_t}^1 p dF(p)$  which is smaller than  $\lim_{a_t \searrow \overline{a}_t} \int_{\gamma/R_t}^1 p \frac{dF(p)}{1 - F(\gamma/R_t)}$ . Hence,  $\lim_{a_t \nearrow \overline{a}_t} r_t/R_t > \lim_{a_t \searrow \overline{a}_t} r_t/R_t = \lim_{a_t \searrow \overline{a}_t} r_t/R_t$  and our result follows.

Figure 4 illustrates proposition 2. It depicts the equilibrium rates,  $R_t$ ,  $r_t$ , and  $\rho_t$  (y-axis) as functions of  $a_t$  (x-axis), given  $z_t$ . The corporate loan rate monotonically decreases with bank assets almost everywhere, but there is a break for  $a_t = \overline{a}_t$ , when assets reach the banking sector's absorptive capacity. Above this threshold, a credit crunch occurs and the corporate loan rate suddenly jumps to  $\widetilde{R} \equiv \lim_{a_t \searrow \overline{a}_t} R_t$ , with  $\widetilde{R} > \overline{R}$ . Notice that, from a *partial* equilibrium perspective,  $\widetilde{R}$  is high enough to restore banks' incentives and re-ignite the inter-bank market. But this is not sustainable as a rational expectation general equilibrium, since by issuing interbank claims banks would be able to raise their supply of corporate loans and, ultimately,  $R_t$  would go down below  $\overline{R}$ . It follow that the autarkic equilibrium is the only inter-bank market equilibrium that is consistent with the general equilibrium when  $a_t > \overline{a}_t$  and that, while necessary, the condition  $R_t \ge \overline{R}$  is not sufficient to rule out inter-bank market freezes.

Figure 4 also shows that, in autarky, bank inefficiencies materialize themselves by a widening of the interest rate spread, which is due to the simultaneous increase in the corporate loan rate and the fall in the return on bank asset/equity. Everything is now in place to define the decentralized recursive, competitive, general equilibrium of the economy.

**Definition 1 (Recursive decentralized general equilibrium):** A decentralized recursive general equilibrium is defined by the pricing functions {  $R_{t+i}(a_{t+i}, z_{t+i}), r_{t+i}(a_{t+i}, z_{t+i}), \rho_{t+i}(a_{t+i}, z_{t+i})$ }  $z_{t+i}$ }  $z_{t+i}$  and decision rules {  $k_{t+i}(a_{t+i}, z_{t+i}), h_{t+i}(a_{t+i}, z_{t+i}), c_{t+i}(a_{t+i}, z_{t+i}), y_{t+i}(a_{t+i}, z_{t+i}), \overline{a}_{t+i}(z_{t+i}), a_{t+i+1}(a_{t+i}, z_{t+i})$ } such that (i) the representative entrepreneur maximizes profit (1) taking prices { $R_{t+i}$ } $_{i=0}^{i=+\infty}$  as given, (ii) the representative household maximizes utility (4) subject to budget constraint (5) taking prices { $r_{t+i}$ } $_{i=0}^{i=+\infty}$  as given, and (iv) all markets clear at each date.



## 3.4.1 Equations of the Model

Our model can be summarized by the equations in table 2. There are two regimes in the economy: one where the inter-bank market functions (normal times) and another where the inter-bank market is frozen (crisis times). Which regime prevails depends on the quantity of financial assets  $a_t$ , relative to the banking sector's absorptive capacity,  $\overline{a}_t$ .

Table 2: Equations of the model  $^{10}$ 

#### 3.4.2 Good versus Bad Credit Booms

As we already discussed, banking crises break out whenever banks' assets are in excess of banks' absorptive capacity. Since these assets feed the supply of corporate loans, this means that credit booms may be bad, in the sense that they may lead to crises. Only few credit booms are prone to crises, though, as the majority reflect productivity gains in the real sector. To tell apart the bad and the good credit booms, we define the probability of a crisis at a *n*-period horizon as the joint probability that the banking sector's total assets exceed its absorptive capacity in t + n (i.e. that  $a_{t+n} > \overline{a}_{t+n}$ ) and not before (i.e.  $a_{t+i} \leq \overline{a}_{t+i}$  for i = 1, ..., n - 1).

