Business Cycles and the Balance Sheets of the Financial and Non-financial Sectors*

Alonso Villacorta†

February 2022

Abstract


Keywords: Macro finance, financial frictions, financial crises, balance sheet channel, financial intermediation.

JEL codes: E44, E32, G01.

*First draft: August 2016. I am grateful to Monika Piazzesi and Martin Schneider for their invaluable guidance. I also thank Juliane Begenau, Saki Bigio, Sebastian Di Tella, Andres Drenik, Daniel Garcia-Macia, Eran Blass-Hoffman, Han Hong, Matteo Iacoviello, Simon Gilchrist, Sebnem Kalemli-Ozcan, Arvind Krishnamurthy, Patrick Kehoe, Pablo Kurlat, Tim Landvoigt, Moritz Lenel, Claudio Michelacci, Diego Perez, Alessandra Peter, Adriano Rampini, Cian Ruane, Anatoli Segura, Javier Suarez, Chris Tonetti, Stijn Van Nieuwerburgh, and Lucciano Villacorta for helpful comments. Financial support from the Bradley Fellowship Program through a grant to SIEPR and the Macro Financial Modeling Initiative is acknowledged.

†Department of Economics, University of California, Santa Cruz. Telephone: (+1) 650-798-7703. Email: avillaco@ucsc.edu. Website: https://sites.google.com/site/alonsovillacorta/.
1 Introduction

During the 2008 financial crisis, banks suffered large losses to their asset values which led to a banking crisis and massive bailouts, followed by a long-lasting recession. This paper analyzes the importance of the health of banks’ balance sheets for the severity of economic recessions. I ask, are banks “special” in the sense that shocks that affect banks’ balance sheets matter more for real economic activity than shocks that affect other firms?\textsuperscript{1} For the policy debate, it is crucial to understand the mechanisms by which shocks that affect banks or firms propagate to real activity. For instance, is recapitalizing banks more effective than recapitalizing other borrowers during recessions?

There is empirical evidence that both shocks that affect firm net worth (firm credit channel\textsuperscript{2}) and shocks to bank net worth (bank lending channel\textsuperscript{3}) propagate through balance sheet constraints and affect activity. However, few general equilibrium frameworks consider both channels simultaneously. This paper provides a model in which both channels are present and can quantitatively capture the differential dynamics of macroeconomic and financial variables in different recessions. Having a framework that quantitatively considers both endogenous channels is important to speak to whether banks are special and to guide the policy response in the most effective way.

To motivate the analysis, Figure 1 shows two recessions that differentially affected the financial and the non-financial sectors. The figure displays the evolution of aggregate net worth, as measured by the market value of equity, in each sector as a ratio of GDP during the recessions of 2001 and 2008. The whole corporate sector experienced large wealth losses in both recessions (panel 1a). However, the consequences on real activity were different—the recession of 2001 was shorter and milder. Panel 1b shows that while the 2001 recession severely affected the non-financial sector, it mildly affected financial sector balance sheets, which instead severely deteriorated during the Great Recession.\textsuperscript{4}

Moreover, Figure 2 displays the dynamics of both quantities and prices in the bank lending

---

\textsuperscript{1}For the rest of the paper, whenever I refer to banks as special indicates that shocks that affect banks’ balance sheets matter more for activity than shocks that affect firms’ balance sheets.


\textsuperscript{3}See Bernanke and Blinder [1992], Bernanke and Gertler [1995], Kashyap and Stein [2000], Ivashina and Scharfstein [2010] and Chodorow-Reich [2013].

\textsuperscript{4}In his 2012 speech, Fed Chairman Bernanke stressed: “any theory of the crisis that ties its magnitude to the size of the housing bust must also explain why the fall of dot-com stock prices, which destroyed as much or more paper wealth—more than $8 trillion—resulted in a relatively short and mild recession”. See Ben S. Bernanke (2012), ”Some Reflections on the Crisis and the Policy Response”, at the Russell Sage Foundation and the Century Foundation Conference.
market during both recessions. Panel 2a shows the share of bank loans in total firm debt while panel 2b shows the cost of bank loans. Both series show different dynamics during these two recessions. The share of bank loans experienced a substantial drop during the 2008 recession while the change was small in 2001. The cost of bank loans shows a large increase during the Great Recession while a small change in 2001. This differential response of the market of bank lending suggests that the mechanisms in place during these two recessions were different.

In this paper, I develop and estimate a framework to investigate the mechanisms that emerge in different kinds of balance sheet recessions: when banks’ balance sheets are more affected (e.g. 2008) or when firms’ are more affected (e.g. 2001). I propose a dynamic model of financial intermediation with three agents: lenders (households), borrowers who are experts at production (firms), and banks who are experts at monitoring and use their skills to borrow and make risky loans to firms. Contrary to most previous work, frictions in all lending relationships imply that both bank net worth and firm net worth jointly affect equilibrium dynamics. The first theoretical result shows that when bank net worth is not too low relative to firm net worth, changes in firm and bank net worth have the same effect on economic activity, even under frictions between banks and firms. The second theoretical result shows that bank net worth is special when the share of bank net worth is sufficiently low—I refer to these times as banking crises. That is, I uncover a new amplification mechanism—a bank accelerator active only in banking crises—by which shocks to bank net worth have an additional propagation effect on activity beyond that generated by the standard financial accelerator, so they matter more than shocks to firm net worth. This non-
Figure 2: Bank Lending Market

(a) Share of Bank Loans in Total Firm Debt

(b) Cost of Bank Loans

Note. Panel (a): Share of loans in all credit market instruments (outstanding), for non-financial corporations. Source: US Flow of Funds. Panel (b): Weighted average of the cost of new loan issuances (in bps) extended for general corporate purposes and liquidity management. Cost is defined as all-in-drawn spread, which is total (interest plus recurring fees) spread paid over 6 month LIBOR. Source: DealScan database of loan originations. Panel reports the raw series (dashed line) and its smoothed version (solid line).

Linearity in the model leads to different kinds of recessions that are consistent with the findings from the empirical literature on financial crises. Also, this novel bank accelerator implies that policies directed to recapitalize banks are more effective than policies directed to firms during banking crises but not in all types of recessions.

The second contribution is quantitative: I show that the model help in explaining key macroeconomic and financial data. For instance, the recessions of the early 90s and 2008 were accompanied by large drops in bank net worth relative to the 2001 recession and also were associated with steeper and more persistent declines in output and in the quantity of bank loans. The estimation finds that the bank accelerator was active in those banking crises but not in the recession of 2001. And, this new mechanism can explain these key stylized facts.

To illustrate how the mechanism works, it is helpful to consider the model in more detail. In the model, both firms and banks face moral hazard problems that lead to endogenous borrowing constraints. Firms can invest in a technology subject to both idiosyncratic and aggregate shocks, and with constant returns in two factors of production, one which must be financed in advance. Firms have access to two sources of financing: directly from households or intermediated loans from banks. But, due to moral hazard problems, firm net worth constrains the amount of borrowing from both types of financing (debt must be secured by collateral). Households only accept riskless debt. Banks are special in their monitoring skills, which allows them to make firms

5 See Related Literature (sub-section 1.1).
repay additional amounts in good states, and thus extend risky loans. One interpretation is that households only value the riskless part of firm assets as collateral, while banks accept part of firm risky assets to secure their loans. So, intermediated finance, while subject to credit risk, allows firms to raise additional funds. Banks, by lending to multiple firms, pool idiosyncratic risks from their portfolio. This allows banks to borrow additional funds from households since they can use their diversified portfolio as collateral. This bank intermediation process—diversification and creation of collateral—increases the available resources for investment. However, banks’ loan portfolios are exposed to aggregate risk. Thus, banks’ net worth, which secures their borrowing from households, constrains the supply of intermediated funds. As a result, both firm net worth and bank net worth are state variables and can potentially have a differential effect on economic activity.

The state space can be divided into different regions that are endogenously defined by the sizes of bank and firm net worth. Each of these regions characterizes a different regime for economic dynamics. The behavior of the market of bank lending (equilibrium loan spread and volume), the response of aggregates to shocks and the relative importance of bank and firm net worth for activity is different in each regime. For instance, when the bank and firm net worth are not far from the steady state, the economy features the standard financial accelerator that works through the aggregate net worth of banks and firms. Instead, when the share of bank net worth is critically low (i.e. the ratio of bank to firm net worth is below a threshold), the intermediation process becomes disrupted leading to a propagation mechanism additionally to the financial accelerator. That is, the economy can enter into different kinds of recessions depending on whether bank net worth is critically low or not, which I label: (i) aggregate net worth recessions and (ii) banking crises.

Aggregate net worth booms and recessions are defined as times for which the share of bank net worth is not too low. In this regime, the bank loan rate is in an intermediate region: (i) it is not too high so firms borrow up to the limit both from banks and households, and (ii) it is not too low so banks borrow up to the limit from households and lend all funds to firms. That is, all constraints are binding. A financial accelerator appears by which shocks to both bank and/or firm net worth are amplified and have a persistent effect on aggregate investment and output. The novel result is that bank and firm net worth have the same effect on aggregate activity, so that

---

6Only banks can do the diversification of firms risky assets because a monitoring technology is required to hold risks from multiple firms. Thus, while firms cannot use the risky part of their assets as collateral with households, banks can use their risky loans due to diversification.

7The paper focuses on these two types of recessions for which the net worths of banks and firms are not too large so their constraints with households are always binding. However, if the net worth of banks and firms are large enough, all constraints are slack and neither shocks to firm nor bank net worth affect activity.
only the aggregate net worth matters for aggregate dynamics while the distribution is irrelevant.\footnote{Note that this aggregation result is different from the existing literature that assumes frictionless relations between firms and banks. In this regime, financial frictions are always present in the loan market: all constraints are binding.}

In this regime, a drop in bank net worth tightens banks’ constraints and hurts banks’ ability to make loans. The reduction in the supply of loans increases loan rates, which in turn helps banks. The higher cost of loans affects firm financing constraints and hurts activity. Even when firms are willing to borrow at the higher rates, the higher cost of loans reduces the maximum funds firms can get given their net worth (firm constraints get tightened). However, the effect is identical to just a drop in firm net worth. Through the loan rate equilibrium response, banks pass on to firms the costs generated by shocks to their net worth. As a result, in equilibrium, the effect on activity from a drop in bank net worth is equivalent to the effect from a drop in firm net worth. Hence, recapitalizing banks has the same effect than recapitalizing firms.

Similarly, a shock that eases bank funding constraints will produce an expansion of credit supply and propagate to firms’ investment through a reduction in loan rates. This loan rate pass-through mechanism from credit supply shocks to firm borrowing and investment is consistent with the evidence provided by Giovanni, Kalemli-Ozcan, Ulu, and Baskaya [2017], which show using loan-level data for Turkey that the boom phase of the Global Financial Cycle that relaxed bank funding constraints propagated to firms through lower loan rates instead than changes in posted collateral values.

