Booms and Banking Crises[*]

Frédéric Boissay[†]   Fabrice Collard[‡]   Frank Smets[§]
European Central Bank   University of Bern   European Central Bank

This draft: October 1, 2013
First draft: December 2011

Abstract

Banking crises are rare events that break out in the midst of credit intensive booms and bring about particularly deep and long–lasting recessions. This paper attempts to explain these phenomena within a textbook DSGE model that features a non–trivial banking sector. In the model, banks are heterogeneous with respect to their intermediation skills, which gives rise to an interbank market. Moral hazard and asymmetric information in this market may lead to sudden interbank market freezes, banking crises, credit crunches and, ultimately, severe recessions. The model can potentially generate various types of banking crises. But the typical crisis breaks out endogenously, during a credit boom generated by a sequence of small positive supply shocks; it does not result from a large negative exogenous shock. Simulations of a calibrated version of the model indicate that it can mimic the main dynamic patterns of banking crises.

Keywords: Moral Hazard, Asymmetric Information, Saving glut, Lending Boom, Credit Crunch, Banking Crisis

JEL Class.: E32, E44, G01, G21.

Disclaimer: The views expressed in this paper are our own and should not be interpreted as reflecting the views of the European Central Bank or the Eurosystem.

1 Introduction

Recent empirical research on banking crises has highlighted the existence of similar patterns across diverse episodes. Banking crises are rare events. They come along recessions that are deeper and last longer than other recessions. And, more importantly for the purpose of this paper, banking crises follow credit intensive booms; they are "credit booms gone wrong".[1] In the recent macro–economic literature, banking crises mostly result from the propagation and the amplification of random adverse financial shocks. Rare, large enough, adverse financial shocks can account for the first two properties (see *e.g.* Gertler and Kiyotaki, 2009). However, by implying that banking crises may break out at any time in the business cycle, they do not seem in line with the fact that the occurrence of a banking crisis is closely linked to credit conditions (Gorton, 2010, 2012). The third stylized fact therefore calls for an alternative approach.

In this paper, banking crises result from the procyclicality of bank balance sheets that emanates from interbank market funding. During expansions, bank market funding and credit supply increase, pushing down the rates of return on corporate and interbank loans. The lower rates aggravate agency problems in the interbank market, which lead to a reduction in market funding and contractions. The larger the credit boom relative to the possibilities for productive use of loans, the larger the fall in interest rates, and the higher the probability of a bank run in —and therefore of a disastrous freeze of— the interbank market.[2] As in Shin (2008) and Hahm et al. (2011), the behavior of banks (credit in our case) during good times sows the seeds of the next banking crisis.

In our model, banks are heterogeneous in terms of —non–publicly observed— intermediation efficiency. They finance their activities with funds obtained from depositors/shareholders or raised in the interbank market. There exists the usual agency problem in this market as borrowing banks can always divert some of the funds into low return assets that cannot be recovered by the lending banks. The incentives for diversion are stronger for less productive banks and depend on the level of interest rates in the economy. The lower the return on corporate loans, the greater the incentive to engage in fund diversion and hence the greater counter–party risk in the interbank market. The typical run of events leading to a banking crisis is as follows. A sequence of favorable, non permanent, supply shocks hits the economy. The resulting increase in productivity leads to an expansion of credit. The more efficient

---

[1]See Schularick and Taylor (2012), Borio and Drehmann (2009); the notion that banking crises are endogenous and follow prosperous times is also present in Minsky (1977).

[2]Our representation of banking crises as market–based bank runs is consistent with what happened during the 2007–8 financial crisis (see Uhlig, 2010). Shin (2010, Chap. 8), for example, depicts the demise of Northern Rock —a UK bank— in 2007 as primarily originating from the sudden freeze of the short–term funding market, what he refers to as a "modern bank run". A traditional, deposit–based, run on the bank took place as well, but it did so one month later, accounted for only 10% of the bank's fall in total funding, and rapidly stopped because, following the news of the run, the UK authorities pledged 100% deposit guarantees.

banks expand their corporate loan operations by drawing funds from the less efficient banks and the size of the banking sector as a whole increases. The economy booms. But as supply shocks run their course, the probability of imminent reversion to average productivity increases. This slows down corporate demand for loans while at the same time inducing households to accumulate savings in order to smooth consumption. As this saving glut develops, interest rates go down. The rate of return on interbank loans declines making the less efficient banks more prone to borrow and divert those funds. The identity of these banks being unknown to lenders, counter–party risk in the interbank market increases, interbank loans decline, and market finance recedes. The stronger the credit expansion during the booming times, the larger the decline in interest rates and the sharper the agency problem in the interbank market. We show that there is a threshold value of interest rates below which the interbank market freezes, corporate credit collapses and the economy tanks. This threshold can be alternatively expressed in terms of the level of banking assets relative to the level of productivity (output) in the economy, which we call the absorption capacity of the banking sector. Excessive credit creation shifts the economy beyond its absorption capacity, triggering a banking crisis.

There have been many attempts to analyze banking crises within otherwise standard macro–economic models. For instance, Bernanke et al. (1999), Jermann and Quadrini (2012), Christiano et al. (2013), Gertler and Karadi (2011) show how adverse wealth or financial shocks can be amplified by financial market frictions and generate deep and long lasting recessions. We depart from these approaches in that we do not analyze a linearized version of our model but instead characterize the full equilibrium dynamics, inclusive of important and critical non-linearities such as the freezing of interbank markets. This is an important difference because, near the steady state, our model features the traditional financial accelerator that the previous models possess. But away from it (and the large departures from the steady state are the endogenous outcome of a boom-bust endogenous cycle, rather than a big shock) it gives rise to banking crises. Crises are rare but generate particularly large output losses and inefficiencies due to the presence of pecuniary externalities. In this respect, our work relates to the sudden stops literature and the influential work by Mendoza and Smith (2005), Mendoza (2010), Bianchi (2011), Bianchi and Mendoza (2010), which also feature non–linearities and externalities. One important difference with sudden stops models is that they typically consider small open economies, in which the interest rate is exogenous, whereas in our case the endogenous dynamics of interest rates plays a central role. Furthermore, in these models, financial frictions affect non–financial firms and entail excess credit demand (*i.e*, "over–borrowing"). In our case, they affect the banking sector and generate excess credit supply. Our model is also close to Brunnermeier and Sannikov (2012) and He and Krishna-murthy (2012) in that it features powerful non–linear amplification mechanisms. The novelty of our approach is that a banking crisis does not necessarily result from an adverse exogenous

shock, and the size and sign of the shock play a less important role. When the size of the banking sector is large with respect to aggregate output, it is even possible that crises break out endogenously, in the absence of shocks. Another noticeable feature is that we abstract from financial shocks, which are not necessary in the model to generate banking crises. Finally, our model departs from the existing literature in that banking crises are not due to occasionally binding collateral constraints. They are due to occasional financial market runs, which we find to be particularly disruptive and more costly in terms of welfare than the usual collateral constraints.[3] The welfare losses due to the externalities are for example three times as large in our model as in Bianchi (2011).

We calibrate the model on post–WWII US data and the financial cycles in fourteen OECD countries (1870-2008), and assess its quantitative properties. Most of the time bank assets remain below the threshold for banking crises and the model behaves like a standard financial accelerator model. But once in a while —on average every forty years— there is a banking crisis. The model can potentially generate a variety of banking crises but the simulations indicate that the typical crisis breaks out endogenously, during a credit boom generated by a sequence of small positive supply shocks. Thus, as in the data, banking crises are closely linked to credit conditions; they also bring about more severe and longer–lasting recessions.

The paper proceeds as follows. Section 2 documents stylized facts about financial recessions in 14 OECD countries for the period 1870–2008. Section 3 describes our theoretical framework, the micro-foundations of interbank market freezes and the dynamic implications of such events. In Section 4 we solve a calibrated version of the model and describe the typical dynamics to banking crises. Section 5 discusses the ability of the model to account for the stylized facts and the sensitivity of the results to alternative parameterizations. Section 6 evaluates the various welfare costs associated with the existence of financial frictions. A last section concludes.

2 Facts on Financial Recessions

The recent empirical literature has put forward a set of facts that seem to indicate that financial recessions — recessions that are concomitant with banking crises — are special events that significantly differ from other types of recessions. Reinhart and Rogoff (2009) report that banking crises have historically been followed by profound declines in output and employment, with output falling by about 9% from peak to trough and the duration of the downturn averaging roughly two years. This is much more than during ordinary recessions, possibly because financial recessions come along with a credit crunch that amplifies

---

[3]Gertler and Kiyotaki (2013) also develop a model with bank runs and regime switches. While in their model bank runs are deposit–based and unanticipated, in ours they are market–based and agents are fully rational: they perfectly know and take into account the probability that runs will occur in the future.

the downturn, as Claessens et al. (2008, 2011) show. Mendoza and Terrones (2012) and Schularick and Taylor (2012), among others, also provide evidence that banking crises follow credit booms. For instance, Schularick and Taylor (2012) show that high credit growth is a robust predictor of future banking crises in the case of advanced economies. The predictive power of credit may not necessarily reflect a causal relationship, as credit is strongly correlated with other financial phenomena that have also been found to precede banking crises like asset price bubbles or large capital inflows (Kaminski and Reinhart, 1999; Reinhart and Rogoff, 2008, 2009). However, these other factors seem to be more relevant for emerging market economies than for advanced countries (Reinhart and Reinhart, 2009) where credit remains a strong predictor even once past current account deficits, asset prices, or GDP growth are controlled for (Jordà, Schularick and Taylor, 2011a,b; Schularick and Taylor, 2012). The different evolution of the GDP and credit gaps around financial and non–financial recessions shown in Figure 1 illustrates these facts, which we document in more detail below using the data compiled by Schularick and Taylor (2012) and Jordà, Schularick and Taylor (2011a,b).

Figure 1: Dynamics of GDP and credit gaps around recessions
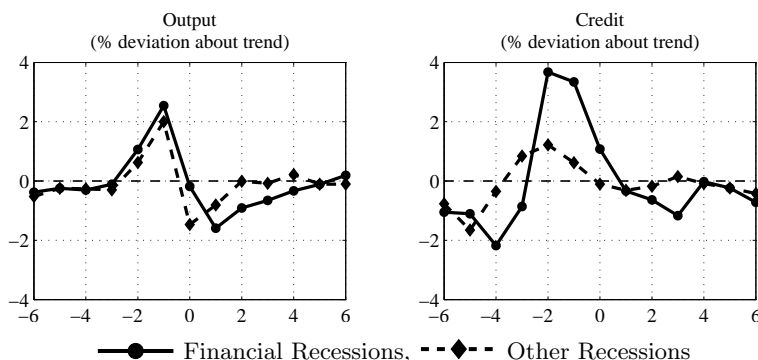
Note: Average dynamics of the Hodrick-Prescott ($\lambda = 6.25$) cyclical component of (log) output and credit 6 years before and after the start of a business cycle (period 0) with and without a banking crisis. The data are from Schularick and Taylor (2012) and Jordà, Schularick and Taylor (2011a,b).

The data set comprises yearly observations from 1870 to 2008 in 14 developed countries for real GDP per capita, total domestic currency loans of banks and banking institutions to non–financial companies and households, the dates of business cycle peaks, and the dates of banking crises. Following Reinhart and Rogoff (2008) and Laeven and Valencia (2008), a banking crisis in the data refers to an event during which the financial sector experiences bank runs, sharp increases in default rates accompanied by large capital losses that result in public intervention, bankruptcy, or the forced merger of major financial institutions. As is common in the literature, we define a recession as a *financial* recession when a banking crisis breaks out between the peak year that precedes it and trough year that ends it (see Jordà, Schularick and Taylor, 2011b). Excluding war times, the data set includes 78 banking crises for a total of 1,736 country/year observations, and includes 161 complete business cycles — *i.e.* periods between two business cycle peaks. We use these data and describe three key

facts on financial recessions.

Table 1: Statistics on recessions

	Financial	Other	All	Severe	Mild
N. country/year obs	1,736	1,736	1,736	1,333	1,333
N. events	41	155	196	54	55
Frequency (%)	2.36	8.93	11.29	4.05	4.13
Duration (years)	2.32***	1.65	1.79	2.46***	1.25
Magnitude ($\Delta_{p,t}y$, %)	-6.84***	-3.75	-4.40	-9.28***	-0.89
Credit crunch					
$\Delta_{p,t}k^{HP}$	-3.94*	-2.09	-2.48	-1.48	-2.67
$\Delta_{p,p+2}k^{HP}$	-3.59*	-1.24	-1.75	-1.69	-2.44
Credit boom					
$\Delta_{p-2,p}k^{HP}$	4.56***	0.01	0.98	1.33	0.40
k_p^{HP}	3.34***	0.62	1.19	1.89	0.99

Note: $\Delta_{p,t}x$ (resp. $\Delta_{p,p+2}x$, $\Delta_{p-2,p}x$) denotes the percentage change of variable x from peak to trough (resp. from peak to peak+2 years, and peak-2 years to peak), where x denotes either GDP per capita (y) or the Hodrick Prescott ($\lambda = 6.25$) cyclical component of credit per capita (k^{HP}). A severe (mild) recession is a recession associated with a peak-to-trough output loss in the top (bottom) three deciles of the distribution. All statistics are averages over the full sample off 1,736 country/year observations, except the peak–to–trough statistics, which are averages over the 161 complete business cycles only (1,333 country/year observations). *,**,*** : the difference between financial (resp. severe) recessions and other (resp. mild) recessions is statistically significant at the 10%, 5%, 1% thresholds.

Fact #1: Financial recessions are rare events. Among the 196 recessions identified in the sample, 41 are associated with a banking crisis, which corresponds to an average probability of a financial recession of 2.36%. Hence, financial recessions are rare events; by comparison, the probability of other types of recession is 8.93% (see Table 1).