**Definition 2 (Probability of a crisis at a** n-period horizon  $\Delta_t^n$ ): Given the data generating process of productivity, the initial state at the end of period  $t(a_{t+1}, z_t)$ , and the optimal asset accumulation rule  $a_{t+i+1}(a_{t+i}, z_{t+i})$ , the probability that a systemic banking crisis next breaks out in period t+n is  $\Delta_t^n \equiv \Pr(a_{t+1} \leq \overline{a}_{t+1}, ..., a_{t+n-1} \leq \overline{a}_{t+n-1}, a_{t+n} > \overline{a}_{t+n}) = \int_{\overline{\varepsilon}_{t+1}}^{+\infty} ... \int_{\overline{\varepsilon}_{t+n-1}}^{+\infty} dG(\varepsilon_{t+1}, z_{t+1}) dG(\varepsilon_{t+1})$ 

<sup>&</sup>lt;sup>10</sup>Few comments are in order here. Summing up the households and entrepreneurs' budget constraints (1) and (5), one gets:  $c_t + a_{t+1} = a_t k_t^{\alpha} h_t^{1-\alpha} + (1-\delta) k_t + r_t a_t - R_t k_t$ . Relations 11a and 11b can be derived from Walras' law and the agents' budget constraints. In particular, in autarky the aggregate intermediation cost amounts to (using 7b and 8b):  $\int_{\gamma/R_t}^1 (1-p) R_t a_t dF(p) = (1-F(\gamma/R_t)) R_t a_t - a_t \int_{\gamma/R_t}^1 p R_t dF(p) = R_t a_t - R_t (a_t - k_t) - a_t (r_t - \gamma F(\gamma/R_t)) = (R_t - r_t) a_t - (R_t - \gamma) (a_t - k_t).$ 

...,  $\varepsilon_{t+n}$ ), where G(.) denotes the cumulative normal n-variate distribution, n > 1, and  $\overline{\varepsilon}_{t+i} \equiv \ln \overline{z}_{t+i} - \xi \ln z_{t+i-1}$ , with  $\overline{z}_{t+i} \equiv \left[ \left( \frac{\vartheta}{1-\alpha} \right)^{\frac{1-\alpha}{\alpha+\nu}} \frac{\overline{R}+\delta-1}{\alpha} a_{t+i}^{\frac{\upsilon(1-\alpha)}{\alpha+\nu}} \right]^{\frac{\alpha+\nu}{1+\nu}}$  being the threshold of productivity in period t+i below which, given the level of financial assets  $a_{t+i}$ , a crisis breaks out.

This probability provides an indicator of financial fragility that is fully consistent with agents' rational expectations and perceived risks. For instance, the ex ante anticipation of a market freeze leads the household to accumulate assets faster so as to smooth consumption should the market indeed freeze. By doing so, however, the household feeds the credit boom, making the crisis more likely ex post. Hence the high crisis probability ex ante.

#### 3.4.3 Equilibrium

We solve the decentralized equilibrium problem numerically using standard global non-linear methods. We discretized the continuous support of  $a_t$  into a 1500-node grid, and that of  $z_t$  into a 31-node grid, and we approximated the autoregressive dynamics of productivity with a first-order Markov chain using Rouwenhorst (1995)'s method.

**Calibration** A period in the model represents a year. Parameters  $\beta$ ,  $\vartheta$ , v,  $\sigma$ ,  $\xi$ ,  $\sigma_{\varepsilon}$ ,  $\delta$  and  $\alpha$ , are standard parameters for a neoclassical growth model and we set them to their usual values (see table 3). The remaining parameters pertain to the banking sector and are the return on storage  $\gamma$ , the cost of diversion  $\theta$ , and the distribution of banks F(.). For tractability reasons we assume that  $F(p) = p^{\lambda}$ , with  $\lambda \in \mathbb{R}^+$ . The parameters of the banking sector are calibrated jointly so that when we simulate the model over 500,000 periods we obtain (i) a systemic banking crisis every forty years on average, i.e. with probability around 2.5% (see table 1), (ii) an average interest rate spread equal to 1.71%, and (iii) an average corporate loan rate of 4.35%. These latter two figures correspond to the averages observed for the US between 1990 and 2011 and, <sup>11</sup> in particular, to the interest rate on mid-size business loans as reported in the US Federal Reserve Bank's Survey of Terms of Business Lending. We thus obtain  $\gamma = 0.96$ ,  $\lambda = 24$ , and  $\theta = 0.1$ .