**Banking crises** are times in which bank net worth is critically low. In this regime, the cost of loans is remarkable high relative to firms’ expected profitability. The equilibrium loan rate rises to the maximum firms are willing to accept given the return on their assets. Even when firms are constrained and borrow up to the maximum from households, they choose to borrow less risky loans than the maximum allowed by banks (firm constraint with households is binding, while constraint with banks is not). Banks, which profit from the high rates, borrow the maximum they can from households given their net worth (bank constraint with households is binding). In this regime, a drop in bank net worth contracts the supply of loans but banks cannot respond by increasing the rates over the maximum firms are willing to accept. The limited ability of banks to raise the loan rate—and pass on the rise in their borrowing costs to firms—leads to a more significant adjustment in the quantity of loans, relative to aggregate net worth recessions. Since banks use their loans to create collateral and get additional funds from households, this reduction in lending implies a loss of intermediation—a drop in collateral created by banks per unit of net worth—that further constrains bank supply of loans and amplifies the initial shock. This feedback effect through the intermediation process is unique of the banking crisis regime and is the source
of the new amplification mechanism in the model, which I refer as the bank accelerator. The bank accelerator implies that a drop in bank net worth has more severe consequences on real activity than a drop in firm net worth. It also implies that banking crises are longer and more severe than other recessions and favors policies directed to recapitalize banks rather than firms during these times.

Different shocks can move the economy into the different regimes. In reality, different kinds of shocks can affect the assets of financial and non-financial firms: some shocks affect relatively more non-financial firms (e.g. the forces that trigger the dot-com bust in 2001), others affect more financial intermediaries (e.g. the housing crisis in 2008). The model focuses on the propagation mechanisms and explains how different shocks amplify, propagate and can differentially affect the real economy depending on their effect on bank and firm net worth. For tractability, in the quantitative exercise I allow for only two different shocks that can move the economy into the different regimes. First, a productivity shock affects the return on assets of firms, which in turn affects bank loan repayments; thus a productivity shock affects both firm and bank net worth. Second, a bank-specific shock that affects only bank net worth; this shock tries to capture other bank businesses such as residential mortgage lending, which are absent in the model. While the bank shock does not directly affect the assets of the firm, it can indirectly propagate by its effects on the market of bank lending.

In order to quantify the importance of the bank accelerator, I estimate the model using U.S. data on output (real GDP) and the aggregate market value of bank net worth from 1985 to 2015. The estimation is based on a maximum likelihood approach. The model is dynamic and features non-linearities due to the different regimes. I use a particle filter, as suggested by Fernández-Villaverde and Rubio-Ramírez [2007], to deal with the non-linear system.

First, given the estimated parameters and latent variables, I identify when did the U.S. enter into banking crises from the lens of the model. The estimation identifies the recession of the early 90s and 2008 as banking crises, as both are associated with large drops in the relative share of bank net worth; while the recession of 2001 is not.

Second, I quantify the importance of the non-linear bank accelerator—which is only activated during banking crises—by comparing the dynamics of the model relative to an alternative in which this mechanism is shut down and only the standard financial accelerator remains. In this alternative exercise, the flow of credit through the financial sector is not disrupted, for example as occurred in the 2001 recession. I find that the bank accelerator explains on average 40% of the fall in output and 78% of the fall in bank net worth during the Great Recession. Since the bank accelerator restricts banks’ ability to lever and profit from the high spreads, bank net
worth recovery is slower when the economy enters into the banking crises regime. Thus, the new mechanism induces additional persistence and implies longer recessions associated with banking crises. The endogenous dynamics can account for output and bank net worth remaining below their trend for several years after the Great Recession. As a result, the average one-year forecast errors increase, by 25% for output and 60% for net worth, when the bank accelerator is not considered. Similar improvements are found for the recession of the early 90s.

Finally, the model can explain the differential dynamics observed in the share of bank loans in total firm financing and the cost of credit during the recessions of 1990, 2001, and 2008. These series are not used in the estimation; thus they serve as external validation of the model. The estimated model generates much larger drops in the share of bank loans in the recessions of 1990 and 2008, for which the bank accelerator is active, relative to the recession of 2001. These dynamics are consistent with the observed share of bank loans in firm borrowing. Also, the estimated model can capture the pronounced increase in the loan rate in 2008 and the milder increases in 2001 and the early 1990s. Interestingly, the estimated model can produce such differential response in both identified banking crises: the large increase in the 2008 recession and the small change in the early 90s. These two financial recessions are different because of two reasons. First, the drop in bank net worth is much larger in 2008. Second, in the model, what matters to activate the bank accelerator is the loan rate relative to the return on assets of firms. The estimation finds that the early 90s recession was associated with a relatively larger productivity shock that reduced firm expected returns, whereas the Great Recession was mainly triggered by a specific shock to banks. Thus, in the early 90s, a small increase in the loan rate was sufficient to make firms contract their demand of bank loans below their limits, which triggered the mechanism that can simultaneously explain the drop in bank loans and the persistence in output and net worth.

1.1 Related Literature

This paper builds on the literature that stresses the monitoring role of banks. In Diamond [1984] or Krasa and Villamil [1992], banks appear as delegated monitors since diversification reduces duplicative monitoring costs, but there is no role for bank net worth in intermediation. Instead in Holmstrom and Tirole [1997] the intensity of monitoring depends on bank and firm net worth, but there is no diversification. These models are static, the net worth of agents is exogenous, and they do not focus on macroeconomic dynamics.

Dynamic models in which the net worth of borrowers plays a role are pioneered by Bernanke and Gertler [1989] and Kiyotaki and Moore [1997]. Bernanke et al. [1999] embed those features
in a macroeconomic framework suitable for quantitative analysis, in which constrained borrowers are entrepreneurs in the non-financial sector. These models feature the financial accelerator as propagation mechanism, where the net worth of borrowers—the productive agents—matters for dynamics. Gertler and Kiyotaki [2010], He and Krishnamurthy [2013], Brunnermeier and Sannikov [2014], Begenau [2019] and others consider an intermediary sector. These papers assume no frictions between banks and firms, so that only the sum of bank and firm net worth—the consolidated borrowing sector—matters for the dynamics. These models help understanding why balance sheets play a role when financial frictions limit the availability of funds to borrowers. However, they are silent about what is particularly special about banks’ balance sheets and banking crises: they cannot address whether shocks to banks matter more for activity than shocks to other borrowers and what a redistribution from other borrowers to banks would entail.

Iacoviello [2015] includes frictions between banks and firms. In his model, entrepreneurs borrow from banks subject to a constraint—without banks, entrepreneurs cannot get any outside funding—while banks face frictions in their borrowing from households. His estimation assume constraints are always binding and attributes 2/3 of the decline in GDP during the Great Recession to financial shocks. Instead, my paper documents that non-linearities are important in an estimated model in which constraints only bind occasionally. These non-linearities are particularly relevant to explain the empirical differences between banking crises from other recessions (e.g. 2001 vs 2008).

A few recent papers include frictions between banks and firms and study non-linear dynamics with occasionally binding constraints. The closest paper to mine is Rampini and Viswanathan [2019], which develops a theoretical model with endogenous constraints in which both bank and firm net worth jointly determine economic dynamics. In their model, banks are better able to enforce collateralized claims than households, so they can lend more to firms. But, the additional amount banks lend can only be financed out of their own net worth, giving a role to bank net worth. A key difference in my model is that banks, by diversifying risk and creating collateral, channel additional funds from households to firms. This channel makes bank net

---

9 More recent contributions that feature a financial accelerator at the firm level and quantify the importance of different financial shocks include: Jermann and Quadrini [2012], Liu, Wang, and Zha [2013], Christiano, Motto, and Rostagno [2014], or Bigio [2015].

10 See Brunnermeier, Eisenbach, and Sannikov [2012] for an overview of the literature of macroeconomic models with financial frictions.

11 Moreover, to be able to speak about the composition of firm debt, I allow firms to borrow from banks as well as other lenders. I find that frictions in the bank lending market are important even when firms are allowed to borrow directly from households.
worth *special* and is the source of the new feedback mechanism in the model.\textsuperscript{12,13} I show that this new mechanism is quantitatively important by estimating my model. Elenev, Landvoigt, and Van Nieuwerburgh \[2021\] and Mendicino et al. \[2019\] provide quantitative models that are used to study optimal capital requirements. They find that restrictions on bank leverage trades off stability against restricting credit provision. These papers generate slower recoveries from recessions that are accompanied by a credit crunch.\textsuperscript{14} But, they are silent about the differential effect from a drop on bank versus firm net worth and whether is more effective to recapitalize banks or firms.

In other papers that consider occasionally binding constraints (e.g. Elenev et al. \[2021\], He and Krishnamurthy \[2013\] and Guerrieri and Iacoviello \[2017\]), the borrowing constraint of firms is binding during recessions. Instead, in my model, firms face two types of constraints: on their borrowing from banks and from households. Importantly, during banking crises, firms’ constraints with banks are slack, while firms’ and banks’ constraints with households bind. The fact that firms do not borrow up to the maximum affects banks’ ability to create collateral and, in turn, constrains further banks’ ability to borrow, inducing the new amplification mechanism.

This paper is also related to the empirical literature on financial crises, for instance, Bordo et al. \[2001\], Cerra and Saxena \[2008\], Reinhart and Rogoff \[2009\], or Jordà et al. \[2013\]. This literature generally concludes that recessions accompanied by financial crises are more severe and persistent than other recessions. More recently, Romer and Romer \[2015\] and Krishnamurthy and Muir \[2016\] highlight the importance of distinguishing banking crises by their severity. Both papers find that crises associated with more severe measures of financial distress are associated with slower recoveries. My model generates banking crises and other financial recessions. The endogenous mechanism in my model explains why are banking crises more severe and persistent, and also associates the severity of the downturn with the decline in bank equity and the resulting cost of credit intermediation. Thus, my paper provides a theory that supports the empirical

\textsuperscript{12}In the absence of diversification in my model, bank net worth would still matter for aggregate dynamics, but the bank accelerator would disappear: banks would not be special and recapitalizing them would not be more effective.

\textsuperscript{13}Rampini and Viswanathan \[2019\] emphasize that Holmstrom and Tirole \[1997\], in which firm and bank capital (net worth) are exogenous, conclude: “A proper investigation ... must take into account the feedback from interest rate to capital values. This will require an explicitly dynamic model along the lines of Kiyotaki and Moore \[1997\].” Let me add that Holmstrom and Tirole \[1997\] also add to their conclusion: “Another caveat concerns that intermediary’s projects are perfectly correlated ... It is just a way of avoiding the extreme conclusion that all intermediation can be carried without own capital. We see the issue of diversification, the degree of leverage and the intensity of monitoring as closely linked, complementary choice variables that deserve more careful study in the future.” Rampini and Viswanathan \[2019\] and my paper provide dynamic models where the joint evolution of firm and bank capital and rates are endogenously determined. I contribute by emphasizing the role of diversification and its relationship with bank capital.

\textsuperscript{14}Mendicino et al. \[2019\] in particular stress that firm and bank defaults are correlated due to the structure of contracts, so that these types of “twin” recessions are more likely than under standard default risk models.
literature. Moreover, I provide a structural framework to identify banking crises episodes. For instance, the estimation identifies the recessions of 1990 and 2008 as banking crises which are accompanied with large equity declines, being the 2008 recession more severe. In a recent paper, Baron, Verner, and Xiong [2020] construct a new dataset on bank equity and non-financial equity returns for 46 countries over the period 1870-2016. Consistent with my model, they show that large bank equity declines are associated with substantial credit contractions and output gaps. They also highlight that large bank equity declines can uncover a number of historical banking crises not accompanied by banking panics, including the recession of the 1990s described as a prominent episode of severe non-panic banking distress associated with a prolonged recession and credit crunch.