Fact #2: Financial recessions are deeper and last longer than other recessions. In our sample, financial recessions last on average 2.32 years, about eight months longer than other recessions. During these episodes real GDP per capita falls almost twice as much, from peak to trough, than during other recessions (6.84% against 3.75%). These differences are statistically significant at the 1% level and, importantly, they are unlikely to be due to severe recessions causing non–performing loans, bank capital losses, and ultimately banking crises. Indeed, should it be the case, we would observe that banking crises tend to break out *after* the recessions start, once bank capital losses have accumulated. In our sample, we find the opposite: in 32 out of 41 financial recessions the banking crisis breaks out either in the year *before* or in the very first year of the recession. Table 1 also shows that the credit gap falls by more during financial recessions (-3.94%) than during other recessions (-2.09%). We find similar differences when we look at the level of credit (not reported): during financial recessions credit falls by 1.24%, which is unusual as during recessions credit growth usually

decelerates but remains positive (see, *e.g.*, Claessens et al., 2008). Altogether, these results suggest not only that financial recessions are accompanied with a credit crunch but also that, in the case of financial recessions, it is the severity of the recession that results from financial distress —and not the other way around.

Fact #3: Financial recessions follow credit booms. Banking crises, and *a fortiori* financial recessions, do not hit at random (Gorton, 1988) but instead break out in the midst of a credit boom (Schularick and Taylor, 2012). The bottom of Table 1 illustrates this point. It shows that credit grows significantly faster before financial recessions than before other recessions (4.56% against 0.01%). Credit is also 3.34% above its trend in the peak year that precedes a financial recession, whereas it is only 0.62% above trend in the case of other recessions. All these differences are statistically significant at the 1% threshold. In contrast, there is no such difference between mild and severe recessions (last two columns of Table 1). In other words, only financial recessions are associated with credit booms, which suggests that a country's recent history of credit growth helps predict a financial recession. To test

Table 2: Prediction of recessions

	Banking crisis	Recession		
	ST model (15)	all	financial	severe
Sum of lag credit coefficients	19.32*	0.20	15.39*	-2.66
N. obs	1,272	1,272	1,272	1,078
N. groups	14	14	14	12
N. events	78	196	41	58
Pseudo R^2	0.123	0.053	0.127	0.099

Note: All models include country fixed effects and the lagged growth rates of real credit and real GDP per capita (the coefficients are reported in Section D.1 of the companion technical appendix). First column: replication of Schularick and Taylor (2012, page 1052, Table 6, model (15)); the dependent variable is a dummy equal to one in the country/year where a banking crisis breaks out. Other models: the dependent variable is a dummy equal to one in the country/year where a recession starts. Last column: as in Table 1 a severe recession is a recession associated with a peak-to-trough growth rate of real GDP per capita in the bottom three deciles of the distribution; Denmark and Germany are excluded because of the fixed effects. Standard errors in parentheses. *: Significant at the 1% thresholds.

this idea formally we build upon Schularick and Taylor (2012) and estimate the probability that a recession starts in a given country in a given year as a function of the country's lagged real credit growth, and check whether credit indeed matters. To control for potential omitted variables and reverse causality we also include five lags of real GDP growth in the model. The results are reported in Table 2. For comparison, the first column replicates Schularick and Taylor's logit model for banking crises; the other three columns correspond to our model for all types of recessions, as well as for financial and severe recessions separately. The results essentially confirm the statistics in Table 1. The sum of the lag coefficients is positive and significant in the case of financial recessions only. Credit helps predict financial recessions,

but it is of no use to predict other types of recessions, be they severe or not. We conclude that financial recessions are indeed special, that they are credit booms gone bust. This result is important because it establishes the role of credit as an endogenous source of financial instability —and not merely as an amplifier of exogenous shocks.

3   The Model

We consider a closed economy populated with one representative risk averse household, one representative risk neutral competitive firm, and a mass one of heterogeneous, risk neutral, and competitive banks.

3.1   The Representative Firm

The representative firm lives for one period. It produces a homogeneous good that can be either consumed or invested by means of capital, k_t, and labor, h_t, according to a constant returns to scale technology represented by a production function $z_t F(k_t, h_t)$ that satisfies standard Inada conditions. The level of total factor productivity (TFP), z_t, is assumed to follow an exogenous AR(1) process of the form

$$\log z_t = \rho_z \log z_{t-1} + \varepsilon_t,$$

where $|\rho_z| < 1$ and ε_t is an normally distributed innovation with zero mean and standard deviation σ_z. Variations in productivity are the only source of uncertainty and ε_t is realized at the beginning of period t, before the firm decides on its production plan. Capital, k_t, depreciates at rate $\delta \in (0, 1)$. The firm is born with no resources and must borrow k_t from the banks at a gross corporate loan rate R_t at the beginning of the period. The corporate loan is repaid at the end of the period. The firm also rents labor services from the household at rate w_t. The production plan is decided so as to maximize profit

$$\pi_t = z_t F(k_t, h_t) + (1 - \delta)k_t - R_t k_t - w_t h_t. \tag{1}$$

3.2   The Representative Household

The economy is populated by an infinitely lived representative household that has preferences over consumption, c_t, represented by the utility function:

$$\mathbb{E}_t \sum_{\tau=0}^{\infty} \beta^\tau u\left(c_{t+\tau}\right), \tag{2}$$

where $u\left(c_t\right)$ satisfies the usual regularity conditions (*i.e.*, $u'(c) > 0$, $u''(c) < 0$, $u'(0) = \infty$, $u'(\infty) = 0$), $\beta \in (0, 1)$ is the psychological discount factor, and $\mathbb{E}_t\left(\cdot\right)$ denotes the expectation

operator which is taken over $\{\varepsilon_{t+\tau+1}\}_{\tau=0}^{+\infty}$. For expositional convenience, we assume that the household supplies inelastically one unit of labor; we will relax this assumption in the quantitative analysis of the model (see Section 4.1). The household enters period t with assets a_t, which yield a state contingent gross return r_t. There is no friction between the household and the banks, which implies that banks' funding structure and leverage will be indeterminate. Hence, bank leverage will not be a source of financial fragility in our model.[4] Since the composition of a_t is indeterminate, a_t can be thought of either as bank deposit or as bank equity. To ease exposition, we will refer to it as bank deposit, and to r_t as the gross return on bank deposit. Following the macro–finance literature (*e.g.* Gertler and Kiyotaki, 2009; Gertler and Karadi, 2011) we assume that there exist frictions between the household and the firm that prevent the household from financing the firm directly. The household earns unit wage w_t from supplying her labor, receives profits π_t from the firm, and gets a lump–sum transfer χ_t that corresponds to banks' financial intermediation costs (see next section). She therefore decides on her consumption/saving decisions maximizing her utility (2) subject to the budget constraint

$$c_t + a_{t+1} = r_t a_t + w_t + \pi_t + \chi_t. \tag{3}$$

3.3   The Banking Sector

The banking sector is at the core of the model and plays a non–trivial role because of two specific features. First, banks are heterogeneous with respect to their intermediation technology — some banks are more efficient than others, which potentially gives rise to an interbank market. Second, the banking sector is subject to both asymmetric information and moral hazard problems, which impair the functioning of the interbank market. Overall, banks engage in two types of activities. They make traditional retail banking operations, which consist in collecting deposit from households and in lending the funds to the firm. Banks also make wholesale banking operations by issuing interbank claims, through which they reallocate assets among themselves —typically from the least efficient to the most efficient banks.

3.3.1   Banks

There is a continuum of one–period, risk–neutral, competitive banks that raise deposits a_t from the household at the end of period $t-1$. The banks that operate in period t are born at the end of period $t-1$ and die at the end of period t.[5] When they are born, banks are

---

[4]For an extension with endogenous bank leverage, see Section E.2 of the companion technical appendix.

[5]We will assume in a moment that banks are heterogeneous and that their types are private information. The assumption of one–period living banks is made to preserve this asymmetry of information over time. An alternative —but equivalent— approach would be to allow banks to live infinitely and, in order to rule out potential reputation effects, to assume that the types are randomly drawn afresh every period.

all identical and raise the same amount of deposits, a_t. They then draw a random, bank–specific, intermediation skill and become heterogeneous. Let p denote the bank with skill p, and assume that the ps are distributed over interval $[0, 1]$ with cumulative distribution $\mu(p)$, satisfying $\mu(0) = 0$, $\mu(1) = 1$, $\mu'(p) > 0$. Bank p must pay an intermediation cost $(1 - p)R_t$ per unit of corporate loan at the end of the period, so that the net unit return of the loan is pR_t. These costs are rebated to the household in the form of lump–sum transfers that amount to χ_t, so that there is no dead–weight loss in the economy. These intermediation costs reflect the prospection, screening, and monitoring costs that banks face either as loan originators or as loan servicers.[6] Banks also have, as outside option, the possibility to invest funds in their own project. We assume that this project yields a constant gross return γ per unit of good invested and is worth more than just letting the good depreciate (*i.e.*, $\gamma \geqslant 1 - \delta$). To fix ideas, we will refer to this outside option as storage.

Bank heterogeneity gives rise to an intra–periodic interbank market, where the least efficient banks lend to the most efficient ones at gross rate ρ_t. This rate is the same for all borrowers; otherwise the borrowers that promise the lowest returns would not attract any lender. Moreover, in equilibrium ρ_t must be strictly lower than R_t; otherwise no bank would lend to the firm and the Inada conditions on the firm's production function would imply that R_t is infinite. Similarly, ρ_t must be bigger than γ; otherwise no bank would lend to other banks. It follows that, in an equilibrium, storage is an inefficient technology as it must be the case that $\gamma < R_t$. Banks take ρ_t and R_t as given. Given these rates, bank p decides whether, and how much, it borrows or lends. By convention, we will call the banks that supply funds on the interbank market the "lenders" and those that borrow the "borrowers". Let ϕ_t be the — endogenous and publicly observable — amount borrowed on the interbank market per unit of deposit by a borrower p, with $\phi_t \geqslant 0$. Since ϕ_t is the ratio of market funding to traditional funding, we will refer to it as banks' "market funding ratio". Note that one cannot interpret ϕ_t as a leverage ratio because (i) market funding is only one of the two components of bank liabilities —the other one being deposits— and (ii) the bank deposit structure is indeterminate. Let $r_t(p)$ be bank p's gross return on deposit, with

$$r_t(p) \equiv \max\left\{pR_t(1 + \phi_t) - \rho_t\phi_t, \rho_t\right\}. \tag{4}$$

It is equal to $pR_t(1 + \phi_t) - \rho_t\phi_t$ when bank p borrows $\phi_t a_t$ from other banks at cost ρ_t and lends $(1 + \phi_t)a_t$ to the firm for return pR_t. It is equal to ρ_t when, instead, bank p decides to

---

[6]The way we model these costs is not critical, in the sense that we would find qualitatively similar results (i) if they were dead–weights (see Boissay et al., 2013), (ii) if they were paid at the beginning of the period, (iii) if p were the fraction of the loan that the firm repays to bank p, or (iv) if p were the probability that the firm reimburses bank p (see Boissay, 2011). We do not introduce these features in the model so as to stay the closest possible to the textbook neoclassical model, where all deposit are channeled to firms and firms do not default.

lend the deposit it collected to other banks. Bank p chooses to be a borrower when

$$pR_t (1 + \phi_t) - \rho_t \phi_t \geqslant \rho_t \iff p \geqslant \bar{p}_t \equiv \frac{\rho_t}{R_t}. \qquad \text{(PC)}$$

Inequality (PC) is the participation constraint of bank p to the interbank market as borrower, rather than as lender, and pins down the type of the marginal bank $\bar{p}_t$ that is indifferent between the two options. Banks with $p < \bar{p}_t$ delegate financial intermediation to more efficient banks with $p \geqslant \bar{p}_t$. In a frictionless world, all banks with $p < 1$ would lend to the most efficient bank, so that $\bar{p}_t = 1$. This bank would then have an infinite market funding ratio ($\phi_t \to +\infty$) and the economy would reach the First Best allocation. In this model, two frictions on the interbank market —moral hazard and asymmetric information— prevent the economy from reaching the First Best allocation.

Moral Hazard: We assume that the proceeds of the storage technology are not traceable and cannot be seized by creditors. As a consequence, interbank loan contracts are not enforceable, banks may renege on paying their interbank debt and walk away with the funds. A bank that walks away with $(1 + \phi_t)a_t$ can invest in the storage technology and gets $\gamma(1 + \theta\phi_t)a_t$ as pay–off, with $\theta \in [0, 1]$. Following the current practice (*e.g.* Hart, 1995, Burkart and Ellingsen, 2004), we refer to such an opportunistic behavior as "diversion". From a corporate finance literature viewpoint (*e.g.* Tirole, 2006), this is a standard moral hazard problem: (i) the gain from diversion increases with ϕ_t, (ii) the opportunity cost of diversion increases with bank efficiency p and (iii) with the corporate loan rate R_t. Features (i) and (ii) imply that efficient banks with "skin in the game" are less inclined to walk away than highly leveraged and inefficient banks. Feature (iii) is similar to feature (ii), but in the "time–series" (as opposed to "cross–sectional") dimension; it implies that banks are more inclined to walk away when the return on corporate loans is low. This latter feature captures recent empirical evidence that banks tend to take more risks when interest rates are low (Maddaloni and Peydro, 2011).

Asymmetric Information: The ps are privately known so that lenders can neither observe *ex ante* nor verify *ex post* the borrowers' skills, and therefore ignore the borrowers' private incentives to divert funds. As a result, the loan contracts signed on the interbank market are the same for all banks. Neither ϕ_t nor ρ_t depends on p.[7]

By limiting the borrowing capacity of the most efficient bank ($p = 1$), moral hazard will give less efficient banks room to borrow; hence a positive wedge between R_t and r_t. Moral

---

[7]To see this, consider 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 $\rho_t(\tilde{p}) = \rho_t \; \forall \tilde{p} \in [0, 1]$. It is easy to see that such a menu of contracts cannot be revealing because any borrower p (*i.e.* with $pR_t > \rho_t$) claiming being of type $\tilde{p}$ would 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 $\phi_t(\tilde{p})$, independent of its type. It is equally easy to see 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 retained earnings. Then the net profit of bank p would be $\eta_t(\tilde{p})(1 + \phi_t(\tilde{p})) pR_t$ and, in equilibrium, this bank would pick the contract that yields the highest $\eta_t(\tilde{p})(1 + \phi_t(\tilde{p}))$, independently of its own p.

hazard is not enough to generate market freezes though. Uncertainty about the quality – and therefore some selection– of borrowers is also needed. Hence, both moral hazard and information asymmetry will be necessary to generate banking crises in the model.