<sup>&</sup>lt;sup>11</sup>Because precautionary savings play an important role in our model there is a significant gap between the deterministic steady state and the stochastic steady state. It is therefore more acurate to calibrate the model based on simulations, rather than based on the deterministic steady state.

	Values	Source/Target
Discount factor	$\beta = 1/1.03$	Standard value
Risk aversion	$\sigma = 5$	Standard value
Production function	$\alpha = 0.3$	Standard value
Labour disutility	$\vartheta = 0.96$	Labour normalized to one in steady state
Frish elasticity	v = 1/3	Standard value
Standard dev. productivity shock	$\sigma_{\varepsilon} = 0.018$	Standard value
Autocorr. coeff. productivity	$\xi = 0.9$	Standard value
Capital depreciation rate	$\delta = 0.1$	Standard value
Bank distribution - $F(p) = p^{\lambda}$	$\lambda = 24$	R - r = 1.71%; SBC every 40 years; $R = 1.045$
Diversion cost	$\theta = 0.1$	R - r = 1.71%; SBC every 40 years; $R = 1.045$
Contract enforceability	$\gamma=0.96$	R - r = 1.71%; SBC every 40 years; $R = 1.045$

#### Table 3: Calibration

Based on this calibration, we obtain an interbank loan rate of 0.86% ( $\rho = 1.008$ ) in the long run and a threshold of 2.43% for the corporate loan rate ( $\overline{R} = 1.02$ ).

**Optimal Asset Accumulation Rule** The household's optimal asset accumulation rule  $a_{t+1}$   $(a_t, z_t)$  is represented in Figure 5 against the 45 degree line for four selected productivity levels, namely, for the lowest, the average, the highest productivity levels and for the level  $a_t = 1.04$ . These rules are continuous almost everywhere, except at the points when the total financial assets reach the banking sector's absorptive capacity, i.e. when  $a_t = \overline{a}_t$ . In those instances the economy switches regime.

The decision rule second from below is the one that prevails when productivity is at its average level ( $z_t = 1$ ). Its intersection with the 45 degree line in point O corresponds to the average steady state of the economy. As in the neoclassical growth model, the household smooths his consumption over time by accumulating relatively more (less) financial assets when productivity is above (below) its mean. His speed of accumulation is also all the faster when the level of productivity is high and, therefore, likely to reverse back to its mean. What is specific to our model, though, is that in such good times the household tends to accumulate even faster than in the frictionless economy. The reason is that the household also reckons with the higher odds that a banking crisis occurs, should the absorptive capacity of the banking sector fall below the level of banks assets after a fall in productivity. By adding volatility into the economy, the possibility of a crisis leads the household to accumulate for precautionary motives. As we will show next section, precautionary motives are strong in this model.

The optimal decision rules provide a first insight into the dynamics of the model. Starting from steady state O, there are two opposite ways the economy may enter a systemic banking crisis. One is if it experiences an unusually large negative technology shock, say down to point S. Because such



Figure 5: Optimal Decision Rules

a shock instantly reduces banks' absorptive capacity below the current level of assets, the crisis breaks out on impact of the shock and is, therefore directly related to the shock. The other way is to experience an unusually long spell of positive technology shocks, for example up to point U, so that total assets increase so much that, after some time, they outgrow banks' absorptive capacity.

# 4 Results

The model is simulated to provide with a first quantitative assessment of the mechanisms leading to a banking crisis. Our findings are twofold. First, we are able to replicate the three stylized facts described in section 2. In accordance with our calibration, the model generates rare systemic banking crises — on average one in forty years. The recessions coupled with banking crises are significantly deeper and more protracted than the other recessions. They are also preceded by credit intensive booms, whereas regular recessions are not. Second, we find that most banking crises break out *endogenously* after a gradual build—up of financial imbalances, reflected in excessive credit and a disproportionately large banking sector. Typically, they follow upon an *unusually long sequence of small and positive* transitory technology shocks. Only a few crises are triggered by unusually large and negative technology shocks.