**Layout.** Section 2 describes the main equations that drive the dynamics of aggregates used in the quantitative exercise. Section 3 describes the model. Section 4 characterizes the state space and the partition that defines the different regimes: i) aggregate net worth booms and recessions, and ii) banking crises. Section 5 describes the equilibrium dynamics in each regime. Section 6 discusses the quantitative exercise. Section 7 concludes. All proofs and appendices are in the Supplemental Material.

## 2 The role of bank net worth: a stylized model

In this section, I present the equations that drive the dynamics of main aggregate variables in my model. The purpose of this section is to provide the intuition behind the main equations used in the estimation without deriving them from first principles.

Financial frictions imply a relationship between investment and the net worth of borrowers. In models with financial constraints, the borrowers’ net worth determines the collateral value that firms can provide to their creditors. Thus, models of financial frictions imply

\[ K_t \leq \lambda_t N_t, \]

where \( K_t \) captures the value of aggregate assets/capital, \( N_t \) is aggregate net worth of borrowers and \( \lambda_t \) is a multiplier associated with financial constraints; when the constraint is binding \( \lambda_t \) represents leverage. Borrowers invest \( 1/\lambda_t \) per unit of capital from their own net worth and the rest, \( B_t = K_t - N_t \), is the value of debt borrowed from creditors. Denoting \( R \) as the promised return to creditors and \( K_{t+1}^R \) as the return on capital, we have the following law of motion for net
worth
\[ N_{t+1} = R_{t+1}^{K} K_t - RB_t. \] (2)

Equations (1) and (2) show the dynamic complementarities between capital and net worth, which are the main forces behind financial friction models: Shocks that affect net worth \( N_t \) affect investment, capital and output, which in turn affect future values of net worth, capital, and output. Through this mechanism, even temporary shocks can generate cycles.

Two important questions arise: i) which agents’ net worth is represented by \( N_t \)?, and ii) what determines the constraint multiplier \( \lambda_t \)? First, both banks and firms are borrowers that invest and share the returns and risks of real projects, while facing frictions from other creditors. In my model, both firm net worth and bank net worth matter so \( N_t \) represents the sum of their net worth.\(^{15}\)

Second, the degree of financial frictions, captured by \( \lambda_t \), captures the functioning of the financial sector. Banks act as intermediaries which alleviate frictions and help funds to flow from creditors to borrowers. During normal times, funds flow from creditors to firms through banks, implying high levels of borrowing and investment per unit of net worth (\( \lambda_t \) high). During banking crises, banks intermediation process is disrupted and credit per unit of net worth falls, implying a reduction in investment (\( \lambda_t \) low).\(^{16}\)

Banks are simultaneously borrowers and lenders, and have frictions on both sides of their balance sheets. Banks’ net worth determines the tightness of their balance sheets constraints and their debt capacity, which in turn determines the supply of funds intermediated to firms. As a result, in my model the multiplier \( \lambda_t \) is endogenous and a function of the relative net worth of banks \( \beta_t = N_t^B / N_t \)
\[
\lambda_t = \begin{cases} 
\bar{\lambda} & \text{if } \beta_t \geq \bar{\beta}_t, \\
\lambda(\beta_t) & \text{if } \beta_t < \bar{\beta}_t \text{ (banking crises)},
\end{cases}
\] (3)
where \( \lambda(\beta_t) \leq \bar{\lambda} \) and \( \lambda(\beta_t) \) is increasing in \( \beta_t \). This specification differentiate recessions depending on the magnitude of bank net worth losses. The persistence and severity of the recession depends on the evolution of \( \beta_t \). Denoting the portfolio of bank loans as \( L_t \) with corresponding return \( R_{t+1}^L \)

\(^{15}\)Gertler and Kiyotaki [2010] interpret \( N_t \) as a consolidated net worth of borrowers and their calibration matches an average leverage ratio across financial and non-financial sectors. Instead, Liu et al. [2013] and Christiano et al. [2014] use the market value of non-financial firms’ equity to estimate their model.

\(^{16}\)Papers like Jermann and Quadrini [2012] or Liu et al. [2013] find financial shocks are important in explaining business cycle variations. Their financial shocks interpreted as “collateral shocks” capture exogenous shocks to \( \lambda_t \). Other papers like Christiano et al. [2014] include “risk shocks” which are exogenous variations to the dispersion of firm profitability, which impact lending contracts and reduce the amount of borrowing and investment per unit of net worth, that is: shocks to \( \lambda_t \).
(loan rate), the evolution of bank net worth follows

\[ N_{t+1}^B = R_{t+1}^L L_t - RD_t, \tag{4} \]

where \( D_t \) is the aggregate debt of banks. The evolution of \( \beta_t \) follows from (2) and (4).

The model described by equations (1)-(4) allows to differentiate recessions in which firms balance sheets are affected and total net worth \( N_t \) drops (e.g. the 2001 recession) from recessions in which, in addition, bank net worth is severely affected and \( \beta_t \) drops (e.g. the 2008 recession).

Finally, the model has implications for the dynamics of the volume and cost of bank loans that vary across recessions. For instance, the volume of bank loans decrease more during banking crises relative to other recessions, while loan rates increase relatively more. Denoting the share of bank loans in total firm debt with \( share_t^L \), the model implies functions

\[ share_t^L = s^L(N_t, \beta_t), \tag{5} \]
\[ R_{t+1}^L = r^L(N_t, \beta_t), \tag{6} \]

where \( s^L(.) \) is increasing in \( \beta_t \) and \( r^L(.) \) is decreasing in \( \beta_t \). In the next section, I micro-found the equations presented in this section.

### 3 Model

In this section, I develop a model where financial frictions constrain the flow of funds to the productive sector. I consider a dynamic economy that is populated by three types of agents: a representative household, a continuum of firms, and a continuum of banks. Time is discrete and the horizon infinite.

Firms face two types of borrowing constraints: \( i \) they can only borrow through riskless debt from households; and, \( ii \) they can borrow through risky loans from banks, but just up to a fraction of the value of their risky assets. Banks also face a borrowing constraint with households: they can only borrow through riskless debt. Figure 3 illustrates the constraints. For the ease of exposition, I take these borrowing restrictions as given. However, in Appendix B, I show that they arise as part of a contracting problem with moral hazard.
3.1 Representative household

The representative household has linear utility and discounts the future by the factor $1/R$. The household represents unconstrained agents assumed to have large wealth and willing to offer any amount of funds for a return $R$.\(^{17}\) However, households only buy riskless assets: bank deposits or riskless firm bonds.\(^{18}\) Also, the household is owner of a factor of production $H_t = 1$ that is in fixed supply and rents to firms at price $w_t$.

3.2 Firms (entrepreneurs)

There is a unit measure of agents who run firms, which can be thought of as owners or entrepreneurs. Throughout the paper, I say entrepreneurs or firms to refer to those agents. Firms are indexed by $i \in \mathbb{I} = [0, 1]$. Firms have access to a technology with two factors of production: $k$ and $h$. For each $k_{i,t-1}$ units invested at time $t-1$ and $h_{i,t}$ at time $t$, firm $i$ produces at time $t$

$$y_{i,t} = z_{i,t}k_{i,t-1}^{\alpha}h_{i,t}^{1-\alpha},$$

where $z_{i,t}$ denotes an idiosyncratic productivity shock distributed Bernoulli($Z_t$). Thus, a fraction $1 - Z_t$ of firms fail on their projects and have zero production. Moreover, if the project fails ($z_{i,t} = 0$), in addition to the loss of production, a fraction $\kappa$ of $k_{i,t-1}$ is lost.

\(^{17}\)The riskless rate $R$ can be interpreted as the return in a low-productivity project that all agents can access. 
\(^{18}\)Appendix B shows that, due to moral hazard faced by firms and banks, households with assumed linear utility only buy riskless assets because they are uninformed, so they only accept those contracts.
**Factors of production:** The factor $k_{i,t-1}$ represents all factors of production that need to be purchased in advance to start the project, while $h_{i,t}$ are the factors of production that can be acquired at the same time of production. It is common to think that $k$ represents physical capital which need to be purchased in advance, while $h$ represents labor. However, we could think that also part of the labor input must be hired and financed in advance.\(^\text{19}\) For convenience, I refer to the assets $k$ as just capital. In the estimation of the model, the factor $k$ is treated as a latent variable that represents any factor needed to be financed in advance: physical or organizational capital, or part of the workforce.

A firm that starts period $t$ with invested capital $k_{i,t-1}$ takes as given the price $w_t$ and decides how much labor $h_{i,t}$ to hire before the realization of the shock $z_{i,t}$. If the project turns out to be unsuccessful ($z_{i,t} = 0$), firms do not produce and those $h_{i,t}$ workers cannot be reallocated to other firms.\(^\text{20}\) Since the aggregate amount of the factor $H$ is in fixed supply and normalized to one, market clearing in the market for $H$ implies that the return per unit of capital invested by each entrepreneur is

\[
R^K_{i,t} = \begin{cases} 
1 + \alpha K_{t-1}^{\alpha-1} & \text{with prob. } Z_t \\
1 - \kappa & \text{with prob. } 1 - Z_t
\end{cases}
\tag{7}
\]

where $K_{t-1}$ denotes the aggregate investment in capital at $t - 1$. From the perspective of each entrepreneur, there are constant returns in their investment. However, there are decreasing returns on aggregate capital $K_{t-1}$. If all entrepreneurs increase their investment at $t - 1$, then at $t$ they compete for the factor $H$, increasing its price $w_t$ and reducing the return on $k_{t-1}$.\(^\text{21}\)

The relevant prices for firms’ borrowing and investment decisions are the return on assets $R^K_{i,t+1}$ and the borrowing rates (defined next). Hence, I directly use the equilibrium condition (7) which arises from the clearing of the frictionless market for the factor $H$. Since the focus of the paper is on firms’ and banks’ balance sheets, throughout the paper I abstract from the factor of production $H$ and its price $w_t$. Those aspects are described in Appendix A.

The fraction of firms $Z_{t+1}$ that succeed in their projects ($z_{i,t+1} = 1$) is stochastic: it follows an i.i.d. process with expected value $\overline{Z}$ and lowest possible realization $Z_L > 0$. The probability $Z_{t+1}$ is the only source of aggregate risk in the model.\(^\text{22}\)

Entrepreneurs start period $t$ with net worth $n^F_{i,t}$ and can raise financing from both households...\(^\text{15}\)

\(^{19}\)Models in which the financial constraint directly limits the labor input includes: Jermann and Quadrini [2012], Bigio [2015], or Arellano et al. [2019].

\(^{20}\)This assumption of no reallocation of workers is used for simplicity since it keeps the process of firms’ return on capital as binomial.

\(^{21}\)The assumption of decreasing returns on aggregate $K$ is used to ensure the economy converges to a steady state. However, the main results of the paper do not rely on this assumption.