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 with $p < \overline{p}_t$ —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 ϕ_t, which can therefore also be interpreted as lenders' funding tolerance, *i.e.* the market funding ratio above which a bank refuses to lend or, in Holmström and Tirole's language, the borrower's pledgeable income.[8] The program of bank $p \geqslant \overline{p}_t$ thus consists in maximizing its return on equity $r_t(p)$ (see relation (4)) with respect to ϕ_t subject to constraint (IC). Proposition 1 follows.[9]

Proposition 1 (Optimal Market Funding Ratio) *The IC constraint binds at the optimum of the borrowing bank p, which thus exhausts its borrowing capacity:* $\phi_t = \frac{\rho_t - \gamma}{\gamma \theta}$.

The positive relationship between ϕ_t and ρ_t is a critical feature of the model. When ρ_t decreases, the net present value of corporate loans rises and the least efficient banks enter the demand side of the market. Since the private incentives to divert are higher for those banks, lenders tolerate a lower market funding ratio (ϕ_t goes down). In the limit case when $\rho_t = \gamma$, there is no demand for interbank loan because no borrower can commit to repay. The market funding ratio ϕ_t and the type of the marginal bank $\overline{p}_t$ fully describe banks' optimal decisions.

3.3.2 Interbank Market

The equilibrium of the interbank market is characterized by the gross return ρ_t that clears the market. We look for an equilibrium where $\rho_t > \gamma$ so that $\phi_t > 0$ and trade takes place. Since a mass $\mu\left(\overline{p}_t\right)$ of banks lend a_t, the aggregate supply of funds is equal to $\mu\left(\overline{p}_t\right) a_t$. Likewise a mass $1 - \mu\left(\overline{p}_t\right)$ of banks borrow $\phi_t a_t$, and aggregate demand is equal to $\left(1 - \mu\left(\overline{p}_t\right)\right) \phi_t a_t$. The market clears when (using relations (PC) and Proposition 1):

$$\underbrace{a_t \mu\left(\frac{\rho_t}{R_t}\right)}_{\text{supply}} = \underbrace{a_t \overbrace{\left(1 - \mu\left(\frac{\rho_t}{R_t}\right)\right)}^{\text{extensive margin}} \overbrace{\frac{\rho_t - \gamma}{\gamma \theta}}^{\text{intensive margin}}}_{\text{demand}} \iff R_t = \Psi(\rho_t) \equiv \frac{\rho_t}{\mu^{-1}\left(\frac{\rho_t - \gamma}{\rho_t - \gamma(1-\theta)}\right)}. \tag{5}$$

[8]One could indeed recast the moral hazard problem into a setup *à la* Holmström and Tirole (1997), whereby borrowers may misuse the funds and enjoy private benefits at the expense of their creditors. *Stricto sensu*, the pledgeable income is the highest income that can be pledged without jeopardizing the borrower's incentives, *i.e.* $\rho_t(\rho_t - \gamma)a_t/\gamma\theta$.

[9]The proofs of propositions are reported in Appendix A.

Aggregate supply increases monotonically with ρ_t, whereas aggregate demand is driven by two opposite forces. On the one hand, aggregate demand decreases with the interbank loan rate because fewer borrowers demand funds when the cost of funds increases; this is the "extensive margin" effect. On the other hand, a rise in ρ_t also exerts a positive effect on aggregate demand because each borrower is then able to borrow more; this is the "intensive margin" effect. At the aggregate level, this latter effect more than offsets the extensive margin effect when the marginal bank's externality affects a large mass of borrowers, *i.e.* when ρ_t is small enough. It follows that the aggregate demand curve bends backward, increasing with ρ_t for small values of ρ_t (see Figure 2). One can check that $\Psi(\rho_t)$ is strictly convex, goes to infinity as ρ_t approaches γ, is greater than R_t when ρ_t approaches R_t, and reaches a minimum for some value $\rho_t = \overline{\rho} > \gamma$. Hence there exists a threshold $\overline{R} \equiv \Psi(\overline{\rho})$ for R_t below which there is no equilibrium with trade, and above which there are two equilibria with trade. This threshold is the minimum corporate loan rate that is necessary for the banks to accept to lend to each other. Figure 2 illustrates this point and depicts the shifts in aggregate supply and demand as R_t falls from R_H (equilibrium E) to R_L (equilibrium A), with $R_L < \overline{R} < R_H$. Following the fall in the corporate loan rate, the supply curve shifts to the right while the demand curve shifts to the left. Given the initial equilibrium rate $\rho_t = \rho_E$, demand falls below supply. Market clearing then requires that ρ_t goes 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 private incentive to divert cash, lenders require borrowers to reduce their demand for loan. By construction, this intensive margin effect is the strongest when $R_t < \overline{R}$. It follows that, ultimately, aggregate demand decreases and excess supply goes further up. This process feeds itself and goes on until the market freezes, in point A, where by convention we set $\rho_t = \gamma$. Notice that this type of market freeze is reminiscent of Mankiw (1986)'s credit market collapses. In Mankiw (1986), changes in the interest rate also alter the composition and the quality of the pool of borrowers. One important difference, though, is that, in our case, banks can be on either side of the loan market and choose which side they stand on. So, following an increase in the interest rate low quality borrowers switch to the supply side and the average quality of the remaining borrowers improves, whereas in a typical adverse selection model, like Mankiw's, quality would deteriorate. This peculiar borrower selection mechanism explains why low —not high— interest rates are conducive to market freezes. This point is very important because low interest rates will be generated by an increase in the supply of credit by banks, and therefore by a supply–driven credit boom, which will ultimately explain why, in our model, banking crises occur after a boom in credit.

When $R_t \geqslant \overline{R}$ the aggregate supply and demand curves intersect at points U and E. In this case it seems reasonable to restrict attention to E and to rule out U as not tatônnement–stable (see Mas–Colell et al., 1995, Chapter 17.H). At this point net aggregate demand is

Figure 2: Interbank market clearing
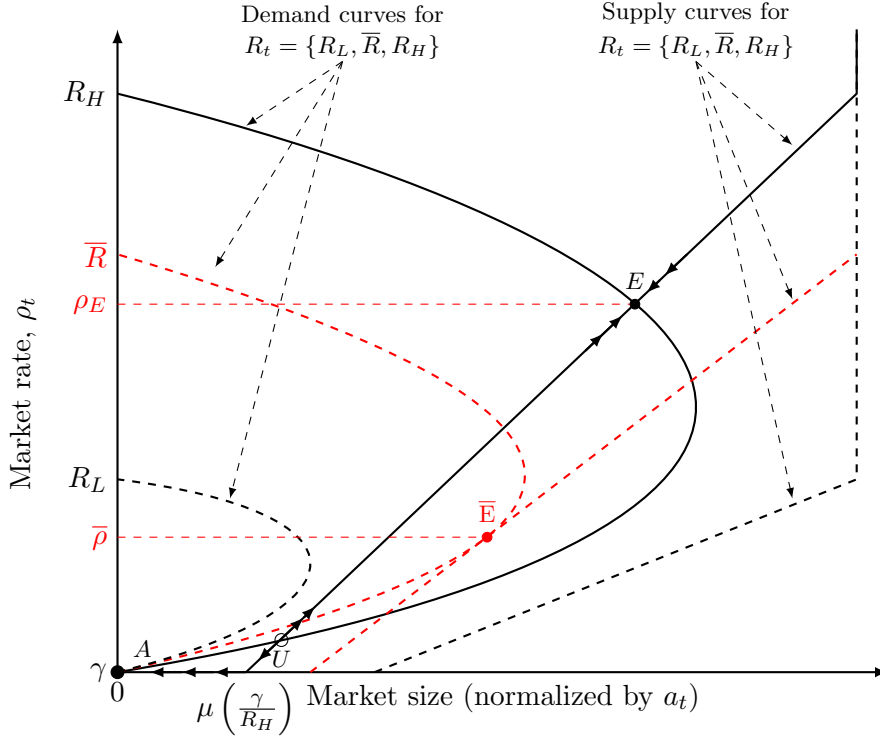
indeed a decreasing function of ρ_t, and following any small perturbation to ρ_t a standard price tatônnement process would push ρ_t away from ρ_U. Similar reasoning shows that point E is a stable equilibrium. Since in this equilibrium trade takes place, we will refer to E as *normal times*.

In equilibrium A, in contrast, no trade takes place. Both aggregate loan demand and supply are equal to zero because borrowers have no pledgeable income ($\phi_t = 0$) and lenders are indifferent between interbank loans and storage. In this case, there exists a cut–off $\overline{p}_t = \gamma/R_t$ such that banks with $p < \overline{p}_t$ store, banks with $p > \overline{p}_t$ are better off lending to the firm, and bank $\overline{p}_t$ is indifferent. We will refer to such no–trade equilibrium as a *banking crisis*.[10] This equilibrium always exists —irrespective of R_t, is stable, and therefore co–exists with E whenever E exists. Because it is Pareto–dominated by E, though, we rule it out in this case by assuming that banks always coordinate on E.[11]

---

[10]As will be clear later, this situation corresponds to a regime switch, in which the economy experiences a large drop in output. Regime switching can be thought of as a way to endogenize rare exogenous shocks to disaster risk (*e.g.*, see Gourio, 2012) or to generate non–linear dynamics like in catastrophe theories (*e.g.*, see Varian (1979) for an introduction).

[11]Two comments are in order. First, note that when $R_t \geqslant \overline{R}$ we could possibly generate sunspot equilibria by randomizing between outcomes A, U, and E, with a sunspot variable coordinating the beliefs of the banks. This would neither complicate the numerical resolution of the model nor change in a material way its qualitative properties. But, from a quantitative viewpoint, sunspot equilibria would not be consistent with Fact #3 that banking crises do not hit at random. For a discussion on the selection of the Pareto–dominant equilibrium in games with multiple Pareto–rankable Nash equilibria, see Cooper et al. (1990). Second, also note that rationing equilibria (with excess aggregate supply) are ruled out by bank competition. Indeed, in a rationing

Using relations (4) and (5), we can complete the description of the banking sector and derive the sector's return on deposit:

$$r_t \equiv \int_0^1 r_t(p)\, \mathrm{d}\mu(p) = \begin{cases} R_t \int_{\bar{p}_t}^1 p \frac{\mathrm{d}\mu(p)}{1-\mu(\bar{p}_t)} & \text{, if an equilibrium with trade exists} \\ \\ R_t \left(\frac{\gamma}{R_t} \mu\left(\frac{\gamma}{R_t}\right) + \int_{\frac{\gamma}{R_t}}^1 p\, \mathrm{d}\mu(p) \right) & \text{, otherwise.} \end{cases} \tag{6}$$

Relation (6) reflects the fact that in normal times the least efficient banks delegate financial intermediation to a mass $1 - \mu(\bar{p}_t)$ of more efficient banks, each of which lending to the firm a multiple $1 + \phi_t = (1 - \mu(\bar{p}_t))^{-1}$ of their initial assets at effective return pR_t. In the no–trade equilibrium, in contrast, banks do not delegate intermediation, and a mass $\mu(\gamma/R_t)$ of the banks use the storage technology.

3.3.3 Aggregate Supply of Corporate Loans

We are now in the position to describe the aggregate supply of corporate loans by the banking sector. The supply of corporate loans depends on whether or not the interbank market functions. When interbank trades are possible, all bank assets are channeled to the firm, so that the supply of corporate loans is a_t. When the interbank market is frozen, banks with $p < \gamma/R_t$ use the storage technology and the aggregate supply of corporate loans is $\left(1 - \mu\left(\frac{\gamma}{R_t}\right)\right) a_t$; denoting this supply by k_t, one gets:

$$k_t = \begin{cases} a_t & \text{, if an equilibrium with trade exists} \\ \\ \left(1 - \mu\left(\frac{\gamma}{R_t}\right)\right) a_t & \text{, otherwise} \end{cases} \tag{7}$$

Even though the use of the storage technology during crisis times is optimal from a private point of view, it is inefficient because the return on storage is less than the marginal return of the firm's project. It follows that, during crises, banks inefficiently divert assets away from the firm — a phenomenon we will refer to as a credit crunch. Proposition 2 follows.

Proposition 2 (Interbank Market Freeze and Credit Crunch) *An interbank market freeze is accompanied with a credit crunch.*

3.4 Decentralized General Equilibrium

A general equilibrium of the economy is defined as follows.

---

equilibrium, some inefficient banks would be rationed; and since these banks only have storage as relevant outside investment opportunity they would always be better off by deviating from the rationing equilibrium loan contract. Hence our model differs from the standard Stiglitz and Weiss (1981) adverse selection model with monopolistic banks.