#### 4.1 Typical Path to Crisis

The aim of this section is to describe the conditions under which systemic banking crises occur. As we discussed above, in the present model banking crises may a priory break out in bad as well as in good times, and it is not clear which type of shocks (i.e. negative/positive, large/small, short/long lived) are the most conducive to crises. Starting from a date t, we look for the dynamics of the economy leading to a banking crisis in year t + 40. The choice of year t + 40 is motivated by the probability of experiencing a banking crisis of about 2.5% in the data. We simulate 500,000 dynamic paths of 60 years starting from the average steady state in t (i.e.  $z_t = 1$ ), represented by point O in Figure 5. We then select among all these paths those that feature a banking crisis in year t + 40, and compute the average underlying path of the technological shock associated with these paths. This gives us the typical sequence of technology shocks leading to a crisis. We then feed the model with this sequence of shocks. Such a typical path is reported in the left panel of Figure 6. The red part of the depicted path corresponds to crisis periods, the black one is associated with normal times. One of the most striking results that emerges from this experiment is that the typical banking crisis is preceded with a long sequence of positive and relatively small technology shocks. This actually reveals one important and interesting aspect of the model: the germs of the crisis lie in productivity being above trend for an unusually long time. For example, in Figure 6 (left panel) productivity is less than 4% above trend in the 10 year run up to the crisis. The reason is that such



<u>Note:</u> Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis. Gray plain (Dashed) line : Median (Average) dynamics of  $\bar{s}_t$ . Thin dashed line: long-run average.

long sequence of shocks gives the household time to over-accumulate assets, beyond the banking sector's absorptive capacity. The one- and two- year ahead probabilities of crisis rise during the expansion phase (Figure 7) reflect the build up of these financial imbalances.<sup>12</sup> For instance, the 1-step ahead probability rises from an initial 0 to 0.75 one year before the banking crisis breaks out. Likewise, the 2-step ahead probability reaches 0.28 two years before the crisis. The corporate loan rate is above its steady state at the start of the simulation, but gradually decreases as productivity gains fade off and credit supply goes up. As the corporate loan rate diminishes, banks' incentives to divert cash increase, which erodes trust among banks. As a result, banks deleverage and the banking sector as a whole becomes less efficient. Hence the gradual rise in the interest rate spread. The crisis busts when  $R_t$  eventually reaches  $\overline{R}$ . At this point in time, banks do not have access to the interbank market anymore and have to reduce their supply of corporate loans. The spread spikes from 2% to almost 4%. It takes the banking sector three years to reduce its size below its absorptive capacity (see right panel of Figure 6), and therefore for the inter-bank market to recover. After the crisis, all financial variables return back to their steady state levels. Finally, Figure 8 illustrates the evolution of macroeconomic variables. In the run up to the crisis, the positive wealth effect associated with technological gains leads the household to both consume and invest more. The household also increases his supply of labor in order to take advantage of the high real wage. Hence, output increases too, up to more than 11% above its steady state just before the crisis. As a result, the typical crisis breaks out in the midst of unusually good times, which end abruptly with a credit crunch. The credit crunch triggers a sizeable drop in aggregate productivity, investment, and the real wage, which in turn affects labour supply, consumption, and output negatively. Three years of crisis are enough to bring the economy back down to its average steady state.

<sup>&</sup>lt;sup>12</sup>The crisis probabilities are constructed following definition 2. Hence, the 2–step ahead probability in period t reports the probability that a banking crisis will break out in period t + 2, conditional on the event that there is no crisis in period t + 1.



<u>Note</u>: Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis. Thin dashed line: long–run average.



<u>Note</u>: Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis. Thin dashed line: long–run average.

## 4.2 The Economy in Normal Times

Next, we want to analyze the properties of our model in the neighborhood of the average steady state. The idea is to show that our model behaves most of the time like the standard neoclassical growth model. To do so we report in figures 9–11 the dynamics of the economy after a 1–standard deviation positive technology shock.<sup>13</sup> Output, consumption, investment and hours worked first increase on impact, and then gradually go back to steady state. This pattern is standard and there is no systemic banking crisis. The reason is clear. The economy is started from the average steady state level, which corresponds to a business as usual situation. The corporate loan rate rises on impact, which mitigates counterparty fears on the inter-bank loan market and relaxes banks' borrowing constraints. As banks leverage up, the aggregate demand for inter-bank loans and the inter-bank loan rate both increase. The inefficient banks then switch from the demand to the supply side of the market, which works to raise borrowers' overall quality and reduces the moral hazard problem further. Notice that this leverage cycle generates a financial accelerator but, as is standard (e.g. Carlstrom and Fuerst, 1997, Bernanke, Gertler, Gilchrist, 1998), this mechanism has little impact from a quantitative viewpoint. After the first period, the corporate loan rate

 $<sup>^{13}</sup>$ Again we report the average and the median of the distribution of the dynamic path in the economy, as obtained from 500,000 simulations of the model.

gradually falls below steady state as the household accumulates assets, and eventually converges back up to steady state. Because the shock is small, at no point in the dynamics does  $R_t$  fall below  $\overline{R}$ .