\(^{22}\)In Section 6, I extend the model to include other aggregate autocorrelated shocks in order to match the data.
and banks through the issuance of riskless debt or risky loans. I denote with $b_{i,t}$ the amount of funds borrowed through riskless debt and $l_{i,t}$ the amount borrowed through risky loans. Total capital invested by entrepreneur $i$ at period $t$ is

$$k_{i,t} = n_{i,t}^F + b_{i,t} + l_{i,t}. \tag{8}$$

**Firm borrowing constraints.** Firms have two sources of financing, each subject to a different constraint. First, the amount of riskless debt that they can issue is limited by the safe part of their assets, which is determined by the return in case of failure ($z_{i,t} = 0$). Given the risk-free rate $R$, we have that the total riskless promise $Rb_{i,t}$ is limited by

$$Rb_{i,t} \leq (1 - \kappa)k_{i,t}. \tag{9}$$

Second, firms can borrow through risky loans from banks. I denote with $R_{i,t+1}^L$ the promised loan rate, which is repaid in case of project success. There is a competitive market of bank loans, so all entrepreneurs, with ex-ante identical probability of success, face the same loan rate $R_{t+1}^L$. Thus, the stochastic loan rate promised at $t$ denoted $R_{i,t+1}^L$ is

$$R_{i,t+1}^L = \begin{cases} R_{t+1}^L & \text{if } z_{i,t+1} = 1 \\ 0 & \text{if } z_{i,t+1} = 0. \end{cases} \tag{10}$$

Firms are also constrained in the amount of risky loans they can issue. Their promised pay-off must be secured by the fraction $\phi$ of the risky part $\kappa$ of their capital. Given the loan rate $R_{t+1}^L$, the total promise from the risky loans issued $R_{i,t+1}^L l_{i,t}$ is limited by

$$R_{t+1}^L l_{i,t+1} \leq \phi \kappa k_{i,t}. \tag{11}$$

The parameter $\phi$ represents the monitoring technology of banks. Due to their monitoring advantage, banks accept loans that pay differently in case of success or failure. An interpretation is that they are willing to accept the risky part $\kappa$ of firm assets $k_{i,t}$ as collateral, but only up to the fraction $\phi$.23

---

23Appendix B describes the moral hazard problem faced by firms and the role of bank monitoring. It also shows the implementation of contracts by a mix of riskless debt and zero-recovery risky loans. Appendix E extends the constraint to be cash-flow based, so that firms can also use a fraction of their output $y_{i,t}$ as collateral.
Discussion. Notice that in case of failure there is zero recovery value for the loan. This is just a normalization. Given the binomial feature of the project, any loan with positive recovery value can be divided into a safe security that promises the recovery value and a risky loan with zero recovery. The safe debt \( b_{i,t} \) already captures all safe securities (or loans’ recovery value) that can be issued, which could be held by either households or banks. The share of firms’ safe debt held by households or banks is undetermined in the model. For the sake of exposition, I assume all \( b_{i,t} \) is held by households. This is also without loss of generality, we could equivalently assume that all \( b_{i,t} \) are extended by banks, which borrow these amount from households. This indirect implementation of the lending that can (equivalently) be directly provided by households is also present in Holmstrom and Tirole [1997] and Rampini and Viswanathan [2019], and they, too, focus on the direct implementation. Therefore, in addition to simplifying notation and analysis, assuming that all \( b_{i,t} \) is held by households highlights the crucial difference of intermediated finance in my model relative to these previous papers. As we will see below, a key difference between \( l_{i,t} \) and \( b_{i,t} \) is that \( l_{i,t} \) represents the part of intermediated finance for which the net worth of the bank is essential.

Investment and the down-payment required. How much capital can firms invest for given net worth \( n_{i,t}^F \)? Borrowing constraints imply that the amount of capital is constrained by the net worth of the firm. By combining the borrowing constraints (9) and (11) with the flow of funds restriction (8) we get

\[
k_{i,t} \leq \left( \frac{1}{1 - \frac{(1-\kappa)}{R} - \frac{\bar{\phi}\kappa}{R_{i,t+1}}} \right) n_{i,t}^F.
\]

(12)

The expression above represents the maximum leverage of firms when borrowing up to the limit from households and banks. The denominator represents the minimum funding that needs to be provided by an entrepreneur per unit of capital, since the fractions \((1 - \kappa)/R\) and \(\bar{\phi}\kappa/R_{i+1}\) of each unit of capital can be financed by raising funds from households and banks, respectively.

Importantly, we can see from (12) that the maximum investment depends on the net worth of firms \( n_{i,t}^F \) and on the loan rate \( R_{i,t+1}^L \). Both objects are endogenous in the model.

Net worth dynamics and the firm problem. In order to limit firms’ ability to save and overcome financial constraints, I assume that firms exit with probability \( \tau_x \). Firms have the same discount rate as households \( 1/R \). Thus, upon exiting, firms just become households. Given linear utility and that firms are always indifferent between consuming and saving at rate \( R \) (which is always possible by choosing a negative \( b_{i,t} \)), we can assume that firms postpone consumption
until the time they exit.

Surviving firms start the period with net worth $n_{i,t}^F$, they borrow funds from households and banks and invest in assets $k_{i,t}$, implying the following law of motion for net worth

$$n_{i,t+1}^F = R_{i,t+1}^k k_{i,t} - Rb_{i,t} - R_{i,t+1}^L l_{i,t}. \tag{13}$$

Therefore, surviving firms maximize their expected wealth upon exiting, given the law of motion of net worth, borrowing constraints, and taking as given the loan rate $R_{i,t+1}^L$. Thus, at time $t$, a firm with net worth $n_{i,t}^F$ maximizes

$$\sum_{s=t+1}^{\infty} (1 - \tau_s)^{s-t-1} \tau_s (1/R)^{s-t} E_t n_{i,s}^F$$

subject to (8), (9), (11) and (13).

3.3 Banks

There are a continuum of agents who run financial firms, I refer to these agents as banks. Banks are special because, unlike households, they accept risky loans from firms.\footnote{Appendix B shows that banks are willing to accept contracts with risky payoffs because, unlike households, they have the ability to monitor firms.} By lending to multiple firms, banks pool idiosyncratic risks and use their diversified portfolio as collateral to borrow additional funds from households.

Banks are indexed $j \in \mathbb{J} = (1, 2]$. Bank $j$ starts period $t$ with net worth $n_{j,t}^B$, borrows from households by issuing riskless debt $d_{j,t}$ and lends to firms. I denote $l_{j,t}$ as the total amount of funds lent by bank $j$, thus $l_{j,t} = \int_{i \in I} l_{i,j,t} di$ where $l_{i,j,t}$ is the specific loan from bank $j$ to firm $i$. The balance sheet of the bank is described by the budget constraint

$$l_{j,t} = n_{j,t}^B + d_{j,t}. \tag{14}$$

I denote with $\tilde{R}_{t+1}^L$ the total return on banks’ assets (return on their portfolio of loans), then

$$\tilde{R}_{t+1}^L = \frac{\int_{i \in I} R_{i,t+1}^L l_{i,j,t} di}{l_{j,t}}. \tag{15}$$

Banks’ debt must be riskless, i.e. debt must be secured by the lowest realization of return on assets, which implies the borrowing constraint
\[ R_{d,j,t} \leq \tilde{R}_{t+1}^L l_{j,t}, \]  
(16)

where \( \tilde{R}_{t+1}^L \) denotes the lowest possible return on banks’ assets \( \tilde{R}_{t+1}^L \). By diversifying risks, banks increase the lowest return \( \tilde{R}_{t+1}^L \) and relax their constraints. Since banks have zero cost to pool over different firms, in equilibrium they perfectly diversify idiosyncratic risks, thus

\[ \tilde{R}_{t+1}^L = \int_{i \in I} R_{i,t+1}^L di = Z_{t+1} R_{t+1}^L \]  
(17)

and the lowest return on the portfolio is

\[ \tilde{R}_{t+1}^L = Z_t R_{t+1}^L. \]  
(18)

While banks diversify away the idiosyncratic risk from their portfolio, they hold the aggregate risk in their balance sheets: the loan rate \( R_{t+1}^L \) is known at \( t \), but the fraction of firms that default on their loans \( Z_{t+1} \) is stochastic. Thus, the presence of aggregate risk in the economy limits banks’ leverage. The issuance of safe debt requires loss-absorption capacity against non-diversifiable risk which is provided by banks’ net worth. Therefore, the size of banks’ balance sheets depends on their net worth: from (14) and (16) we get

\[ l_{j,t} \leq \frac{1}{1 - \frac{Z_t R_{t+1}^L n_{j,t}^B}{R}}. \]  
(19)

The denominator in (19) represents the minimum funding that needs to be provided by a bank for each unit of assets. This down-payment is directly related to the return on the bank’s portfolio of loans, which determines the collateral value of its assets. Importantly, we can see from (19) that the maximum size of the bank’s portfolio depends on its net worth \( n_{j,t}^B \) and the loan rate \( R_{t+1}^L \). Both are endogenous in the model.

**Net worth dynamics and the bank problem.** Like firms, banks discount time at rate \( 1/R \), exit with probability \( \tau_x \) and maximize the expected value of their wealth upon exiting. The net worth of a surviving bank \( j \), who provides loans \( l_{j,t} \) and issues deposits \( d_{j,t} \), evolves as

\[ n_{j,t+1}^B = \tilde{R}_{t+1}^L l_{j,t} - R_{d,j,t} = Z_{t+1} R_{t+1}^L l_{j,t} - R_{d,j,t}. \]  
(20)
A bank that starts period $t$ with net worth $n_{j,t}^B$ maximizes

$$\sum_{s=t+1}^{\infty} (1 - \tau_x)^{t-1} \tau_x (1/R)^{s-t} \mathbb{E}_t n_{j,s}^B,$$

subject to (14), (16) and (20).

**Intermediation: Diversification and collateral provision.** Bank intermediation is key in the provision of riskless assets that can be used as collateral to help allocate resources to the most productive sector. Firms are only able to borrow from households against the part of their assets that is not affected by idiosyncratic shocks, that is the fraction $(1 - \kappa)$. While firms cannot use the remaining risky part of the assets to borrow from households, banks can pool those risky payoffs, backing up their loans, and borrow from households against the fraction $Z_L/Z$ of the expected value of their loan portfolio. In that sense, banks’ ability to share and pool firms’ risk allows banks to provide *new safe collateral.*

Through this intermediation process, by which banks provide loans and create collateral, additional funds flow from households to firms—*funds that could not flow directly.* Importantly, the size of intermediation—the amount of funds banks intermediate from households—depends on the size of the loan portfolio that banks can manage and use to create collateral. On the asset side, banks can intermediate only over the loans firms demand, which is constrained by firms’ net worth. On the liability side, because banks hold aggregate risk, the size of their portfolio is limited by banks’ net worth. Therefore, in equilibrium, the size of intermediation fluctuates depending on the endogenous evolutions of the net worth of banks and firms.

### 3.4 Aggregation

I denote with uppercase letters aggregate variables.