Definition 1 (Decentralized general equilibrium) *A decentralized general equilibrium is a sequence of prices $\mathscr{P}_t \equiv \{R_{t+i}, r_{t+i}, \rho_{t+i}, w_{t+i}\}_{i=0}^{\infty}$ and a sequence of quantities $\mathscr{Q}_t \equiv \{c_{t+i}, y_{t+i}, k_{t+i}, h_{t+i}, a_{t+i}\}_{i=0}^{\infty}$ such that for a given sequence of prices, $\mathscr{P}_t$, the sequence of quantities, $\mathscr{Q}_t$, solves the optimization problems of the agents, and for a sequence of quantities, $\mathscr{Q}_t$, the sequence of prices, $\mathscr{P}_t$, clears the markets. The market clearing conditions in the interbank and credit markets are given by equations (5) and (7); and those of the labor and good markets by $h_t = 1$ and the usual resource constraint $y_t = c_t + i_t$.*

Since in equilibrium the household supplies one unit of labor, the level of production is $z_t f(k_t) \equiv z_t F(k_t, 1)$, the marginal productivity of capital is $z_t f_k(k_t) \equiv z_t \partial F(k_t, 1)/\partial k_t$, and the aggregate demand for capital (and credit) is $k_t = f_k^{-1}((R_t + \delta - 1)/z_t)$. It follows that the market clearing condition on the corporate loan market takes the form
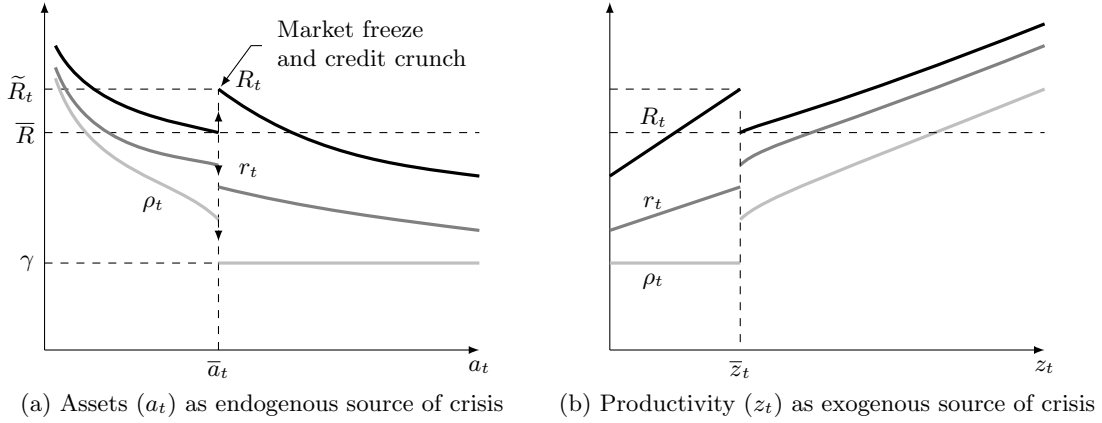
$$
f_k^{-1}\left(\frac{R_t + \delta - 1}{z_t}\right) = \begin{cases} a_t \text{ , if an equilibrium with trade exists} & (a) \\[2ex] a_t - \mu\left(\frac{\gamma}{R_t}\right) a_t \text{ , otherwise.} & (b) \end{cases} \tag{8}
$$

Relation (8) yields the equilibrium R_t as a function of the two state variables of the model, a_t and z_t. It also points to the two–way relationship that exists between the interbank loan market and the retail corporate loan market. We already showed that the way the interbank operates depends on whether or not $R_t \gtrless \overline{R}$. Likewise, whether or not the interbank market operates affects the supply of corporate loans and R_t. The corporate loan market equilibrium must therefore be solved sequentially. We first solve (8) for R_t assuming that an interbank market equilibrium exists (using 8a) and check that $R_t \geqslant \overline{R}$. In the negative, the interbank market equilibrium with trade cannot emerge (the interbank market freezes), and banks reduce their supply of loans to the firm. The corporate loan rate, R_t, then solves (8b). Note that in the no–trade equilibrium R_t may well be above $\overline{R}$; for example, it is the case for $\tilde{R}_t$ in Figure 3. But this rate is irrelevant for the existence of interbank trades because, in the absence of coordination failures, the only rate that matters in this respect is the one that would prevail in the equilibrium with trade if it existed. (Remember that the interbank market freezes if and only if an equilibrium with trade cannot be sustained in the first place.) It follows that condition $R_t < \overline{R}$ is neither necessary nor sufficient for the interbank market to freeze. Hence Proposition 3.

Proposition 3 (Interbank loan market freeze) *The interbank loan market freezes if and only if $a_t > \overline{a}_t \equiv f_k^{-1}((\overline{R} + \delta - 1)/z_t) \Leftrightarrow z_t < \overline{z}_t \equiv (\overline{R} + \delta - 1)/f_k(a_t)$.*

The threshold $\overline{a}_t$ is the maximum quantity of assets that the banking sector can allocate efficiently, that is, without the storage technology being used. Above this threshold counter–party fears are so prevalent that no trade takes place in the interbank market. In the rest of the paper we will refer to $\overline{a}_t$ as banks' absorption capacity. It increases with the level

Figure 3: Interest Rates

(a) Assets (a_t) as endogenous source of crisis

(b) Productivity (z_t) as exogenous source of crisis

of TFP ($\partial \overline{a}_t / \partial z_t > 0$) because following an increase in TFP the equilibrium corporate loan rate goes up and banks' opportunity cost of diversion rises, which alleviates the moral hazard problem. Symmetrically, there is a threshold $\overline{z}_t$ for the level of TFP below which the interbank market freezes; and this threshold increases with banks' assets ($\partial \overline{z}_t / \partial a_t > 0$). Proposition 3 thus suggests that market freezes result either from the *endogenous* over–accumulation of assets by the household (when, given z_t, a_t rises beyond $\overline{a}_t$), or from an *exogenous* adverse productivity shock that reduces banks' absorption capacity (when, given a_t, z_t falls below $\overline{z}_t$), as Figure 3 illustrates; or from a combination of both endogenous and exogenous factors. This variety of banking crises is a novel feature that will have important consequences for the macroeconomic dynamics. To see this, we report in Figure 4 a stylized representation of the household's optimal asset accumulation rules for three different levels of z_t (low, average, and high).

Figure 4: Optimal Decision Rules

17
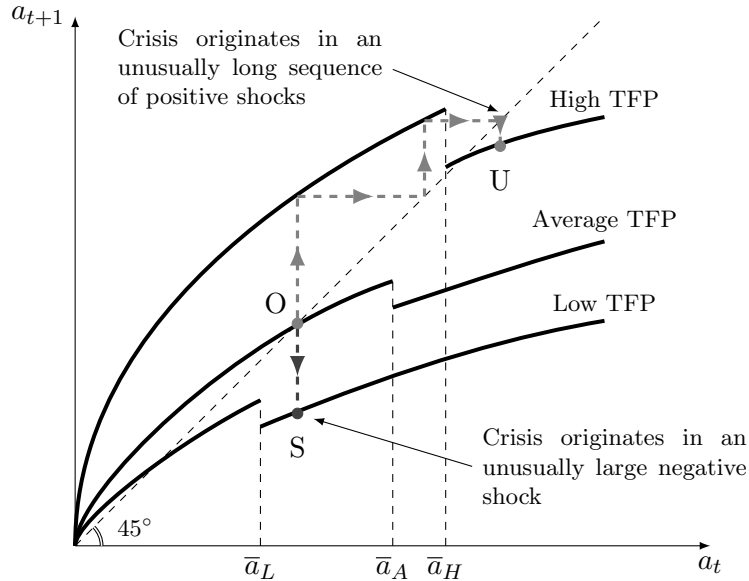
Given z_t, the optimal saving rule is discontinuous at $a_t = \bar{a}_t$. The upper branch (with $a_t \leqslant \bar{a}_t$) corresponds to normal times, where the interbank market functions well and all available assets in the economy are used to finance the firm. The lower branch (with $a_t > \bar{a}_t$) corresponds to crisis times. During a crisis, banks reduce their supply of loans to the firm, the firm reduces its demand for labor and the wage drops. The household dis-saves so as to sustain her consumption; hence the discontinuity in the decision rule when $a_t = \bar{a}_t$. To illustrate the variety of banking crises in the model, we focus on two polar examples of crises: S and U (it will become clear later that crises in the model are in effect "in–between" these two polar cases). The crisis in S arises exogenously. Starting from the average steady state O, the economy experiences an exogenous large negative technology shock, which brings the economy down to S. Because this shock instantaneously reduces banks' absorption capacity below the current level of assets, the crisis breaks out on impact. This crisis is purely driven by the bad realization of the shock and the amplification mechanism that the shock triggers. This is the channel usually depicted in the literature. The crisis in U is of a very different nature, as it arises endogenously. Starting from the average steady state O, the economy experiences an unusually long sequence of positive TFP shocks, which leads the household to accumulate assets above banks' absorption capacity (point U).[12] Even though crises may follow credit booms, not all credit booms lead to crises. A better indicator of crisis is the model probability that a crisis breaks out in the next period, as derived in Proposition 4.

Proposition 4 (Crisis probability) *Given the state of the economy (z_t, a_t) at the end of period t, the probability that a banking crisis breaks out in period $t+1$ is $\mathbb{P}(z_{t+1} < \bar{z}_{t+1} | z_t, a_t) = \Phi\left(\log \bar{z}_{t+1} - \rho_z \log z_t\right)$, where $\Phi(\cdot)$ denotes the normal c.d.f. As long as the optimal accumulation decision rule is increasing, $\mathbb{P}(z_{t+1} < \bar{z}_{t+1} | z_t, a_t)$ is an increasing function of $a_t / \bar{a}_t$.*

To the extent that this probability is fully consistent with general equilibrium effects and agents' rational expectations, it can be considered as the "true" crisis probability. In normal times, it increases as the economy approaches its absorption capacity (*i.e.*, as a_t gets closer to $\bar{a}_t$), as in this case the interest rates go down and the moral hazard problem worsens (see Figure 3. In Section 5.2 we use this probability to construct early warnings of banking crises, and compare the prediction performance of such warnings with warnings derived from credit aggregates.

---

[12]In theory, this second type of crisis does not even require any (positive) shock. To emphasize this point we present in Section E.1 of the technical appendix a purely deterministic version of the model —*i.e.* with a constant level of TFP— and show (i) that the model can generate endogenous booms and bust cycles and (ii) that the length of those deterministic cycles varies with the severity of financial frictions (for example, the larger θ, the higher the frequency of crises).

4 Quantitative Analysis

The aim of this section is to investigate the quantitative properties of the model. To bring our model closer to the workhorse macroeconomic model we extend it in two directions; we allow for endogenous labor supply decisions, and we introduce technological progress that is consistent with the observed long term productivity growth in advanced economies. These features are standard and do not change the qualitative properties of the model. The equations characterizing the general equilibrium of the economy are reported in Appendix B.

4.1 Calibration of the Model

Technology is represented by a constant–returns to scale Cobb–Douglas function, $z_t F(K_t, h_t; \Psi_t) \equiv z_t K_t^\alpha (\Psi_t h_t)^{1-\alpha}$ with $\alpha \in (0,1)$, where technological progress Ψ_t is exogenous and grows at the constant gross rate $\psi > 1$. In the sequel, lower(upper)–case letters will refer to variables that are (not) deflated for growth. The household is endowed with preferences over consumption, C_t, and hours worked, h_t, represented by a Greenwood et al. (1988) utility function,

$$\mathbb{E}_t \sum_{\tau=0}^{\infty} \tilde{\beta}^\tau \frac{1}{1-\sigma} \left(C_{t+\tau} - \vartheta \Psi_{t+\tau} \frac{h_{t+\tau}^{1+v}}{1+v} \right)^{1-\sigma},$$

where $v \geqslant 0$ is the inverse Frish labor supply elasticity. $\tilde{\beta}$ denotes the discount factor in the economy not deflated for growth. The presence of the technological progress term in the utility function ensures the existence of a balanced growth path (for another application see for example Jaimovich and Rebelo (2009). As is well known, in this class of utility functions wealth does not affect labor supply decisions. We only chose this specification for a practical reason as it yields a closed form for the banking sector's (detrended) absorption capacity that greatly simplifies the numerical solution of the equilibrium,

$$\bar{a}_t = \Gamma z_t^{\frac{1+v}{v(1-\alpha)}}, \text{ with } \Gamma \equiv \left(\frac{1-\alpha}{\vartheta} \right)^{\frac{1}{v}} \left(\frac{\alpha}{\overline{R}+\delta-1} \right)^{\frac{v+\alpha}{v(1-\alpha)}}. \tag{9}$$

The model is calibrated on a yearly basis to be consistent with the facts presented in Section 2. The calibration, reported in Table 3, is most standard. We set $\tilde{\beta}$ so that the household discounts the future at a 3% rate per annum in the economy deflated for growth. We assume that the labor supply elasticity is 2, such that $v = 0.5$. The labor dis–utility parameter ϑ is such that the household would supply one unit of labor in a deterministic version of the model. The risk aversion parameter σ is set to 4.5, which lies within the range of estimated values. The capital elasticity in the production function α is set to 0.3, and we assume that capital depreciates at a 10% rate per annum ($\delta = 0.1$). To calibrate the data generating process of TFP we proceed in three steps. First, we use the model to back out a time series for the logarithm of TFP that is consistent with Fernald (2012)'s annual series of output,

capital and labor. We then fit a linear trend to this time series and obtain a long term annual growth rate of TFP of 1.2% ($\psi = 1.012$). Finally, we use the detrended TFP series to estimate the AR(1) process in relation (3.1), and obtain $\rho_z = 0.9$ and $\sigma_z = 1.77\%$. Given the importance of parameters σ, ν, ρ_z and σ_z for the dynamics of our model, we will discuss the sensitivity of the model to changes in their values.

The remaining parameters pertain to the banking sector, and include the return on storage γ, the diversion technology parameter, θ, and the distribution of banks $\mu(\cdot)$. For tractability reasons we assume that $\mu(p) = p^\lambda$, with $\lambda \in \mathbb{R}^+$. The parameters of the banking sector are calibrated jointly so that *(i)* the spread between the real corporate loan rate and the implicit real risk free rate equals 1.7%, *(ii)* the real corporate loan rate equals 4.4%, and *(iii)* a financial recession occurs on average every 42 years. For statistics *(i)* and *(ii)* we use the real lending rate on mid–size business loans for the US between 1990 and 2011, as reported in the US Federal Reserve Bank's Survey of Terms of Business Lending, and the real Federal Fund rate. We obtain $\gamma = 0.94$, $\lambda = 25$, and $\theta = 0.093$. Based on this calibration, the model generates an average interbank loan rate of 0.90% and an implied threshold for the real corporate loan rate of 2.62% (*i.e.*, $\overline{R} = 1.0262$).