Figure 9: Impulse response to a 1 Standard Deviation Technology Shock (I)



<u>Note:</u> Plain line : Average across 100,000 simulations. Dashed line: Average path of  $\overline{a}_t$  across simulations.





<u>Note:</u> Plain line : Average impulse response function across 100,000 simulations in the model, Dashed line: Average Impulse Response across simulations in the RBC model.

# 4.3 Understanding Crises

The above results reflect an important property of the model: banking crises are rare events, which only occur when specific conditions are met. The typical path to crisis described in section 4.1 is one example of such conditions. However, this path is only one among many that lead to crises. In



## Figure 11: Impulse response to a 1 Standard Deviation Technology Shock (III)

<u>Note:</u> Plain line : Average path across 100,000 simulations in the model, Dashed line: Steady state value.

this section we want to make the point that, in fact, what matters is not how imbalances build up, but rather whether they have built up. We consider a situation where the economy has reached a steady state associated with a productivity level that is 4% above its average (i.e. with  $z_t = 1.04$ ). We do not make any assumption about the sequence of shocks that led to this situation. What is important is that, at this point, the level of assets  $a_t$  is simultaneously low enough to make sure that the interbank market functions well, and high enough to make sure that imbalances have had the time to build up. This steady state, which we will refer to as the starting steady state, corresponds to point T in Figure 5. We then simulate 500,000 dynamic transition paths back to the average steady state (point O). The dynamics are reported in Figures 12–14. Notice that not all these transition paths feature a systemic banking crisis and, for those that do, crises do not all break out at the same point in the transition. Hence, to illustrate banking crises, we report not only the usual average path (dashed line), but also the median path (plain line).<sup>14</sup> The upper ( resp. lower) horizontal line refers to the starting (resp. average) steady state T (resp. O).

The starting steady state has two important features: (i) The level of financial assets is 70% higher than in the average steady state and (ii) the probability of crisis at a one year horizon is 25% (whereas it is null in the average steady state). These two features are related. The household

<sup>&</sup>lt;sup>14</sup>The mean path is the usual representation. However it averages out the effects of the financial crises, which do not always break out along the transition. This is the reason why, in the context of our model, the median path is more informative.



Figure 12: Dynamics toward a Banking Crisis (I)

<u>Note</u>: Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis, This dotted line: initial steady state, Thin dashed line: average steady state.

accumulates savings as a buffer not only against an anticipated fall in productivity back to trend, but also against a probable banking crisis in the following year. By doing so, however, he pushes the banks closer to their absorptive capacity, making the crisis indeed probable. Because the household does not internalize this adverse spiral and the fact that, through his savings decisions, he exerts negative externalities on the banking sector, he tends to over–accumulate assets. These negative externalities are the root cause of financial imbalances.

For the dynamics of the crisis, we focus on the median path. The crisis breaks out in the first period into the transition, as banks' absorptive capacity falls below banks' assets. The mechanism goes as follows. All things equal, by reducing the demand for corporate loans the fall of productivity back to its long-run average exerts a downward pressure on the corporate loan rate. Since a low corporate loan rate is detrimental to banks' incentives, counterparty fears rise. Every bank must then deleverage in order to keep access to the market. At the aggregate level, however, deleveraging reduces the demand for interbank loans and the interbank loan rate. As the latter goes down, inefficient banks switch from the supply to the demand side of the market, which spreads counterparty fears further. This adverse liquidity spiral feeds itself until the interbank market freezes, making the whole financial intermediation process less efficient. On impact, the crisis materializes itself as a sudden 10% fall in the credit to assets ratio and a dramatic 25% drop in the size of the banking sector. While the entrepreneur faces a credit crunch and a rise in the corporate loan rate, the household faces a fall in the return on bank equity.