As firms and banks exit over time with probability $\tau_x$, I assume that every period a new set of firms and banks enter and replace them so that the set of firms and banks remains $I = [0,1]$ and $J = (1,2]$. Every period, a measure $\tau_x$ of banks and firms enter the market. Each new firm (bank) is endowed with random initial net worth with mean equal to the fraction $\xi$ of the average net worth of existing firms (banks). We have, by aggregating (13) and (20) and considering the
entry and exit of firms, that

\[ N^F_t = (1 - \tau_x + \xi \tau_x) \left[ R^K_t K_{t-1} - R B_{t-1} - Z_t R^L_t L_{t-1} \right], \tag{21} \]

\[ N^B_t = (1 - \tau_x + \xi \tau_x) \left[ Z_t R^L_t L_{t-1} - R D_{t-1} \right], \tag{22} \]

where \( R^K_t \) denotes the aggregate return on capital

\[ R^K_t = (1 - \kappa) + Z_t \kappa + Z_t \alpha K^\alpha_{t-1}. \tag{23} \]

I define the parameter \( \tau \equiv \tau_x - \xi \tau_x \) since only the aggregate net dividend payout of banks and firms matter for equilibrium dynamics. For simplicity, the model assumes this payout policy to be a constant ratio of net worth.

**Assumption 1.** \( R(1 - \tau) < 1 \)

Assumption 1 ensures that firms and banks do not accumulate net worth fast enough to get away the constraints in steady state. Throughout the paper, I maintain this assumption.

### 3.5 Equilibrium

Given initial values of firm and bank net worth \( \{ n^F_{i,0}, n^B_{j,0}; i \in I, j \in J \} \), an equilibrium is described by a set of stochastic processes adapted to the filtration generated by \( Z \): aggregate capital and loans \( \{ K_t, L_t \} \), aggregate net worth of firms and banks \( \{ N^F_t, N^B_t \} \) and loan rates \( \{ R^L_{t+1} \} \), and a set of stochastic processes for each firm \( i \in I \) and bank \( j \in J \) (each adapted to their information): \( \{ n^F_{i,t}, k_{i,t}, b_{i,t}, l_{i,t} \} \) and \( \{ n^B_{j,t}, l_{j,t}, d_{j,t} \} \), such that:

1. Taking prices as given, allocations solve the optimization problems of firms and banks.
2. The market for bank lending clears

\[ \int_{i \in I} l_{i,t} di = \int_{j \in J} l_{j,t} dj = L_t. \tag{24} \]

3. Aggregate capital and output are given by

\[ K_t = \int_{i \in I} k_{i,t} di, \tag{25} \]

\[ Y_t = \int_{i \in I} y_{i,t} di = Z_t K^\alpha_t. \tag{26} \]

\( \text{25} \)The filtration generated by \( Z \) and \( z_i \) in the case of firms.
4. Aggregate net worth of firms and banks follow (21) and (22), respectively.

The deep-pocket feature of households with exogenous discount rate implies the demand for riskless debt or deposits is residual: it is perfectly elastic, and adjusts to satisfy the aggregate supply from firms and banks that earns the rate $R$. Aggregate consumption is also residual: linear utility and the access to the investment with return $R$ implies that households are indifferent between current and future consumption. Thus, this consumption decision is irrelevant.$^{26}$

4 Solving the equilibrium

4.1 Benchmark without borrowing constraints

Without borrowing constraints balance sheets do not play any role. Firm owners can finance their capital investment by issuing equity to households. Thus, firms, banks and households frictionlessly share risks, and linear utilities imply that any spread is arbitrated away, thus in equilibrium $E_t R_{t+1}^K = Z R_{t+1}^L = R$. Decreasing returns in aggregate capital implies there is an efficient allocation $K_t^{FB}$ which, using (23), solves

$$E_t R_{t+1}^K = Z \alpha (K_t^{FB})^{\alpha - 1} + (Z \kappa + (1 - \kappa)) = R. \quad (27)$$

Thus,

$$K_t^{FB} = \left[ \frac{\alpha Z}{R - (Z \kappa + (1 - \kappa))} \right]^{\frac{1}{1-\alpha}}.$$

Since aggregate shocks $Z_t$ are i.i.d. the economy instantaneously arrives to its efficient level of capital and remains there at all periods. Firm and bank net worth grow at expected rate $R(1 - \tau)$, so from assumption 1 their aggregate net worth would tend to zero.

4.2 Linear policies, aggregation, and the state space

In a competitive equilibrium, firms and banks take $R_{t+1}^K, R_{t+1}^L$ as given, and face dynamic problems with linear constraints in net worth and constant returns on their investments, which implies optimal policies are linear in their individual net worth. This allows aggregation across firms

$^{26}$Since my focus is on firms and banks balance sheets, I do not include the objects related to the household in the definition of equilibrium nor derive them in next sections.
and across banks and a simplification of the state space: we need to keep track of the aggregate net worth, separately, of firms and banks, but the distribution across each kind of agent is not important. As a consequence, the state space is summarized by the aggregate net worths of firms and banks $s_t = (N^F_t, N^B_t)$. Equivalently, I express the state space with the total net worth in both sectors $N_t = N^F_t + N^B_t$ and the share of banks' net worth $\beta_t = N^B_t / N_t$.

**Discussion.** The literature has mainly focused on models where the aggregate net worth of constrained borrowers, $N$, is the relevant variable which determines the allocation of capital and economic dynamics, while the distribution of net worth between different borrowers does not play any role. In the models of Bernanke et al. [1999], Jermann and Quadrini [2012] or Liu et al. [2013], constrained borrowers are entrepreneurs/firms and their net worth is the relevant variable; intermediaries/banks are merely a veil or not present at all. Thus, these models are represented by the particular case in which $N^B = 0$ or $\beta = 0$. Instead, in the models of He and Krishnamurthy [2013] or Begenau [2019] the borrowers are financial experts, intermediaries or banks, which directly manage the productive assets of the economy. Their net worth is the relevant state variable and models are calibrated to match net worth or leverage data of the financial sector: firms are not present in these cases, thus they correspond to the case when $N^F = 0$ or $\beta = 1$. Finally, Gertler and Kiyotaki [2010] or Brunnermeier and Sannikov [2012] model explicitly both a financial sector and a productive sector, but assume a frictionless relationship between them, which implies that the relevant variable is the total net worth $N = N^F + N^B$, while $\beta$ does not play any role.

In my model, frictions between banks and firms imply that they might have different borrowing and investment strategies and the conditions of their balance sheets may play different roles on investment dynamics. Consequently, the dynamics of the economy additionally depend on the net worth distribution $\beta$.

**Division of state space.** As depicted in Figure 3, the model is characterized by three different constraints: i) firms are constrained in their borrowing from households, ii) firms are constrained in their borrowing from banks, and iii) banks are constrained in their borrowing from households. The state space can be divided into different regions according to which constraints are binding and which are not. In each of these regions the net worth of banks and firms play different roles and each region features different economic dynamics.

---

27 In my model, this also requires that there is no friction between banks and firms, so that banks can continue lending even if firms’ net worth is zero. This would happen if firms could use as collateral with banks all the risky part of capital ($\bar{\phi} = 1$) and their cash-flows ($\phi_y = 1$, where $\bar{\phi}_y$ is defined in Appendix E).
I assume banks and firms borrow from households up to the limits of their constraints. Hence, the state space is characterized by two different regimes, as depicted in Figure 4:

(i) The region where also firms borrow up to the limit from banks. I refer to it as the regime of aggregate net worth booms and recessions.

(ii) The region where firms borrow below the maximum from banks. I refer to it as the regime of banking crises.

This partition is endogenous: agents optimally decide whether to borrow up to the limit of their constraints or not. Thus, the boundaries depend on the agents policies and vice versa. First, I characterize firm optimal borrowing decisions which determine the boundary that divides the two regimes. Then, I proceed by describing the dynamics of the model in the first regime (aggregate net worth booms and recessions). Finally, I explain the dynamics in the second regime (banking crises).²⁸

²⁸ Appendix C describes firms’ and banks’ dynamic problems and characterizes when constraints bind.
4.3 Firms’ optimal borrowing decision

By assumption firms borrow up to the maximum from households, so (9) is binding and we can write the law of motion of net worth in (13) as:

\[ n_{i,t+1}^F = (R^K_{i,t+1} - (1 - \kappa)) k_{i,t} - R^L_{i,t+1} l_{i,t}, \]  

which, from (7) and (10), implies

\[ n_{i,t+1}^F = \begin{cases} (\kappa + \alpha K_{t-1}^{\alpha-1}) k_{i,t} - R^L_{i,t+1} l_{i,t} & \text{if } z_{i,t+1} = 1 \\ 0 & \text{if } z_{i,t+1} = 0 \end{cases}. \]  

We can see from (28) and (29) that the decision of firms to borrow from banks depends on the loan rate \( R^L_{t+1} \) and on the return on capital \( R^K_{i,t+1} \). As usual in models with constraints, even when firms have linear utilities, their value functions are concave in the aggregate net worth \( N^F_t \). However, from (10) and (29) we see that firms pass on to banks the risk of the assets financed by loans. Therefore, firms’ borrowing decision from banks is determined by the spread between the loan rate and the expected firm return on assets, and independently of their marginal value of wealth. Building on this intuition, the following proposition characterizes the optimal firm demand for loans. Appendix D formally shows this result.

**Lemma 1.** Let \( R^K_{t+1} = E_t R^K_{i,t+1} \) be the firms’ expected return on capital defined in (23). There exists a threshold \( \overline{R}^L_{t+1} \) such that firms’ loan demand is

\[ L_{i,t} = \begin{cases} \overline{L}_{i,t} & \text{if } R^L_{t+1} < \overline{R}^L_{t+1} \\ \in [0; \overline{L}_{i,t}] & \text{if } R^L_{t+1} = \overline{R}^L_{t+1} \\ 0 & \text{if } R^L_{t+1} > \overline{R}^L_{t+1} \end{cases}, \]  

where

\[ \overline{L}_{i,t} = \frac{\phi - \kappa}{R^L_{t+1}} \left( 1 - \frac{1}{\phi} \frac{\kappa}{R^L_{t+1}} \right) n_{i,t}^F \]  

represents the maximum loan amount allowed by constraint (11). The threshold \( \overline{R}^L_{t+1} = \frac{R^K_{t+1} - (1 - \kappa)}{\overline{\mu}(1 - \frac{\kappa}{R^L_{t+1}})} \) is a function of firms’ expected return on capital \( R^K_{t+1} \).

Lemma 1 characterizes firms’ optimal borrowing from banks. It states that there is an upper-
bound for the loan rate such that if the rate is lower that this bound then firms borrow up to
the maximum, if it is greater firms wouldn’t borrow at all, while if it is equal to the upper-bound
firms are indifferent between borrowing or not. This linear feature of the borrowing decision
arises due to the constant return on capital that each individual firm faces. Each individual firm
compares their expected return on capital $R_{K}t+1$ with the loan rate $R_{L}t+1$.