Table 3: Calibration

Parameter		Values
Discount factor (deflated for growth)	$\beta \equiv \tilde{\beta}\psi^{-\sigma}$	0.970
Inverse of Frish elasticity	υ	0.500
Labor disutility	ϑ	0.944
Risk aversion	σ	4.500
Capital elasticity	α	0.300
Capital depreciation rate	δ	0.100
Growth factor	ψ	1.012
Standard dev. productivity shock	σ_z	0.0177
Persistence of productivity shock	ρ_z	0.900
Bank distribution; $\mu(p) = p^\lambda$	λ	25
Diversion cost	θ	0.093
Storage technology	γ	0.9417

Note that we do not calibrate the model to replicate a priori any of Facts #2 (financial recessions are deeper and longer) and #3 (financial recessions follow credit booms) precisely because we want to discuss later the ability of the model to replicate these facts. We solve the model numerically using a collocation method, allowing for discontinuities in the asset accumulation decision rule at the points where the economy switches regime, *i.e.*, when $a_t = \overline{a}_t$.[13]

---

[13]Details on the numerical solution method are provided in Section B of the companion technical appendix.

4.2 The Economy in Normal Times

We analyze the response of the economy to a positive one standard deviation productivity shock. Figure 5 compares the dynamics in our model (thick plain line) with that in a frictionless model (dotted line), which is obtained by setting either $\gamma = 0$ (no moral hazard), $\theta = 0$ (no moral hazard), or $\lambda = +\infty$ (all banks have $p = 1$).[14] This frictionless version of our model corresponds to a standard Real Business Cycle (RBC) model. The responses

Figure 5: Impulse Response to a One Standard Deviation Technology Shock

—— Model (baseline), —●— Frictionless economy, ▇ 66% Confidence band around baseline.
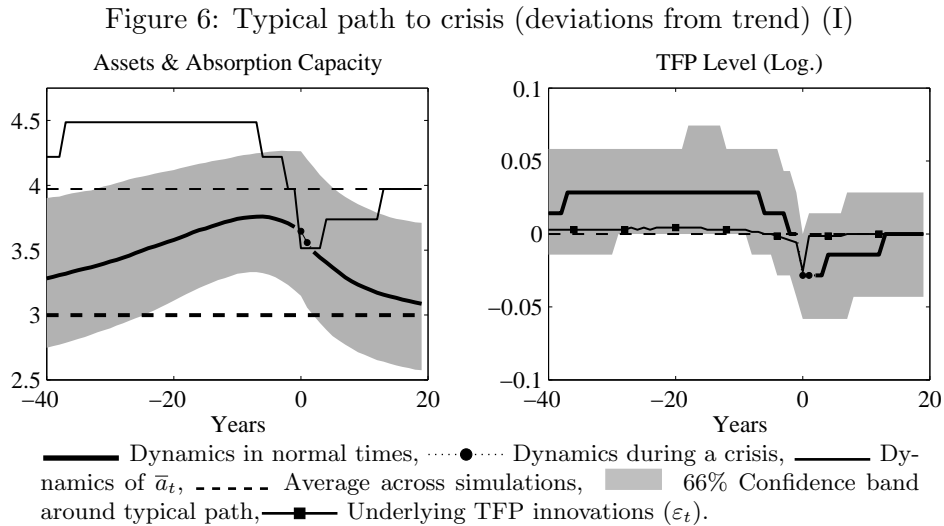Note: Impulse responses are computed using 100,000 Monte Carlo simulation of the model.
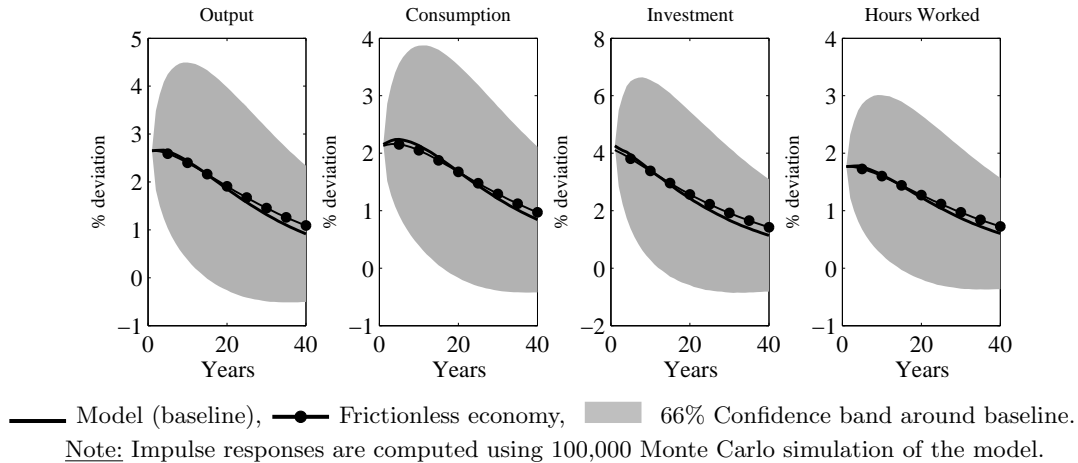
of output, hours worked, consumption, and investment are essentially the same in the two models. Our model only exhibits slightly more amplification than the RBC model due to a financial accelerator mechanism that tends to magnify the effects of the shock. In particular, following the shock the corporate loan rate rises, which relaxes banks' borrowing constraints in the interbank market. As the aggregate demand for interbank loans increases, the equilibrium interbank loan rate goes up and inefficient banks switch from the demand to the supply side of the market, so that the moral hazard problem recedes; hence the financial accelerator. However, this accelerator mechanism is hardly noticeable. In particular, the economy does not experience any banking crisis after the shock. Indeed, even though the corporate loan rate eventually falls below its steady state level as the household accumulates assets, at no point does it fall below $\overline{R}$ (not shown). Most of the time, the model behaves like a standard financial accelerator model; banking crises are indeed rare events that occur under specific conditions, as we show in the next section.

4.3 Typical Boom–Bust

This section describes the typical conditions under which banking crises, and hence financial recessions, occur in our model. As already discussed in Section 3.4, banking crises can be

---

[14]In all these cases savings are channeled to the most efficient bank $p = 1$, either as an equilibrium result or because $p = 1$ is the only type of bank in the economy.

driven either by exogenous productivity shocks, or by the household's asset accumulation behavior, or by a combination of these two factors. Which factor typically leads to a banking crisis is a quantitative question, which we settle by simulating the calibrated version of the model. Starting from the average steady state we simulate the model over 500,000 periods, and identify the years when a crisis breaks out as well as the path of the main aggregates —in deviation from their trend— 40 years before and 20 years after each crisis. We then compute, for each variable, the median of the distribution of paths, and refer to this median path as the "typical path". The left panel of Figure 6 describes the endogenous dynamics of banks' assets as well as banks' absorption capacity along this typical path, and the right panel describes the exogenous dynamics of TFP and of its innovations. By convention, period 0 corresponds to the period when the crisis breaks out.

Figure 6: Typical path to crisis (deviations from trend) (I)

Dynamics in normal times, ⋯●⋯ Dynamics during a crisis, —— Dynamics of $\bar{a}_t$, ----- Average across simulations, ▨ 66% Confidence band around typical path, —■— Underlying TFP innovations (ε_t).

The first important result that emerges from Figure 6 is that the typical banking crisis is *not* due to an unusually large, exogenous, negative shock. There is a negative TFP innovation in the year of the crisis (right panel, squared markers); but it is a 1.44–σ_z shock, which is neither unusual nor statistically different from zero at the 5% threshold. Following this shock, TFP and banks' absorption capacity fall to 2.8% and 11.5% below trend, respectively. But lower levels are reached 18.5% of the time in the simulations. This suggests that exogenous TFP innovations alone cannot explain why financial recessions are rare. The left panel shows that banks' assets are disproportionately large at the time the crisis hits; they are 25% above their average steady state and 68% above the steady state associated with the current TFP level. Had banks not been so large in the first place, the fall in their absorption capacity would not have triggered any crisis. This indicates that the banking sector is fragile when it is over–sized, as in this case even a standard adverse shock can trigger a crisis. Banks' assets rarely grow 68% oversize: this happens in only 6.7% of the time in our simulations. So, banking crises are rare in our model because it is unlikely that banks' absorption capacity
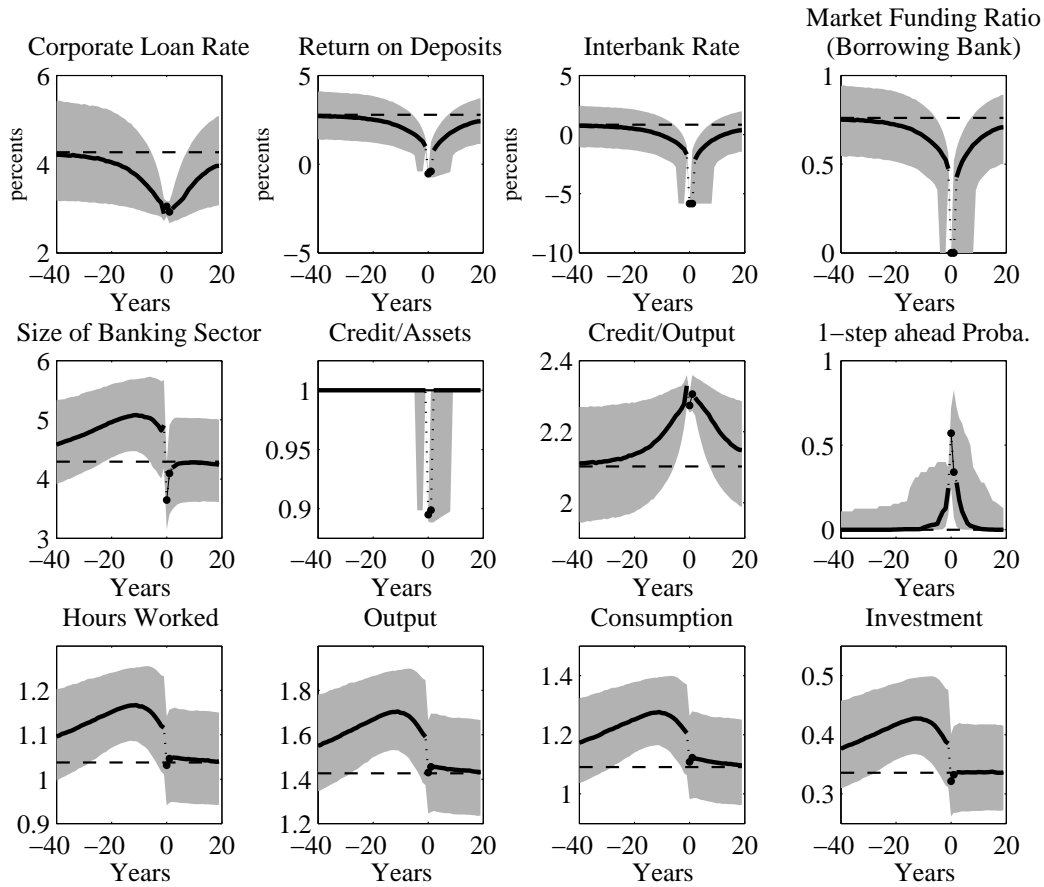
falls 11.5% below steady state at the same time as banks' assets are 68% above steady state. This prompts the question why the banking sector can grow so large. Since in our model banks' assets are entirely financed by the household, banks' expansion results from an excessive accumulation of savings or, as Bernanke (2005) coined it, from a saving glut.[15] In a neoclassical framework like ours, where savings are driven by the household's consumption smoothing behavior, a saving glut can only develop following an unusually and unexpectedly long string of positive TFP innovations, as the right panel of Figure 6 illustrates (squared markers). To understand how things work it is useful to decompose the pre–crisis period into two phases. During the first phase, TFP raises above trend and stays above it for a long time (right panel, plain line). These productivity gains result from the sequence of positive TFP innovations, which are small but yet large enough to overcome the mean reverting dynamics of TFP. These are good times, where the economy is far from its absorption capacity, the moral hazard problem is benign, crises are unlikely (see middle right panel of Figure 7), and it is optimal for the household to accumulate savings. This first phase lasts long, which gives the household enough time to build–up an unusually large stock of assets. This is the phase during which the saving glut develops and feeds the credit boom (see Figure 7).

The second pre–crisis phase starts when TFP reverts back to its trend.[16] As productivity gains peter out, the economy gets closer to its absorption capacity, output reverts to trend and the credit/output ratio rises. The equilibrium corporate and interbank loan rates both fall significantly below their average level; counter–party fears rise; the banking sector's absorption capacity diminishes; and in the year that precedes the crisis, the crisis probability goes up to 0.25. The household realizes this, and maintains her bank deposit at a relatively high level to hedge against the loss of revenues she would incur should the crisis materialize. But she does not internalize that, by doing so, she also brings the economy closer to its absorption capacity and makes the crisis even more likely; an externality that we dub "the saving glut externality". In this context of financial fragility, the $1.44-\sigma_z$ negative TFP shock in period 0 is enough to bring the banking sector down. As it breaks out, the crisis manifests itself by a sudden reversal of the economy, with the size of the banking sector falling by 25% and the credit to assets ratio falling by around 10%. As banks cut their supply of credit, the spread rises from 2pp to above 3.5pp (not shown). This rise in the spread is due to the return on bank deposit falling by more than the corporate loan rate. The latter is indeed relatively stable during the crisis, because the positive effect of the credit crunch essentially offsets the negative effect of the $1.44-\sigma_z$ negative TFP shock. Finally, Figure 7 shows that

---

[15]Another way to obtain a saving glut would be to consider a two–country model, where the banking sector would be more developed (*i.e.* would have a lower θ) in one of the countries, and would attract the developing countries' savings, as in for instance Caballero et al. (2008), Mendoza et al. (2009), and Boissay (2011).

[16]It should be clear that, given the positive 1.2% slope of the underlying productivity growth trend, productivity also increases during this second phase, but at a slower pace than its trend; we illustrate this in Section D.2 of the companion technical appendix by reporting the typical path to crisis with the underlying TFP trend.

Figure 7: Typical path to crisis (deviations from trend) (II)

Corporate Loan Rate

Return on Deposits

Interbank Rate

Market Funding Ratio (Borrowing Bank)

Size of Banking Sector

Credit/Assets

Credit/Output

1−step ahead Proba.