On the real side (see Figure 14), investment falls by 20% at the beginning of the crisis.<sup>15</sup> The entrepreneur then reduces his demand for labour, which exerts pressure on the equilibrium wage, so that labour supply too goes down. The fall in both the return on bank equity and wages affects savings and consumption negatively through a standard wealth effect. Ultimately, output falls by

<sup>&</sup>lt;sup>15</sup>Part of this effect is also attributable to the technology shock that goes back to its long–run average.



<u>Note:</u> Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis, This dotted line: initial steady state, Thin dashed line: average steady state.

about 8% on impact.

The slow decrease in savings helps banks close the gap between the amount of assets they have to process and their absorptive capacity. However, the household still saves too much during the first years of the transition, expecting productivity to continue to fall and the crisis to last. Accordingly, the one-period ahead crisis probability remains above 50% most of the time during the crisis. These precautionary savings delay the adjustment, which explains why the crisis persists for so long. After ten years into the transition, the level of assets eventually passes below banks' absorptive capacity and the economy goes back to normal times.

## 4.4 Financial versus Regular Recessions

Are financial recessions in our model deeper and more protracted than regular recessions, as observed in the data (our second stylized fact)? To answer this question, we simulate 500,000 years and define as recessions the periods with negative output growth during at least two consecutive years. We define as "financial" the recessions that are concomitant with a banking crisis. Figures 15–17 show the evolution of the economy in the 4 years before and after the starting date of the recession (year 0). Again, to illustrate the dynamics better we report the median dynamic path.



Figure 14: Dynamics toward a Banking Crisis (III)

<u>Note:</u> Dark plain (Dashed) line : Median (Average) Dynamics in normal times, Red plain (Dashed) line : Median (Average) Dynamics in a systemic Banking Crisis, This dotted line: initial steady state, Thin dashed line: average steady state.

— To be completed —

### 4.5 Sensitivity Analysis

Finally, we investigate the impact of changes in the parameters on the overall properties of the model. We simulate the model for 500,000 periods and report averages of some key quantities of the model across these simulations. Results are reported in Table 2. The first column reports results for our benchmark calibration.

## 4.5.1 Risk Averse Economies Are Prone to Crises

We first vary the utility curvature parameter  $\sigma$  to from a benchmark 4.5 to 10, therefore increasing the degree of risk aversion in the economy. As we already mentioned, precautionary motives are strong in this model. By making the household more willing to accumulate assets, all things being equal, the increase in  $\sigma$  works to raise the quantity of assets banks have to process without affecting banks' absorptive capacity, and leaves banks more exposed to adverse shocks. Hence, at 5.4% the probability of a crisis is higher than in the benchmark (2.7%). In other terms, the risk averse



<u>Note</u>: Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics in a Recession Featuring a Banking Crisis, Red plain line : Average Dynamics in a Recession without Banking Crisis.



<u>Note</u>: Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics in a Recession Featuring a Banking Crisis, Red plain line : Average Dynamics in a Recession without Banking Crisis.



<u>Note</u>: Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics in a Recession Featuring a Banking Crisis, Red plain line : Average Dynamics in a Recession without Banking Crisis.

economy is paradoxically more prone to systemic banking crises. Moreover, it also experiences deeper crises than the benchmark, with output falling by 25% more from peak to trough, and crises last 1.4 years longer. The main reason is that the economy typically grows bigger, and builds up larger imbalances that make it more difficult to escape crises once they occur. These findings hold irrespective of whether the crisis is preceded by a positive or a negative technology shock. Accordingly, the risk averse economy's banking sector is also less efficient, with an interest rate spread of 2.09pp, against 1.71pp.

#### 4.5.2 Contract Enforceability and Bank Efficiency Improve Financial Stability

The third column of the table reports statistics for an economy where the cost of diversion is lower than in the benchmark (i.e.  $\theta$  is set to 0.2, instead of 0.1). The increase in the cost of diversion involves subtle general equilibrium effects. In the first place, it works to reinforce the moral hazard problem between banks, so that banks must deleverage to keep issuing inter-bank claims. Since the banking sector is less efficient, the spread goes up, and the return on bank equity goes down. As a consequence, the household dissaves, which ultimately works to reduce the supply of credit to the entrepreneur. Hence the corporate loan rate increases, as well as the inter-bank loan rate. Notice that the rise in the corporate loan rate somewhat restores banks' incentives, but this is of second order. Overall, the banking sector is less efficient and its absorptive capacity is lower, as the higher threshold for the corporate loan rate (5.65% against 2.43%) suggests. Accordingly, the probability of a crisis jumps from 2.7% to 5.92%.