4.4 Market of bank loans

Aggregate demand of bank loans. From Lemma 1, we can see that the loan policy function
is linear in firm net worth $n_{F}$. In particular, if firms borrow up to the maximum from banks,
aggregating (31) over all firms, we get the aggregate demand for bank loans:

$$L_{t}^{\text{Demand}}(R_{L}t+1; N_{F}t) = \phi \frac{\kappa}{R_{L}t+1} \left( \frac{1}{1 - \frac{1 - \kappa}{R} - \phi \frac{\kappa}{R_{L}t+1}} \right) N_{F}t.$$ (32)

Aggregate supply of bank loans. Since we are assuming that banks borrow up to the
maximum from households, we have that (19) is binding. So, aggregating over all banks, we get the aggregate supply of bank loans:

$$L_{t}^{\text{Supply}}(R_{L}t+1; N_{B}t) = \frac{1}{1 - \frac{z_{L}R_{L}t+1}{R}} N_{B}t.$$ (33)

Market clearing. Figure 5 illustrates the equilibrium in the market of bank loans for the two
possible regimes described in figure 4. In panel 5a, the market clears at a rate lower than the
threshold $R_{L}t+1$ so firms borrow up to the maximum from banks. Instead, in panel 5b, the bank
net worth and the supply curve are significantly low. In this case, the market clears at the
maximum rate $R_{L}t+1$ and firms borrow below the limit of their constraints.

We can see from (32) that a decrease in firms’ net worth, tightens firms’ constraints, and
contracts the demand for loans, pushing downwards the loan rate. While, from (33), a decrease
in banks’ net worth, tightens banks’ constraints, and contracts the supply of loans, pushing
upwards the loan rate. We can see from figure (5) that, for given $R_{K}t+1$, the effects of a decrease
in the bank net worth share $\beta_{t}$ are different in each regime. In case (5a), a drop in $\beta_{t}$ generates
an increase in the equilibrium loan rate $R_{L}t+1$. While, in case (5b), a drop in $\beta_{t}$ cannot increase
$R_{L}t+1$ since it is already at the maximum firms are willing to accept. In this case, the implied
contraction of the supply would be matched by a stronger adjustment in the equilibrium quantity
of loans. While this analysis is done in partial equilibrium as it is taking as given the return on
capital $R^K_{t+1}$, these differentiated response in the lending market will be the driving force of the differentiated transmission of shocks and equilibrium dynamics in the two different regimes.

**Empirical evidence.** This increase in the demand elasticity of firms during times of very low bank net worth will be key for the appearance of the bank accelerator in my model. Estimation of demand elasticities is hard due to the lack of data and endogeneity due simultaneity of demand and supply functions. A recent paper by Shen and Zhang [2021], provides a first estimate of loan demand elasticities. They show that loan demand from banks was more elastic during the Great Recession relative to the post-recession periods.\(^{29}\)

### 4.5 The partition

As described in above, the equilibrium loan rate $R^L_{t+1}$ depends negatively on the share of bank net worth $\beta_t$. Then, the maximum loan rate $\bar{R}^L_{t+1}$ is associated with a threshold $\bar{\beta}_t$ such that, for $\beta_t$ below this threshold, if firms would borrow from banks up to the maximum then the market would clear at a loan rate higher than $\bar{R}^L_{t+1}$. Thus, when $\beta_t < \bar{\beta}_t$ the economy moves into the banking crises regime (see figure (5b)).

**Lemma 2.** There exists a function $\bar{\beta}(N)$ of total net worth such that for $\beta_t < \bar{\beta}(N_t)$ the economy enters into the banking crises regime. In this regime, the equilibrium loan rate is at its upper bound $R^L_{t+1} = \bar{R}^L_{t+1}$ and firms borrow from banks below the limit of the constraint.

\(^{29}\)They show the results for time-varying elasticities for loan demand from banks, still not for firms’ demand.
Lemma 2 defines the partition of the state space that defines the banking crises regime. Notice that the threshold $\beta(N_t)$ depends on the total net worth. This is because, as the next section shows, total net worth $N_t$ will determine the return on capital, which in turn determines the maximum loan rate $R_{t+1}^L$.

**Definition 1.** The region of the state space that is associated with banking crises is defined as the set $\{(N_t, \beta_t) : \beta_t < \beta(N_t)\}$.

Figure 4 illustrates the partition of the steady state. The function $\beta(N)$ is upward sloping as a decrease in $N$ increases the firms’ return on capital and so the maximum rate at which firms would be willing to borrow from banks.

## 5 Regimes

### 5.1 Aggregate net worth booms & recessions

First, I describe the regime in which $\beta_t > \beta(N_t)$, so the loan rate $R_{t+1}^L$ is below $\overline{R}_{t+1}^L$. In this regime, both firms and banks borrow up to the limit of their constraints (all constraints bind). It is common in the literature to focus in models’ local behavior around a steady state where constraints are always binding. This has the advantage that dynamics can be described by using approximation techniques. For small shocks the economy lives around such steady state and the approximation may be good. As described below, this regime features the standard financial accelerator which works through the aggregate net worth of firms and banks, thus I name it the regime of aggregate net worth booms and recessions.

Aggregating (8) we have that the aggregate capital allocation in the economy equals the aggregate amount of funding raised by firms from their three sources of financing: their aggregate net worth, borrowing from households and bank loans, that is

$$K_t = N_t^F + B_t + L_t.$$  \hspace{1cm} (34)

From the market clearing condition for bank loans (24) and banks’ budget constraint (14), we have $L_t = N_t^B + D_t$, so we can rewrite (34) as

$$K_t = (N_t^F + N_t^B) + (B_t + D_t).$$  \hspace{1cm} (35)

This equation captures the end financing flow from investors to firms’ productive projects in the economy. The first term in parenthesis accounts for the overall net worth of firms and banks,
and the second term includes the overall funds from households that flow into firms through the issuance of riskless assets: $B_t$ flows directly through firm bonds and $D_t$ flows indirectly through banks.

By assumption in this regime we have that all constraints are binding, thus aggregating (9) and (16) we get

$$B_t = \frac{(1 - \kappa)K_t}{R},$$

$$D_t = \frac{Z_L R_{t+1}^L L_t}{R}. \tag{37}$$

These expressions state that the amount of borrowing from households equals the safe pay-offs (safe collateral) firms and banks can promise, respectively, discounted at rate $R$.

Also, firms borrow up to the limit from banks, so aggregating firm constraint (11) leads to

$$R_{t+1}^L L_t = \phi \kappa K_t. \tag{38}$$

Substituting this expression into (37), we can rewrite the aggregate amount of bank debt as

$$D_t = \frac{Z_L \phi \kappa}{R} K_t. \tag{39}$$

Note, by comparing (37) and (39), that the amount of safe debt banks issue equals the fraction $Z_L$ of the risky payoffs banks obtain from their loans, which, in this regime, coincides with the part of the risky assets that firms at maximum can promise.

Combining (35), (36) and (39), we obtain the following result.

**Proposition 1.** Given state $(N_t, \beta_t)$, if firms borrow up to the maximum from both households and banks, and banks borrow up to the maximum from households, the equilibrium aggregate capital is

$$K_t = \overline{\lambda} \left(N_t^F + N_t^B\right) = \overline{\lambda} N_t, \tag{40}$$

with

$$\overline{\lambda} = \left(1 - \frac{1}{R} \left(\phi Z_L \kappa + (1 - \kappa)\right)\right)^{-1}. \tag{41}$$

The proposition states that aggregate capital investment is determined by total net worth $N_t$. That is, both firm net worth and bank net worth affect the aggregate capital investment but their effect is the same. In this regime, since both banks and firms borrow up to the maximum of their constraints, they both use all of their net worth to allocate as much funds as they can in firms’
projects. As a result, what matters for overall capital investment is the amount of funds together they can borrow from households. For instance, the parameter $\lambda$ in equation (41) represents the \textit{overall leverage} of banks and firms. The overall “downpayment”, $\lambda^{-1}$, that they require to finance a unit of capital corresponds to the difference between the unit investment cost and the overall amount that is finance through safe securities placed to households. The latter consists of the sum of safe debt funding raised directly by firms $(1 - \kappa)$ and that raised by banks $\phi Z_L K$. Banks, by diversifying risks, provide additional collateral per unit of capital but only from the part of assets that they are able to intermediate, which is limited by $\phi$, the fraction of risky assets that firms can promise.

From the laws of motion of firm and bank net worth (21) and (22), we get that the evolution of total net worth follows

$$N_{t+1} = (1 - \tau) \left( R^K_{t+1} K_t - R(B_t + D_t) \right). \tag{42}$$

Using (36), (39) and (40), we can rewrite (42) as

$$N_{t+1} = (1 - \tau) \left( (R^K_{t+1} - R) \lambda + R \right) N_t. \tag{43}$$

\textbf{Irrelevance of net worth distribution.} Equations (40) and (43) show the dynamic complementarities between capital and total net worth. We can see that the transmission mechanism stressed by Bernanke and Gertler [1989] or Kiyotaki and Moore [1997] works through the aggregate balance sheet of firms and banks. A fall in total net worth $N_t$ implies a reduction in capital $K_t$ (from (40)) and affects future net worth $N_{t+1}$ (from (43)). This mechanism propagates the aggregate shocks to $Z_{t+1}$ as follows: because both banks and firms are levered, a temporary negative shock to $Z_{t+1}$, and to the return on capital $R^K_{t+1}$ (see (23)), reduces total net worth $N_{t+1}$ and capital $K_{t+1}$. The effects persist because the fall in future capital makes future cash-flows, total net worth, and investment lower than it would otherwise be. This channel—which represents the financial accelerator—induces persistence through balance sheets. However, balance sheet cycles are driven by total net worth, $N_t$, while the distribution of net worth in each sector, $\beta_t$, is irrelevant.\footnote{In the same way, a dynamic amplification channel, as in Kiyotaki and Moore [1997], would appear if we include some type of asset illiquidity which would affect asset prices; for example, by introducing capital adjustment costs. Note that if the price of capital is determined by its aggregate demand $K_t$, as it is common in the literature, the amplification mechanism of the dynamic channel will also work only through the total net worth $N_t$, while $\beta_t$ will be irrelevant.}

Note that this irrelevance result does not imply that banks do not play a role. Banks are
essential for the creation of safe collateral and intermediation of funds. For instance, the multiplier $\lambda$ depends on the parameter $\phi$, which captures the friction between firms and banks. So, $\phi$ represents the monitoring ability of banks, and affects the amount of intermediation in the economy. A more financially developed economy (high $\phi$) will achieve higher investment than a less developed one (low $\phi$) for the same total net worth.

**Redistribution between borrowers: Bank lending vs firm balance sheet channel.** We can analyze now the effects of a redistribution of funds from banks to firms (or vice versa), that is: a drop in $\beta_t$. While the loss in bank net worth would trigger a negative bank lending channel, the gains in firms net worth would trigger a positive firm balance sheet channel. It turns out that, in this regime, the effects of these channels on aggregate investment and output cancel out.

The intuition is the following. On the one hand, the increase in firms' net worth slackens firms' constraints and lowers the (shadow) borrowing costs for firms. On the other hand, the drop in banks' net worth tightens banks' constraints and raises the (shadow) borrowing costs for banks. Through the increase in the equilibrium loan rate, banks pass on these costs to firms, rendering no effect on total borrowing or investment.