Hours Worked

Output

Consumption

Investment

—— Dynamics in normal times, ·····●····· Dynamics during a crisis, —— Dynamics of $\overline{a}_t$, − − − − Average across simulations, ▢ 66% Confidence band around typical path.

<u>Note:</u> The size of the banking sector corresponds to the sum (as opposed to consolidation) of all banks' assets and therefore includes interbank funding. Formally, we report $a_t + (1 - \mu(\overline{p}_t)) \phi_t a_t$.

the banking crisis ends a long period of economic expansion where output, investment, hours worked, and consumption were more than 10% above their respective trends. As the crisis hits, all aggregates return instantaneously to trend, inducing a large recession. The financial recession is also accompanied with a sudden and large drop in consumption, which we view as another sign of the incapacity of the household to hedge effectively against the crisis.

The confidence band around the typical path contains additional information about its properties.[17] In particular, Figure 6 indicates that TFP is not necessarily below trend in the year that precedes the crisis. In that year, the distribution of TFP is in fact not different from its unconditional (symmetric) distribution. In contrast, the credit/output ratio is always above its mean in that year: it is more than 5% (10%) above its average steady state in the year that precedes the crisis in 95% (60%) of the cases. This suggests that credit is more informative than TFP about the likelihood of a crisis, and confirms the critical role of credit as an endogenous source of financial instability. Therefore, the typical path is an overall fair representation of banking crises in the model.

4.4 Inspecting the Mechanisms

There is nothing systematic about the prevalence of credit boom driven crises. This is a result, which can be partly explained by the asymmetric effects of the household's consumption smoothing behavior on financial stability. In bad times TFP is relatively low and the household dis–saves, which by lowering the TFP threshold $\overline{z}_t$ makes crises less likely (remember that $\partial \overline{z}_t / \partial a_t > 0$). Hence, in bad times the dynamics of asset accumulation tends to stabilize the banking sector. In contrast, in good times TFP is relatively high and the household accumulates assets, which by raising $\overline{z}_t$ makes crises more likely. In this case the dynamics of savings tends to destabilize the financial sector. As already discussed, though, asset accumulation by the household would not be as destabilizing if the household internalized its adverse effects on financial stability. Therefore we would argue that another important reason why most crises follow credit booms in the model is the presence of the saving glut externality. To shed more light on those mechanisms, we compare in Figure 8 the typical paths of assets and TFP in our baseline economy with the typical paths in an economy without the saving glut externality and in an economy with a constant saving rate.[18] The idea here is that if the externality and consumption smoothing are both necessary to generate crises endogenously, then we would expect the typical crisis in those two economies to be generated exogenously,

---

[17] The interested reader is referred to Section D.5 of the companion technical appendix for details about the following discussion on distributions of TFP and the credit/output ratio before a crisis.

[18] The economy without the saving glut externality corresponds to an hypothetical economy in which the household internalizes —all else equal— the adverse effects of her savings decisions. In the case of the constant saving rate economy, the saving rate is set at the average steady state saving rate of the baseline economy. We describe the constant saving rate version of the model in Section C of the companion technical appendix, and explain how we solve the model in the absence of the saving glut externality in Section B.2 of the companion technical appendix.

that is: (i) to be associated with an unusually low level of TFP at the time it breaks out and (ii) not to be preceded by an unusually high level of assets —or credit.

Figure 8: Typical paths to crisis without smoothing or externality (deviations from trend)

——— Baseline; ——◆—— No saving glut externality; —○— Constant saving rate.
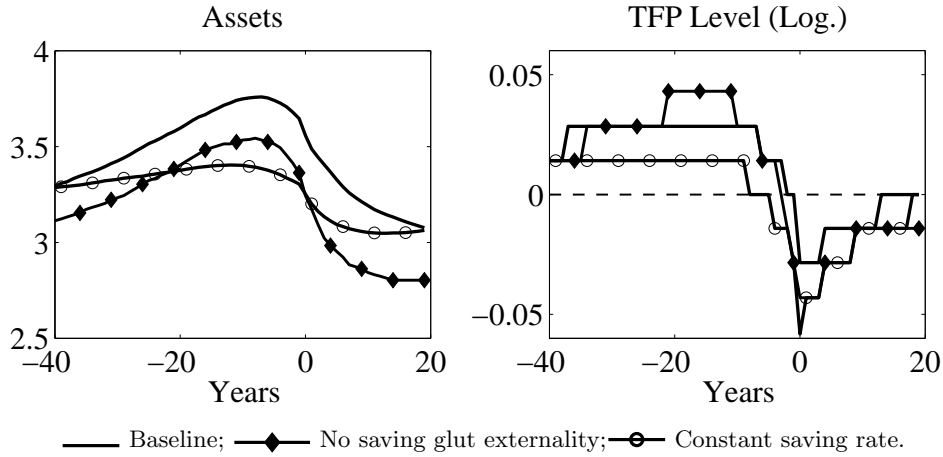
The first result that emerges from Figure 8 is that in the constant saving rate version of the model (circle markers) the typical banking crisis is *not* due to an endogenous boom in credit. As the crisis breaks out bank assets are hardly above their average steady state whereas TFP is significantly —6%— below trend, which happens only 4.9% of the time in the simulations; TFP shocks can thus almost entirely explain on their own why financial recessions are rare. Hence we would argue that consumption smoothing is an important ingredient to generate endogenous banking crises. But again, it is not sufficient. The typical path of TFP in the economy without the saving glut externality (right panel, diamond markers) indeed shows that in this economy too the crisis occurs when the level of TFP is relatively low. The reason is that the household, who now internalizes the adverse effects of her savings, not only saves overall less but also hastily dis–saves as soon as the likelihood of a crisis increases— precisely to avoid that the crisis materializes. To trigger the crisis the negative TFP shocks must therefore surprise the household, and be relatively large; here TFP must fall to 4.3% below trend. We conclude that both the consumption smoothing behavior and the saving glut externality are necessary to generate banking crises endogenously in the calibrated version of the model. Consumption smoothing is important because it causes the development of the saving glut in the first pre–crisis phase and creates the conditions under which even small shocks matter; the saving glut externalities are equally important as it is what prevents the saving glut from resorbing when the economy is on the brink of a crisis.

5 Quantitative Implications

5.1 Financial Recession Statistics

This section assesses the ability of our model to account for the stylized facts on financial recessions, as documented in Section 2.[19] We simulate a 500,000 period time series for output and identify all business cycle peaks and troughs so that the model generates the same percentage of recessions as in the data. A recession is then said to be a financial recession when a banking crisis breaks out between the peak that precedes it and the trough that ends it; we find that all banking crises in the simulations are associated with a recession.

Table 4: Statistics on recessions in the model

|	Financial	Other	All	Severe	Mild
N. events	11,739	44,711	56,450	18,800	18,800
Frequency (%)	2.35	8.94	11.29	3.76	3.76
Duration (years)	2.08	1.39	1.53	2.22	1.04
Magnitude ($\Delta_{p,t}y$, %)	-12.60	-4.98	-6.56	-11.32	-3.28
Credit crunch					
$\Delta_{p,t}k^{HP}$	-9.44	0.29	-1.76	-5.59	0.27
$\Delta_{p,p+2}k^{HP}$	-5.09	0.09	-0.99	-2.97	0.02
Credit boom					
$\Delta_{p-2,p}k^{HP}$	3.70	0.20	0.93	2.33	0.14
k_p^{HP}	3.81	0.11	0.88	2.33	0.06

Note:$\Delta_{p,t}x$ (resp. $\Delta_{p,p+2}x$, $\Delta_{p-2,p}x$) denotes the percentage change of variable x from peak to trough (resp. from peak to peak+2 years, and peak-2 years to peak), where x denotes either output (y) or the Hodrick Prescott ($\lambda = 6.25$) cyclical component of credit (k^{HP}). Statistics based on a 500,000 time period simulation.

For comparison purposes, Table 4 reports the same statistics as Table 1. By construction, the model generates financial recessions with a frequency of 2.35% (Fact #1). These recessions feature a 12.6% drop in output from peak to trough and last on average 2.08 years. This is more than in other recessions, which are about 8 months shorter and where output falls by about half as much (4.98%). These differences are of the same order of magnitude as in the actual data. The model replicates Fact #2 that financial recessions are deeper and last longer than other recessions. However, the magnitude of the average financial recession is larger than in the data. In the Schularik and Taylor (2012) data set, output falls on average by 6.84% from peak to trough and only 10% of the financial recessions are as large as in our model simulations. Three features of the model contribute to these relatively large output losses. First, financial recessions are accompanied by a severe credit crunch, with credit falling by 9.44% from peak to trough (this is more than twofold what is observed in the data). Second,

---

[19]In Section 4.2 we showed that financial frictions in our model hardly matter in normal times, which implies that our model has essentially the same business cycle properties (second order moments) as the textbook RBC model; see Section D.7 of companion technical appendix.

the firm is assumed to refinance its capital through bank loans only, and to roll over those loans every period; it does not retain earnings, and does not issue bonds or outside equity. As a consequence, capital is not pre–determined (only the household's stock of assets is) and the fall in credit implies an equally large fall in capital. Finally, the absence of wealth effects, and therefore of leisure smoothing behavior, makes hours more responsive to fluctuations in wages. As capital falls, labor productivity and labor demand tank, wages go down, and the household reduces her supply of labor, which amplifies the initial fall in output. Hence the severe recession. In Section E.5 of the companion technical appendix, we show that a version of the model with a labor Frish elasticity of 0.5 (instead of 2) would perform better regarding the dynamics of output, but not as well regarding the dynamics of credit.

Figure 9: Dynamics of output and credit gaps around recessions
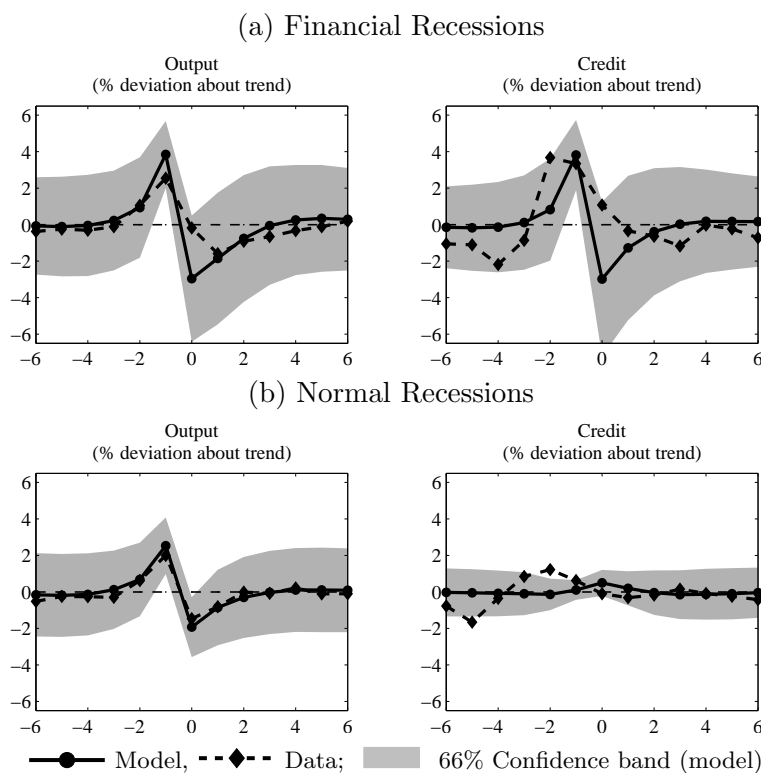
(a) Financial Recessions

(b) Normal Recessions

Note: Average dynamics of the Hodrick-Prescott ($\lambda = 6.25$) cyclical component of (log) output and credit 6 periods before and after the start of a business cycle (period 0) with and without a banking crisis. For the sake of consistency, we treat the simulated series of output and credit as we treat the actual data in Figure 1.

The last two rows of Table 4 show statistics on the evolution of credit before the start of financial recessions. Credit grows by 3.7% in the two years that precede financial recessions (against 4.56% in the data), whereas it does not grow at all before other types of recessions (as in the data). Moreover, at the time the financial recession starts, credit is 3.81% above its Hodrick–Prescott trend, whereas it is basically on its trend in the case of other recessions. As a way to illustrate these results, we report in Figure 9 the dynamics of output and credit gaps around normal and financial recessions, both in the model and in the data. The figure

indicates not only that financial recessions in the model are associated with a boom–bust cycle in credit —whereas normal recessions are not, but also that the dynamics observed in the data lie within the 66% confidence band of the model. The model is therefore consistent with Fact #3 that (only) financial recessions follow credit booms.

5.2 Crisis Prediction

The probability of falling into a crisis (see Proposition 4) can be used to construct early warnings in the context of our model; but it is not observable in the real world. Our aim in this section is to look into whether this model probability can be backed out from variables that are directly observable in the data, to construct early warnings based on this crisis probability, and to discuss the properties of those early warnings (Type–I and Type–II errors).