A change in the distribution of banks has qualitatively similar effects. In the fourth column we consider an increase in  $\lambda$  from 24 to 40, implying a higher concentration of the distribution towards the top. The banking sector as a whole is more efficient than in the benchmark. Since efficient banks have a lower incentive to divert cash, the moral hazard problem is less stringent and counterparty fears recede. Lenders tolerate higher leverage, aggregate demand and the interbank rate rise, which crowds the less efficient banks out of the demand side of the inter-bank loan market. As a result, the crisis probability drops from 2.7% to 0.8%. When they occur banking crises are however slightly more pronounced and longer.

	<b>D</b> 1 1		0			
	Benchmark	$\sigma$	$\theta$	$\lambda$	$ ho_z$	$\sigma_z$
		10	0.20	35	0.95	0.02
Interbank rate $(\rho)$	0.86	0.23	0.65	1.34	0.75	0.89
Corporate rate $(R)$	4.35	3.70	5.50	3.70	4.10	4.32
$\overline{R}$	2.43	2.43	4.83	0.41	2.43	2.43
Spread $(R-r)$	1.71	2.09	2.89	1.03	1.92	1.77
Probability of a crisis	2.70	5.43	7.34	0.16	4.06	3.35
Average duration	2.64	4.08	5.06	1.87	4.00	2.86
Positive $z$	3.52	5.56	6.77	3.26	5.41	3.91
Negative $z$	2.05	2.25	2.49	1.58	2.39	2.05
Average amplitude	5.90	7.99	8.60	3.86	9.00	6.76
Positive $z$	8.18	10.89	11.46	9.27	12.22	9.80
Negative $z$	4.33	4.37	4.29	2.74	5.32	4.40

Table 2: Sensitivity Analysis

<u>Note</u>: Positive (resp. negative) z rows report information for the case where the crisis is preceded by a positive (resp. negative) technology shocks. All numbers, except for durations, are expressed in percents. All reported numbers are average over a long simulation of 500,000 periods. In the case where the persistence of the technology shock is raise to  $\rho_z = 0.95$ , the standard deviation of the innovation was rescaled so as to maintain the same volatility of TFP.

## 4.5.3 Uncertainty is Conducive to Crises

As the source of uncertainty, the data generating process of the technology shock too plays an important role in terms of financial stability. In columns 5 and 6 in table 2 we consider two experiments. First, we increase the volatility of the shock, leaving its persistence unchanged. The consequences are straightforward: the household accumulates more assets for precautionary motives, the corporate loan rate decreases with respect to benchmark, and the financial sector is more fragile. Next, we increase the persistence of the shock, leaving its volatility unchanged. This change has two opposite effects on financial stability. On the one hand, the household may not

need to accumulate assets as fast as in the benchmark after, say, a positive shock. On the other hand, however, he has also more time to accumulate assets. It follows that imbalances build up more slowly, but also more surely. Because what matters for financial stability is not the speed of accumulation but the time of accumulation (see section 4.1), overall the probability of a crisis increases.



<u>Note:</u> Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics around a Banking Crisis, Red plain line : Average Dynamics around a Banking Crisis triggered by Positive Technology Shock, Dark Dashed Line: Average Dynamics around a Banking Crisis triggered by Negative Technology Shock



Figure 19: Dynamics around a Banking Crisis (II)

<u>Note:</u> Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics around a Banking Crisis, Red plain line : Average Dynamics around a Banking Crisis triggered by Positive Technology Shock, Dark Dashed Line: Average Dynamics around a Banking Crisis triggered by Negative Technology Shock



<u>Note:</u> Period t = 0 corresponds to the period when the Banking Crisis breaks out. Dark plain line : Average Dynamics around a Banking Crisis, Red plain line : Average Dynamics around a Banking Crisis triggered by Positive Technology Shock, Dark Dashed Line: Average Dynamics around a Banking Crisis triggered by Negative Technology Shock

# 5 Conclusion

To be written...