**Risky steady state.** The risky steady state is defined as the limit to which the economy converges if agents expect future risk but shocks are realized at their expected values, i.e. $Z_t = \bar{Z}, \forall t$. I solve for the risky steady state under the assumption that at this limit all constraints bind. At the risky steady state we have, from (43), that

$$(1 - \tau) \left( (R^K_{SS} - R) \bar{\lambda} + R \right) = 1$$

The steady state return on capital, which from (23) equal $R^K_{SS} = \bar{Z}_\alpha (K_{SS})^{\alpha - 1} + (\bar{Z}\kappa + (1 - \kappa))$, satisfies

$$R^K_{SS} = R + \frac{1}{\bar{\lambda}} \left( \frac{1 - R(1 - \tau)}{1 - \tau} \right).$$

Assumption 1 thus implies $K_{SS} < K^{FB}$, that is, firms and banks die fast enough and do not accumulate enough net worth to get away the constraints. For the calibrated parameters in Section 6, the risky steady state effectively belongs to this region in which all constraint bind.

**Representative borrower.** For small shocks the economy lives around the steady state, always inside this region in which all constraints bind. Therefore, a corollary of Proposition 1 follows: we can consolidate banks and firms in one sector, a representative borrower, which is constrained
in its borrowing from households. This borrower owns the total net worth \( N_t \) and requires a down-payment \( 1/\varLambda \) to purchase a unit of capital.

As discussed in section 4.2, the literature mainly has abstracted from one of these sectors or has considered frictionless relationships. Proposition 1 implies that if the focus is on the dynamics of aggregate output and capital such abstraction gives a good approximation of the economic dynamics even when there are frictions among firms and banks, as long as constraints are always binding, that is, as long as firms and banks can share the shocks to their net worth through the equilibrium loan rate. In addition, leverage constraints should be interpreted as a combined friction from both sectors and models should be calibrated to match the overall leverage of the financial and non-financial sectors.\(^{31}\)

### 5.2 Banking crises

The *banking crises* regime is defined as the region of the state space in which firms borrow from banks below the maximum amount. In this regime, \( \beta_t < \overline{\beta}(N_t) \), the equilibrium loan rate \( R_{t+1}^{L} = \overline{R}_t^{L} \), and firms’ constraint with banks is slack. That is, firms use a lower fraction from their risky project payoffs than what they could potentially use to secure bank loans. Let \( \phi_t \) denote the fraction of the risky payoffs, \( \kappa K_t \), firms use to secure their loan promises, then:

\[
\phi_t = \frac{\overline{R}_t^{L} L_t}{\kappa K_t}.
\]

The fraction \( \phi_t \) captures how close the firm is to the constraint’s limit. When firms demand the maximum allowed by the constraint, we have \( \phi_t = \overline{\phi} \), otherwise \( \phi_t < \overline{\phi} \).

Following similar steps to the ones in the previous section, that is combining equations (35), (36), (37) with the new expression (44), we obtain the following result

**Proposition 2.** *Given state \((\beta_t, N_t)\) with \( \beta_t < \overline{\beta}(N_t) \), in the regime of banking crises in which both firms and banks borrow from households up to the maximum of their constraints but firms borrow below the maximum from banks, the equilibrium capital investment is*

\[
K_t = \lambda_t N_t,
\]

*with*

\[
\lambda_t = \left(1 - \frac{1}{\overline{R}_t} \left( \phi_t Z_t \kappa + (1 - \kappa) \right) \right)^{-1},
\]

\(^{31}\)Gertler and Kiyotaki [2010] calibrate parameters to match an average leverage ratio across different sectors (financial and non-financial).
where $\phi_t = \hat{\phi}(\beta_t)$ is an increasing function of $\beta_t$ and $\hat{\phi}(\beta_t(N_t)) = \overline{\phi}$.

Proposition 2 implies that, in the banking crises regime, the dynamics of aggregate capital and output depend on the total net worth $N_t$ and on the the share of banks’ net worth $\beta_t$. In this case, the multiplier $\lambda_t$ depends on $\beta_t$ through the variable $\phi_t$. The intuition is the following. As in the previous regime, both firms and banks allocate their net worth into firms projects. However, in this case, the value of the portfolio of loans that banks hold on their balance sheets corresponds to the fraction $\phi_t < \overline{\phi}$ of the risky part of capital $\kappa$, which they use to create safe collateral. A smaller fraction of firm assets allocated to the portfolio of banks implies smaller benefits from diversification, a drop in safe collateral and in the overall amount borrowed from households per unit of total net worth.

In this regime, the net worth of banks is special: It has a stronger impact on investment and output than firms’ net worth. An increase in either banks’ or firms’ net worth affects the level of capital through its effect on total net worth (direct effect in (45) from an increase in $N_t$). However, in addition, an increase in banks’ net worth increases the share $\beta_t$, which implies a higher multiplier $\lambda_t$.

**Bank accelerator.** A new financial accelerator appears during banking crises. In this regime, shocks to bank net worth impact the intermediation process. A drop in bank net worth tightens banks’ constraints, raises (shadow) borrowing costs for banks and leads to a contraction in the supply of loans. This puts upward pressure on the loan rate, which is at the maximum firms are willing to accept, i.e. $R_{t+1}^L$. Contrary to the regime of aggregate net worth recessions, banks cannot fully pass the increase in (shadow) borrowing costs on to firms. Instead, during these times in which loan rates are significantly high, firms react by cutting their borrowing below the limit of the constraint. Banks have to adjust their balance sheets: the drop in their extended loans reduces the value of the portfolio they were using as collateral to borrow from households, which in turn tightens banks’ constraints further, raises banks (shadow) borrowing costs, and contracts the supply of loans even more, reinforcing and amplifying the effects of the initial shock. I refer to this feedback loop, which leads to a collapse of banks’ balance sheets and severely affects investment and output, as the bank accelerator. The bank accelerator is captured by the change in the multiplier $\lambda_t$. Since less collateral is created per unit of total net worth, the overall leverage of banks and firms falls.

---

$^{32}$In general equilibrium, loan rates will increase further with the drop in bank net worth since the limit $R_{t+1}^L$ increases with the implied lower level of capital.
Redistribution between borrowers: Bank lending vs firm balance sheet channel. The bank accelerator appears in addition to the standard financial accelerator of Bernanke and Gertler [1989]. This new mechanism appears even when total net worth remains constant. In this regime, a redistribution of funds from banks to firms, i.e. a drop in $\beta_t$ holding $N_t$ constant, worsens the recession. This may be counterintuitive as firms are the ultimate borrowers that are financially constrained. But, note that, in this regime, the bank net worth multiplier is larger than the firm multiplier because banks, by expanding the loans over which they diversify risk, can lever and intermediate more funds from households than what firms could directly borrow. In this sense, the effect of the bank lending channel is stronger than the one triggered by the firm balance sheet channel.

Dynamics. The laws of motion of total net worth and bank net worth are described by

$$ N_{t+1} = (1 - \tau) \left( (R_{t+1}^K - R) \lambda_t + R \right) N_t, \quad (47) $$

$$ N_{t+1}^B = (1 - \tau) \left( Z_{t+1} R_{t+1}^L L_t - R D_t \right) = (1 - \tau) \left( Z_t - Z_L \right) \phi_t \kappa K_t, \quad (48) $$

where the last equality on (48) follows from (37) and (44). In addition to the effect of $\beta_t$ in aggregate capital $K_t$, the bank net worth share $\beta_t$ affects the laws of motion of net worth through its effect on $\phi_t = \hat{\phi}(\beta_t)$.

From (47) and (48), and replacing in (23), we can write $\beta_{t+1} = N_{t+1}^B / N_{t+1}$ as:

$$ \beta_{t+1} = \frac{(Z_{t+1} - Z_L) \hat{\phi}(\beta_t) \kappa}{Z_{t+1} (\alpha K_t^{\alpha - 1} + (1 - \hat{\phi}_t(\beta_t)) \kappa) + (Z_{t+1} - Z_L) \hat{\phi}_t(\beta_t) \kappa}, \quad (49) $$

Using that the total asset payoffs in the economy is $R_{t+1}^K K_t = ((1 - \kappa) + Z_{t+1} \kappa + Z_{t+1} \alpha K_t^{\alpha - 1}) K_t$ from which $((1 - \kappa) + Z_L \phi_t \kappa) K_t$ are distributed to households, expression (49) captures how the residual payoffs are distributed between firms and banks. The numerator captures the bank equity claim per unit of the aggregate capital, and the denominator adds the firm equity claim per unit aggregate capital.

How fast does the economy exit a banking crisis? The dynamic of the share of banks’ net worth $\beta_{t+1}$ is endogenous. In this regime, even when the loan rate $R_{t+1}^L$ is high, the recovery of bank net worth is slow because banks severely contract their balance sheets ($L_t$ and $\phi_t$ are low). Notice, from the second expression in (48), that $\beta_{t+1}$ depends positively on $\phi_t$ which, during banking crises, is low and an increasing function of $\beta_t$. This implies that any shock, even temporary, that affects the share of bank net worth and brings the economy to the banking crises regime has additional persistent effects through the bank accelerator: It takes time to exit this
Also, we can see from (49) that $\beta_{t+1}$ is increasing in $Z_{t+1}$, implying that the economy would move to the banking crises region when there is a low enough aggregate shock $Z$. While this shock affects firm assets directly, the indirect effect on banks balance sheet (through firms defaults) is stronger due to the higher leverage of banks. Notice that banks lever over an already leveraged asset (firm loans), thus banks have a higher exposure to aggregate risk.

Corollary 1. Bank accelerator. In the banking crises regime, a decrease in $\beta_t$, holding $N_t$ constant, generates a drop in aggregate capital $K_t$ and expected output $E_t Y_{t+1}$. In addition, banking crises are persistent: $E_t[\beta_{t+1}|\beta_t, N_t]$ is increasing in $\beta_t$.

Figure 6 illustrates the effects of the bank accelerator. The figure plots the equilibrium dynamics for two cases in which unexpected shocks affect bank and firm net worth at $t = 1$, and shows how the economy converges to the risky steady state: all aggregate shocks realize at their expected value, i.e. $Z_t = \bar{Z}, \forall t$. In the first case (green lines), total net worth drops by 10% from its steady state, while the net worth share stays at its steady state, i.e. $N_{t=1} = 0.9N_{SS}, \beta_{t=1} = \beta_{SS}$. In this case, the economy stays in the aggregate net worth recessions region, so the dynamics feature the standard accelerator and the drop in bank net worth leads to a drop in capital and output. In the second case, in addition, the share of bank net worth drops 70% bringing the economy to the banking crises region: $N_{t=1} = 0.9N_{SS}, \beta_{t=1} = 0.3\beta_{SS}$. In this case, the bank accelerator is triggered in addition to the standard accelerator, amplifying the drop in output.

The appearance of the bank accelerator during banking crises implies that the effects of the bank lending channel are stronger than the ones triggered by the firm balance sheet channel. It implies that recessions in which bank net worth is significantly low are more severe and persistent. Moreover, it implies that a government with available resources and the objective of increasing expected output would find more effective to use policies directed to recapitalize banks than policies directed to non-financial firms.