In the model three variables are key to understand the dynamics to crises: the level of bank deposit (or credit), TFP, and the credit/output ratio. (Instead of the credit/output ratio, one could consider the corporate loan rate, since given our production function the latter is a simple transformation of the former.) We simulate the model over 500,000 time periods, and then regress the crisis probability on these variables. The results are reported in Table 5. We

Table 5: Prediction of banking crises

	Model probability (benchmark)	Probability regressions				Logit
		(1) z	(2) a	(3) (a, z)	(4) K/Y	(5) K/Y
R^2	–	0.03	0.55	0.69	0.72	0.38
F-Test	–	0.00	0.00	0.00	0.00	0.00
Type–I (%)	31.43	100.00	72.50	56.87	36.79	35.97
Type–II (%)	4.85	0.00	4.11	5.55	4.92	5.16
N. warnings	30,215	0	22,020	30,439	29,911	31,089

Note: Regressions (1)-(4): linear regression using a logistic transformation of the period t model probability that a crisis breaks out in $t + 1$ as the dependent variable. Regression (5): logit regression using a dummy equal to one when a crisis breaks out in $t + 1$ (and zero otherwise) as dependent variable. All models include period t independent variables (in logarithm) and are conditional upon not being in a crisis in period t. The regression sample includes 468,769 observations and 11,739 crises. Type–I: probability that the model fails to issue a warning. Type–II: probability that the model mistakenly issues a warning. A warning is issued as soon as the —model or fitted— crisis probability is above 12.75%, which threshold is chosen so that the type–II error is less than 5% in the case of model based warnings (first column).

start by including the levels of credit (a) and TFP (z) separately and find —as expected— that the fit is relatively poor. TFP, for example, is significant but explains only 3% of the variations in the crisis probability (regression (1)). The R^2 is higher in the regression with credit (regression (2)), which reflects the endogenous feature of the crises in the model. But it is still relatively low compared to the regressions with both credit and TFP. Indeed, with those credit supply and demand factors together the R^2 goes up to 69% (regression (3)); this result

is robust to using the credit/output ratio (regression (4)). Next, we construct early warnings of crises based on the crisis probabilities. By convention, we issue a warning whenever the —model or fitted— probability of a crisis is above 12.75%. This cut–off value is chosen so that the warnings derived from the model probability (our benchmark) are statistically correct at the 5% threshold (*i.e.* so that the Type–II error is less than 5%). In this case, the model issues almost three times as many warnings as there are crises in the simulations (30,215 warnings for 11,739 crises; see Tables 4 and 5), but also signals two thirds of the crises (the percentage of Type–I errors is 31.43%). Among the regression–based warnings, only those derived from the regression with the credit/output ratio (regression (4)) perform as well, with Type–I and Type–II errors of 36.79% and 4.92%, respectively. One caveat of regression (4), though, is that the dependent variable —the model probability of a crisis— is not observable and has no clear counter–part in the data. So, as a last experiment, we check whether the credit/output ratio remains a good predictor when we instead use a crisis dummy as dependent variable and estimate a Logit regression (regression (5)). We find that the warnings derived from this regression have the same properties as those derived from the model probability. Altogether, these findings therefore suggest that it is reasonable to look at the credit/output ratio to predict crises.

5.3 Sensitivity Analysis

We begin our discussion by evaluating the sensitivity of our results to the key structural parameters of the model. In each case, we simulate the model over 500,000 periods, identify financial recessions, and compare their frequency, average duration, and average magnitude against the baseline in Table 4. The results are reported in Table 6.

We first make the household more willing to smooth consumption over time by raising the utility curvature parameter σ from 4.5 (baseline) to 10. Following positive TFP shocks, the saving glut grows faster and larger than in the baseline model, leaving banks more exposed to adverse shocks. As a result, the probability of a financial recession is higher than in the baseline (4.74% *versus* 2.35%). The average financial recession is however somewhat shorter (1.75 periods) and milder (-10.61%). We next increase labor supply elasticity from 2 (baseline) to 4 ($\nu = 0.25$), thus making the household's labor supply and income more pro–cyclical. As a consequence the household accumulates assets faster and the frequency of financial recessions rises to 3.45%. The impact of the credit crunch on output is also magnified by the larger response of hours worked; ultimately, recessions are thus more severe and last longer.

These first two experiments essentially affect the usual consumption smoothing mechanism. We now assess the sensitivity of our results to the specification of the banking sector. The

Table 6: Sensitivity Analysis: Financial Recessions

| | Baseline | σ | ν | θ | λ | σ_z | ρ_z | Altern. |
		10	0.25	0.15	20	0.025	0.70	TFP
Frequency (%)	2.35	4.74	3.45	5.87	5.73	4.56	4.34	2.32
Duration (years)	2.08	1.75	2.31	1.72	1.84	2.09	2.22	1.99
Magnitude (%)	-12.60	-10.61	-16.33	-9.29	-12.05	-15.40	-17.82	-10.86

Note: The alternative TFP model (last column) is a model where the TFP process is re–estimated correcting for the US productivity slowdown of the 70s. In this case, we re–calibrate ψ to remove long term growth ($\psi = 1$), as well as γ, to have the same crisis probability as in the baseline (2.3%).

fourth column of Table 6 shows the effect of an increase in productivity of banks' diversion technology, as captured by a rise in θ from 0.1 (baseline) to 0.15. This aggravates the moral hazard problem between banks, reduces the absorption capacity of the economy, and increases the probability of a financial recession, which jumps from 2.35% to 5.87%; the recession is milder and shorter than in the baseline. A change in the distribution of banks has qualitatively similar effects. In the fifth column we consider an economy where $\lambda = 20$, instead of $\lambda = 25$ in the baseline, meaning that there are more inefficient banks in the economy. Since those banks have more incentives to divert cash, the moral hazard problem is more stringent and counter–party fears increase. Lenders impose a lower market funding ratio, aggregate demand and the interbank rate drop, the average quality of borrowers deteriorates. As a result, the frequency of financial recessions increases to 5.73%.

Next, we investigate the role of uncertainty. We start by raising σ_z from 0.0177 to 0.025, leaving ρ_z unchanged. The consequences are straightforward: the household accumulates more assets for consumption smoothing motives, the corporate loan rate decreases with respect to baseline, and the financial sector is more fragile. We then decrease the persistence of the shock down to $\rho_z = 0.7$ (instead of $\rho_z = 0.9$ in the baseline), leaving σ_z unchanged. This change exerts two opposite effects on financial stability. On the one hand, savings tend to be more sensitive to shocks and the household accumulates assets faster than in the baseline. On the other hand, periods of high total factor productivity levels are shorter, which implies that the household does not have as much time to accumulate assets. In our calibration, the first effect dominates, and financial recessions are more frequent, last longer and are more protracted. Finally, we re–estimate the TFP shock process while detrending the logarithm of TFP with a linear trend break to fit the US productivity slowdown of the seventies. We find that the technology shock is less persistent ($\rho_z = 0.7$) and less volatile ($\sigma_z = 0.015$). Since our model is not designed to accommodate a break in the TFP growth trend we omit growth and set $\psi = 1$. For comparison purposes, we also adjust the return of storage (γ) so that financial recessions remain as frequent as in our benchmark economy. We find that financial recessions in this model are as large and as long as in the baseline.

6 Welfare Cost of Financial Frictions and Policy Intervention

The aim of this section is to evaluate and decompose the welfare cost of financial frictions —and thereby of the existence of financial recessions. Following Lucas (1987), we measure these welfare costs as the permanent percentage increase in consumption that would be necessary to make the household indifferent between the decentralized equilibrium allocation and the First Best allocation, which would prevail in a frictionless world (here, the RBC model). The results are reported in Table 7. We find that, all in all, financial frictions reduce consumption permanently by 2.20% (first column). These welfare losses can be decomposed

Table 7: Welfare Costs (% of permanent consumption)

Financial frictions FBA - DEA	Deficient institutions FBA - CEA	Externalities CEA - DEA	Financial under-development DEA - NIM
2.20	1.53	0.61	4.61

<u>Note:</u> The welfare associated to an allocation j is measured by $\mathcal{W}^j(\mathbf{c}^j, \mathbf{h}^j) = \mathbb{E} \sum_{t=0}^{\infty} \beta^t U(c_t^j, h_t^j)$ where $\mathbf{c}^j \equiv \{c_t^j\}_{t=0}^{\infty}$ and $\mathbf{h}^j \equiv \{h_t^j\}_{t=0}^{\infty}$. The table reports the permanent percentage difference consumption λ between two allocations j and i that solves $\mathcal{W}^i((1+\lambda)\mathbf{c}^i, \mathbf{h}^i) = \mathcal{W}^j(\mathbf{c}^j, \mathbf{h}^j)$; allocation j is preferred to allocation i if $\lambda > 0$. The welfare cost is computed based on 10,000 Monte–Carlo simulations of the model. FBA: First Best Allocation; DEA: Decentralized Equilibrium Allocation; CEA: Constrained Efficient Allocation; NIM: Equilibrium allocation in a model without interbank market.

into two parts. Some of these losses are due to the presence of externalities; the saving glut externality, and the fact that the household does not internalize that the financial intermediation costs (χ_t) are rebated. We measure the cumulated welfare cost of these externalities as the difference in welfare between the constrained efficient allocation and the decentralized equilibrium allocation. The constrained efficient allocation is defined as the allocation that a central planner would implement if she faced the same constraints as banks. We find that the externalities cost the household 0.61% of permanent consumption, which is large compared to what is usually found in the literature; Bianchi and Mendoza (2010) and Bianchi (2011), for example, report values ranging from 0.05% to 0.2%. However, these welfare losses are still of second order relative to those due to financial markets being deficient. Indeed, we also compute the welfare cost of the market and institutional deficiencies that are responsible for limited contract enforceability (moral hazard) and the lack of transparency (asymmetric information). We measure this cost as the difference of welfare between the First Best allocation and the constrained efficient allocation. We find that deficient institutions cost the household 1.53% of permanent consumption, *i.e.*, more than twice as much as the externalities. Finally, for the sake of completeness, we calculate the welfare loss if interbank loans were not be permitted in the first place. To do so we compare our baseline economy with an hypothetical, financially under–developed economy with no interbank market. As Mendoza et al. (2007), we find that financial under-development has a sizeable, detrimental effect on welfare; it reduces permanent consumption by 4.61%.

The relatively large welfare cost of financial frictions calls for policy intervention. One set of policies are macro-prudential policies that make private agents internalize the effects of their decisions on financial stability. In the model, these policies should typically aim at curbing the supply of corporate loans, to keep the corporate loan rate above $\overline{R}$. There are many ways to achieve this. For example, the regulator could tax banks to reduce their returns on deposit and prevent the development of the saving glut. The profile and timing of the tax is also a critical part of the policy. For instance, a tax rate that is constant throughout the business cycle may make the banking sector more stable, but it may also induce less savings, less capital accumulation, less wealth, and overall reduce welfare. Indeed, such a distorting tax may be very inefficient because, crises being rare, it would be levied mostly in normal times, when the banking sector functions well on its own. We would argue that banks should only be taxed when the probability of a crisis increases (see Figure 7). Taxes should not be raised too early, to let the economy reap the most benefits from high productivity; but they should not be raised too late, to give the household enough time to dis–save and keep assets below banks' absorption capacity. That the welfare cost of externalities is low relative to that of institutions being deficient (see Table 7) however suggests that even an optimal macro–prudential policy would have rather limited effects on welfare. Structural reforms aiming at improving financial contract enforceability and transparency in the banking sector may potentially have a much bigger impact.

7   Concluding Remarks

We offered a simple quantitative macro–economic model where endogenous financial developments could turn the real business cycle round. The model features a non–trivial banking sector, in which banks are heterogeneous with respect to their intermediation skills. This heterogeneity gives rise to an interbank market. Moral hazard and asymmetric information in this market lead to infrequent market meltdowns, banking crises, credit crunches and, ultimately, severe recessions. Our model has the potential to generate a variety of banking crises, and we view it as a step toward understanding the interactions between financial and real business cycles. In this respect, the quantitative properties of the model are encouraging.

For tractability reasons we made some strong assumptions and presented a stylized model, whose narrative of banking crises may not be taken at face value. Indeed, we left out some aspects of the banking sector that would deserve more attention in future extensions. For instance, we assumed that the household cannot finance the firm directly and —by the same token— that the firm cannot issue bonds or equity. A proper investigation of the interactions between financial and real business cycles must take into account the substitution effects that exist between direct and intermediated finance. To keep matters simple we modeled a closed economy. Accordingly, the saving glut, which is key to the dynamics in our model, can only

be fed by the domestic household. Opening the economy could have interesting insights as the saving glut would become global, thus offering the possibility to study how banking crises spread internationally. Finally, we ignored that frictions may exist between the banks and the household. Therefore, we could not pin down banks' optimal funding structure, which in turn prevented us from studying the dynamics of bank leverage. It would be interesting to model bank leverage explicitly so as to get richer dynamics, and be in a position to discuss the effects of macro–prudential policies like changes in bank capital requirements. This and other extensions are left to future research.

References

Bernanke, B., 2005, "The Global Saving Glut and the U.S. Current Account Deficit," Remarks by Governor Ben S. Bernanke at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia.

Bernanke, B., M. Gertler and S. Gilchrist, 1999, "The Financial Accelerator in a Quantitative Business Cycle Framework," Handbook of Macroeconomics, volume 1, Chapter 11, J..B. Taylor and M. Woodford eds, Elsevier.

Bianchi, J., 2011, "Overborrowing and Systemic Externalities in the Business Cycle," American Economic Review, 101(7), 3400–3426.

Bianchi, J. and E. G. Mendoza, 2010, "Overborrowing, Financial Crises and 'Macroprudential' Policy," NBER Working Paper No. 16091.

Boissay, F. 2011, "Financial Imbalances and Financial Fragility", ECB Working Paper No. 1317.

Boissay, F., C. Collard, and F. Smets, 2013, "Booms and Systemic Banking Crises", ECB Working Paper No. 1514.

Borio, C. and M. Drehmann, 2009, "Assessing the risk of banking crises: Revisited", BIS Quarterly Review, March, 29–46.

Brunnermeier, M. and Y. Sannikov, 2012, "A Macroeconomic Model with a Financial Sector," mimeo, Princeton University.

Burkart, M. and T. Ellingsen, 2004, "In-Kind Finance: A Theory of Trade Credit,", American Economic Review, 94(3): 569–590.

Caballero, R., Farhi, E., and P-O. Gourinchas 2008, "An Equilibrium Model of "Global Imbalances" and Low Interest Rates," American Economic Review, Vol. 98, 358-393.

Christiano, L., R. Motto, and M. Rostagno, 2013, "Risk Shocks," American Economic Review, forthcoming.

Claessens, S., A. Kose and M. Terrones, 2008, "How Do Business and Financial Cycles Interact?," IMF working Paper No. WP/11/88.

Claessens, S., A. Kose and M. Terrones, 2011, "What Happens During Recessions, Crunches and Busts?," IMF working Paper No. WP/08/274.

Cooper, R., D. DeJong, R. Forsythe, and T. Ross, 1990, "Selection Criteria in Coordination Games: Some Experimental Results," American Economic Review, vol. 80, No. 1, March, 218–233.

Fernald, J., 2012, "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," mimeo, Federal Reserve Bank of San Francisco.

Gertler, M. and P. Karadi, 2011, "A Model of Unconventional Monetary Policy", Journal of Monetary Economics, forthcoming.

Gertler, M. and N. Kiyotaki, 2009, "Financial Intermediation and Credit Policy in Business Cycle Analysis," Mimeo, Princeton University.

Gertler, M. and N. Kiyotaki, 2013, "Banking, Liquidity and Bank Runs in an Infinite Horizon

Economy," Mimeo, Princeton University.

Gorton, G., 1988,"Banking Panics and Business Cycles," Oxford Economic Papers, New Series, vol. 40, No. 4, 751–781.

Gorton, G., 2010, Interview with "The Region," December, downloadable from `http://www.minneapolisfed.org/pubs/region/10-12/gorton.pdf`.

Gorton, G., 2012, "Some Reflections on the Recent Financial Crisis," NBER Working Paper No 18397.

Gourio, F., 2012, "Disaster Risk and Business Cycles," American Economic Review, 102(6), 2734–2766.

Greenwood, J., Z. Hercowitz and G. Huffman, 1988,"Investment, Capacity Utilization, and the Real Business Cycle," American Economic Review, vol 78, No. 3, 402–417.

Hahm, J–H., H–S. Shin and K. Shin, 2011, "Non-Core Bank Liabilities and Financial Vulnerability," mimeo, Princeton University.

Hart, O., 1995, "Firms, contracts, and Financial Structure," Oxford University Press.

He, Z. and A. Krishnamurthy, 2012, "A Macroeconomic Framework for Quantifying Systemic Risk," mimeo, University of Chicago.

Holmström, B. and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and the Real Sector," Quarterly Journal of Economics, vol. CXII, No 3, August, 663–691.

Jaimovich, N. and S. Rebelo, 2009,"Can News about the Future Drive the Business Cycle?", American Economic Review, Vol. 99, 1097-1118.

Jermann, U. and V. Quadrini, 2012, "Macroeconomic Effects of Financial Shocks", American Economic Review, Vol. 102, 238-71.

Jordà, O., M. Schularick and A. Taylor, 2011a, "Financial Crises, Credit Booms and External Imbalances: 140 Years of Lessons," IMF Economic Review, Vol. 59, 340–378.

Jordà, O., M. Schularick and A. Taylor, 2011b, "When Credit Bites Back: Leverage, Business Cycles, and Crises," NBER Working Paper No.17621.

Kaminsky, G. and C. Reinhart, 1999: "The Twin Crises: The Causes of Banking and Balance of Payments Problems," American Economic Review, Vol. 89, 473-500.

Laeven, L. and F. Valencia, 2008, "Systemic Banking Crises: A New Database," IMF Working Paper No 08224.

Lucas, R., 1987, "Models of Business Cycles," Basil Blackwell eds, New York.

Maddaloni, A. and J–L. Peydro, 2011, "Bank Risk-Taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro Area and US Lending Standards," Review of Financial Studies, Vol. 24, 2121–65.

Mankiw, G. 1986, "The Allocation of Credit and Financial Collapse," Quarterly Journal of Economics, Vol. 101, 455-70.

Mas–Colell, A., M. Whinston, and J. Green, 1995, "Microeconomic Theory," Oxford University Press.

Mendoza, E., 2010, "Sudden Stops, Financial Crises and Leverage," American Economic Review, Vol. 100(5), 1941–66.

Mendoza E., Quadrini, V. and J-V. Rìos-Rull (2007): "On the Welfare Implications of Financial Globalization Without Financial Development," in G. Corsetti and M. Yorukoglu (eds.), NBER International Seminar on Macroeconomics, Chicago: The University of Chicago Press.

Mendoza E., Quadrini, V. and J-V. Rìos-Rull (2009): "Financial Integration, Financial Development and Global Imbalances," Journal of Political Economy, vol. 117(3), pp. 371-416.

Mendoza, E. and K., Smith, 2005, "Quantitative Implications of a Debt-Deflation Theory of Sudden Stops and Asset Prices," Journal of International Economics, Vol. 70(1), 82–114.

Mendoza, E. and M. Terrones, 2012, "An Anatomy of Credit Booms and their Demise," NBER Working Paper No. 18379.

Minsky, H., 1977, "A Theory of Systemic Fragility," in E. Altman and A. Sametz, eds., Financial Crises.

Ravn, M. and H. Uhlig, 2002, "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," Review of Economics and Statistics, May, Vol. 84, No. 2, 371–376.

Reinhart, C. and V. Reinhart, 2009: "Capital Flow Bonanzas: An Encompassing View of the Past and Present," in J. Frankel and F. Giavazzi (eds.), NBER International Seminar in Macroeconomics, Chicago: Chicago University Press.

Reinhart, C. and K. Rogoff, 2008: "Banking Crises: an Equal Opportunity Menace," NBER Working Paper No. 14587

Reinhart, C. and K. Rogoff, 2009, "This Time is Different — Eight Centuries of Financial Folly," Princeton University Press.

Schularick, M. and A. Taylor, 2012, "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," American Economic Review, Vol. 102, 1029–1061.

Shin, H–S., 2008, "Securitisation and Financial Stability," mimeo, Princeton University.

Shin, H–S., 2010, "Risk and Liquidity," Clarendon Lecture in Finance, Oxford University Press.

Stiglitz, J. and A. Weiss, 1981, "Credit Rationing in Markets with Imperfect Information," American Economic Review, vol. 71, No. 3 (June), 393–410.

Tirole, J., 2006, "The Theory of Corporate Finance," Princeton University Press.

Uhlig, H., 2010, "A Model of Systemic Bank Run," Journal of Monetary Economics, vol. 57, 78–96.

Varian, H., 1979, "Catastrophe Theory and the Business Cycle," Economic Inquiry, 17(1), 14–28.

— APPENDIX —

A   Proofs of Propositions

Proposition 1:   The program of a borrowing bank writes

$$\max_{\phi_t} pR_t(1 + \phi_t) - \rho_t\phi_t$$

$$s.t.\ pR_t(1 + \phi_t) - \rho_t\phi_t \geqslant \rho_t$$

$$\gamma(1 + \theta\phi_t) \leqslant \rho_t$$

The participation constraint indicates that only banks with ability $p \geqslant \overline{p}_t \equiv \rho_t/R_t$ will borrow. We focus on this segment of the market. The problem simplifies to

$$\max_{\phi_t} pR_t(1 + \phi_t) - \rho_t\phi_t$$

$$s.t.\ \gamma(1 + \theta\phi_t) \leqslant \rho_t$$

for $p \geqslant \overline{p}_t$. Let us denote by λ the Lagrange multiplier associated to the incentive constraint, the first order conditions are then

$$pR_t - \rho_t = \gamma\theta\lambda$$

$$\lambda(\rho_t - \gamma(1 + \theta\phi_t)) = 0$$

The result follows from λ being strictly positive for all $p > \overline{p}_t$.　　■

Proposition 3:   The market clearing condition (8a) in normal times together with the optimal demand for capital yields the normal times equilibrium corporate loan rate $R_t = z_t f_k(a_t) + 1 - \delta$. The interbank market freezes if and only if this rate is below $\overline{R}$, *i.e.*, if and only if $z_t f_k(a_t) < \overline{R} + \delta - 1$, which can be restated either as $a_t \geqslant \overline{a}_t \equiv f_k^{-1}((\overline{R} + \delta - 1)/z_t)$ or $z_t \leqslant \overline{z}_t \equiv (\overline{R} + \delta - 1)/f_k(a_t)$.　　■

Proposition 4:   At the end of period t, the probability that a crisis breaks out in $t + 1$ is the probability that R_{t+1} falls below $\overline{R}$, conditional on state (z_t, a_t), *i.e.* $\mathbb{P}(R_{t+1} < \overline{R}|z_t, a_t)$. The definition of the probability follows from the firm's first order condition and the definition of $\overline{z}_{t+1}$. Notice that

$$\mathbb{P}(z_{t+1} < \overline{z}_{t+1}|z_t, a_t) = \mathbb{P}\left(z_{t+1} < \frac{\overline{R} + \delta - 1}{f_k(a_{t+1})}\middle| z_t, a_t\right) = \mathbb{P}\left(z_{t+1} < \frac{\overline{R} + \delta - 1}{f_k(G(a_t; z_t))}\middle| z_t, a_t\right)$$

where $G(a_t; z_t) = a_{t+1}$ is the household's optimal decision rule, and $\partial G(a_t; z_t)/\partial a_t > 0$. Let $\omega_t = a_t/\overline{a}_t$ be the deviation of assets from banks' absoprtion capacity; then we have

$$\mathbb{P}(z_{t+1} < \overline{z}_{t+1}\middle| z_t, a_t) = \mathbb{P}\left(z_{t+1} < \frac{\overline{R} + \delta - 1}{f_k(G(\omega_t\overline{a}_t; z_t))}\middle| z_t, a_t\right) = \Phi\left(\ln\frac{\overline{R} + \delta - 1}{f_k(G(\omega_t\overline{a}_t; z_t))} - \rho\ln z_t\right)$$

where $\Phi(\cdot)$ is the Gaussian cdf. Differentiating with respect to ω_t, we get

$$\frac{\partial\mathbb{P}(z_{t+1} < \overline{z}_{t+1}|z_t, a_t)}{\partial\omega_t} = -\frac{\partial G(a_t; z_t)}{\partial a_t}f_{kk}(a_{t+1})a_t\Phi'\left(\ln\frac{\overline{R} + \delta - 1}{f_k(G(\omega_t\overline{a}_t; z_t))} - \rho\ln z_t\right)$$

which, given the concavity of the production function and the fact that the optimal asset accumulation decision is increasing in asset holdings, is positive.　　■

B   Equations of the Model with endogenous Labour Supply

1. $y_t = z_t k_t^\alpha h_t^{1-\alpha} + (\gamma + \delta - 1)(a_t - k_t)$

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

3. $\left(c_t - \vartheta \dfrac{h_t^{1+v}}{1+v} \right)^{-\sigma} = \beta \mathbb{E}_t \left[\left(c_{t+1} - \vartheta \dfrac{h_{t+1}^{1+v}}{1+v} \right)^{-\sigma} r_{t+1} \right]$

4. $h_t = \left(\dfrac{(1-\alpha)z_t}{\vartheta} \right)^{\frac{1}{v+\alpha}} k_t^{\frac{\alpha}{v+\alpha}}$

5. $\bar{a}_t \equiv ((1-\alpha)/\vartheta)^{\frac{1}{v}} \left(\alpha/\left(\overline{R}+\delta-1\right) \right)^{\frac{v+\alpha}{v(1-\alpha)}} z_t^{\frac{1+v}{v(1-\alpha)}}$

6. $i_t = \psi a_{t+1} - (1-\delta)a_t$

7. $y_t = c_t + i_t$

If $a_t \leqslant \bar{a}_t$ (normal times)

8a. $k_t = a_t$

9a. $\dfrac{r_t}{R_t} = \displaystyle\int_{\bar{p}_t}^1 p \dfrac{\mathrm{d}\mu(p)}{1-\mu(\bar{p}_t)}$

10a. $\bar{p}_t = \dfrac{\rho_t}{R_t}$

11a. $R_t = \dfrac{\rho_t}{\mu^{-1}\left(\frac{\rho_t - \gamma}{\rho_t - (1-\theta)\gamma} \right)}$, with $\rho_t > \bar{\rho}$

If $a_t > \bar{a}_t$ (crisis times)

8b. $k_t = a_t - \mu(\gamma/R_t)a_t$

9b. $\dfrac{r_t}{R_t} = \dfrac{\gamma}{R_t}\mu(\gamma/R_t) + \displaystyle\int_{\gamma/R_t}^1 p\,\mathrm{d}\mu(p)$

10b. $\bar{p}_t = \gamma/R_t$

11b. $\rho_t = \gamma$

Few comments are in order here. *(i)* During a crisis assets $a_t - k_t$ are stored for return γ. Since capital depreciates at rate δ, the value added of storage is $\gamma + \delta - 1$, as reflected in equation 1. (Remember that $\gamma + \delta - 1 > 0$.) *(ii)* The good market clearing conditions (7) is derived from Walras' law and the agents' budget constraints. Summing up the household and the firm's budget constraints (1) and (3), one gets: $c_t + \psi a_{t+1} = z_t k_t^\alpha h_t^{1-\alpha} + (1-\delta)k_t + r_t a_t - R_t k_t + \chi_t$, which simplifies to $c_t + i_t = y_t - (\gamma+\delta-1)(a_t-k_t) + (1-\delta)(k_t-a_t) + r_t a_t - R_t k_t + \chi_t \Longleftrightarrow y_t + \chi_t = c_t + i_t + (R_t - r_t)a_t - (R_t - \gamma)(a_t - k_t)$. It can be checked that $\chi_t = \int_{\bar{p}_t}^1 (1-p)(1+\phi_t)R_t a_t\,\mathrm{d}\mu(p)$, and that the resource identity reduces to $y_t = c_t + i_t$. In crisis times, by definition one has $\chi_t = \int_{\gamma/R_t}^1 (1-p)R_t a_t\,\mathrm{d}\mu(p)$, which simplifies to (using 8b and 9b): $\chi_t = (1-\mu(\gamma/R_t))R_t a_t - a_t\int_{\gamma/R_t}^1 pR_t\,\mathrm{d}\mu(p) = R_t k_t - r_t a_t + \gamma(k_t - a_t) = (R_t - r_t)a_t - (R_t - \gamma)(a_t - k_t)$. In normal times, by definition one has $\chi_t = \int_{\bar{p}_t}^1 (1-p)(1+\phi_t)R_t a_t\,\mathrm{d}\mu(p)$, which simplifies to (using 8a and 9a and the fact that in equilibrium $1+\phi_t = 1/(1-\mu(\bar{p}_t))$): $\chi_t = (R_t - r_t)a_t$.