Empirical evidence. The bank accelerator helps in explaining the findings from the empirical literature on banking crises. This literature generally concludes that recessions accompanied by financial crises are more severe and persistent than other recessions (see Bordo et al. [2001], Cerra and Saxena [2008], Reinhart and Rogoff [2009], or Jordà et al. [2013]). Romer and Romer [2015] and Krishnamurthy and Muir [2016] highlight the importance of distinguishing banking crises by their severity. They use a continuous measure of financial distress, the former uses a narrative approach based on a reading of OECD accounts and the latter uses credit spreads, to capture the
Figure 6: Impulse response to exogenous shocks to $N_t$ and $\beta_t$

(a) Net worth $N_t/N_{SS}$

(b) Share of bank net worth $\beta_t$

(c) Output $Y_t/Y_{SS}$

(d) Fraction $\phi_t$

Note: The figure shows the convergence to the risky steady state for two cases in which unexpected shocks affect bank and firm net worth at $t = 1$, there are no other shocks: all aggregate shocks realize at their expected value, i.e. $Z_t = \bar{Z}, \forall t$. In the first case, an unexpected shock decreases total net worth $N_{t=1}$ by 10% below its steady state value $N_{SS}$, but the shock doesn’t affect the bank share of net worth $\beta_t$. In the second case, in addition, the share of bank net worth drops 70% from its steady state. The simulation uses the parameters from Section 6.

rise in the cost of credit intermediation, and find that crises associated with more severe measures of financial distress are associated with slower recoveries. In a recent paper, Baron et al. [2020] construct a new dataset on banks’ and nonfinancial firms’ equity returns for 46 countries over the period 1870-2016. They show that large bank equity declines are associated with substantial credit contractions and output gaps, even after controlling for changes in nonfinancial firms equity.\footnote{Models of a single constrained sector would have troubles explaining this fact.} They also show that the bank lending channel plays a key role and that its effects are highly non-linear, while this is not the case for the nonfinancial firms equity. Large declines in bank equity are followed by sharp credit contraction, but smaller declines and increases in bank equity are followed by muted changes in bank credit. There is also a striking contrast between

\footnote{Models of a single constrained sector would have troubles explaining this fact.}
bank equity and nonfinancial equity shocks, with the latter having a weak predictive power on bank credit growth.

All of these findings can be explained by the bank accelerator. My model generates banking crises and other financial recessions, where the latter would capture times of nonfinancial equity declines. The endogenous bank accelerator explains why banking crises are more severe and persistent, and also associates the severity of the downturn with the decline in bank equity and the resulting cost of credit intermediation. My model captures precisely the evidence in Baron et al. [2020] that show that firm and bank net worth have both predictive power for real output (even controlling for real variables as lagged output growth), which indicates that some type of financial accelerator is present in all types of recessions (and booms)—as in the model’s regime of aggregate net worth booms and recessions—while bank equity has associated a stronger nonlinear effect in recessions in which it drops significantly—as in the model’s regime of banking crises.

Baron et al. [2020] also highlight that large bank equity declines can uncover a number of historical banking crises not accompanied by banking panics, which are missed by narrative approaches. They estimate the model by choosing ad-hoc thresholds for the drop in bank equity that define banking crisis. In the next section, with a similar spirit, I identify when the economy enter into the regime of banking crises from the lens of my model. I then quantify the importance of this new bank accelerator in explaining real output, bank credit, and spreads.

6 Quantitative Exercise

In this section, I exploit the observed non-linear relationship between aggregate bank net worth and output to estimate the underlying parameters using the structure of the model. I then quantify the importance of the bank accelerator highlighted in the previous section. I provide answers to the following questions: i) in which periods was the bank accelerator activated (the US economy entered into the banking crises regime)? ii) how much of the variation in output and bank net worth is explained by this endogenous mechanism?, and iii) can this new mechanism explain the behavior in the bank lending market?

In order to answer these questions, I estimate the model to fit the time series variation of output ($Y$) and banks’ net worth ($N^B$). I measure output using quarterly U.S. real GDP and banks’ net worth as the aggregate market value of equity of U.S. commercial banks (from the U.S. Financial Accounts). The sample covers the period from 1985:Q1 to 2015:Q4.
6.1 Extended model with exogenous shocks

I introduce different shocks that can move the economy into the different regimes. For tractability, the estimation of the model allows for only two different shocks. First, a productivity shock $z^Y_t$ affects the return on assets of firms, which in turn affects bank loan repayments. Thus, the productivity shock affects both firm and bank net worth. In addition, I include an exogenous shock to bank net worth $z^B_t$. The purpose of $z^B_t$ is to capture other bank business such as residential mortgage lending, which is absent in the model. For instance, the Great Recession was triggered by the housing crisis. At the onset of the crisis a big share of banks’ assets were invested in securities related to the housing market, which severely affected banks’ balance sheets. My model abstracts from mortgages and the housing sector, so I capture such variations in the data with $z^B_t$. While the bank shock does not directly affect the assets of the firm, it can indirectly propagate by its effects on the market of bank lending.

Both exogenous shocks, $z^Y_t$ and $z^B_t$, follow AR(1) processes:

$$
\begin{align*}
    z^Y_t &= \rho_Y z^Y_{t-1} + \sigma_Y \epsilon^Y_t, \\
    z^B_t &= \rho_B z^B_{t-1} + \sigma_B \epsilon^B_t,
\end{align*}
$$

where $\epsilon^Y_t$ and $\epsilon^B_t$ are non-correlated and normally distributed innovations.

The estimation uses as observables $(Y, N^B)$, and the estimated model is fully described by the following system:

- aggregate output

$$
Y_t = Z_t K^\alpha_{t-1},
$$

(50)

with $Z_t = \max\{\exp(z^Y_t)Z, Z_L\}$ so that $Z$ and $Z_L$ capture the average and lowest possible aggregate shock, respectively;

- aggregate assets $K_t$ invested by entrepreneurs

$$
K_t = \lambda_t N_t;
$$

(51)

- the multiplier $\lambda_t$ that determines overall leverage of banks and firms follows

$$
\lambda_t = \begin{cases} 
\lambda & \text{if } N^B_t/N_t > \overline{\beta}(N_t) \\
\lambda(N^B_t/N_t) & \text{if } N^B_t/N_t \leq \overline{\beta}(N_t)
\end{cases},
$$

(52)
where \( \overline{\lambda} \) and the function \( \lambda(N_t^B/N_t) \) follow from Propositions 1 and 2;

- the laws of motion of total net worth and bank net worth

\[
N_{t+1} = (1 - \tau) \left( (R_{t+1}^K - R) \lambda_t + R \right) N_t, \tag{53}
\]
\[
N_{t+1}^B = (1 - \tau) (\phi_t Z_t \kappa K_t - R D_t) \exp \left( z_{t+1}^B \right), \tag{54}
\]

where (54) adds to (48) the exogenous shock \( z^B \);

- the return on capital

\[
R_{t+1}^K = (1 - \kappa) + Z_{t+1} \left( \kappa + \alpha K_t^{\alpha-1} \right). \tag{55}
\]

The system can be represented by

\[
\text{obs}_t = f(s_t, Z_t; \theta)
\]
\[
\begin{pmatrix}
    s_t \\
    Z_t
\end{pmatrix} = g(s_{t-1}, Z_{t-1}, \epsilon_t; \theta), \tag{56}
\]

where \( \text{obs}_t = (Y_t, N_t^B) \) is the vector of observables, \( s_t = (N_t, \beta_t) \) represents the vector of endogenous states and \( Z_t = (z_t^Y, z_t^B) \) the vector of exogenous states. The functions \( f \) and \( g \) represent the system of equations (50)-(55).

**6.2 Calibration and Estimation**

First, I set the parameters \((\tau, \kappa, \overline{\phi}, Z_L, \overline{Z}, R)\) such that for given \( \alpha \) the risky steady state matches the following targets: a firm leverage of 2, a share of bank net worth of 0.07, a ratio of output to total net worth of 0.9, an annual return on assets of 7\%, an annual spread on loans of 2\%, and an annual safe rate of 2\%. Second, I estimate the parameter \( \alpha \) and the parameters of the shock processes \((\rho^Y, \rho^B, \sigma^Y, \sigma^B)\) using a maximum likelihood approach. This is a non-linear dynamic system with exogenous states \((z_t^Y, z_t^B)\) and endogenous states \((N, \beta)\) treated as latent variables. I construct the likelihood following Fernández-Villaverde and Rubio-Ramírez [2007] that uses a particle filter as the updating procedure for latent variables (see details of the estimation procedure in Appendix F).

Table 1 reports the calibrated/estimated parameters. The estimated value for \( \alpha \) is 0.27. The parameter \( \alpha \) is the elasticity of output to the factor of production \( K \) in the model. But, recall that \( K \) does not represent physical capital, instead \( K \) represent the amount of funds needed in
advance for production. Thus, the value of $\alpha$ is different than the standard capital share—no data on capital or labor is used. Instead, the estimation treats $K$ as a latent variable which depends on net worth $N_t$ (see (51)). So, $\alpha$ is identified from the autocorrelation functions of output and bank net worth. But, note that $\alpha$ is not just identified from the simple correlation between output $Y_t$ and bank net worth $N_t^{B-1}$. Even in the absence of a direct (causal) effect of net worth on output, any other shock or mechanism that affect output would indirectly affect bank net worth and generate a positive correlation. I control for such other mechanisms or shocks through the presence of the autocorrelated factor $z^Y$. Up to a linear approximation, $\alpha$ is identified from the part of variation of output that cannot be explained by lagged values of output and that is correlated with bank net worth.\(^{34}\) In addition, the parameters lead to a non-targeted bank leverage of 10.7.

Table 1: Estimated/Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>technology parameter</td>
<td>0.27</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>fraction of capital hit by idiosyncratic shocks</td>
<td>0.63</td>
</tr>
<tr>
<td>$\phi$</td>
<td>share of firms’ capital promised to banks</td>
<td>0.42</td>
</tr>
<tr>
<td>$\tau$</td>
<td>net exit probability of banks and firms</td>
<td>0.035</td>
</tr>
<tr>
<td>$Z_L$</td>
<td>lowest possible default</td>
<td>0.87</td>
</tr>
<tr>
<td>$Z$</td>
<td>average default</td>
<td>0.99</td>
</tr>
<tr>
<td>$R$</td>
<td>safe rate</td>
<td>1.005</td>
</tr>
<tr>
<td>$\rho^Y$</td>
<td>autocorrelation, technology shock</td>
<td>0.59</td>
</tr>
<tr>
<td>$\rho^B$</td>
<td>autocorrelation, net worth shock</td>
<td>0.81</td>
</tr>
<tr>
<td>$\sigma^Y$</td>
<td>standard deviation, technology shock</td>
<td>0.008</td>
</tr>
<tr>
<td>$\sigma^B$</td>
<td>standard deviation, net worth shock</td>
<td>0.13</td>
</tr>
</tbody>
</table>

**Stochastic steady state.** Figure 7 shows the stochastic steady state for the calibrated parameter values. Lighter colors correspond to a higher density value, while darker colors correspond to a lower density. The green dashed line defines the threshold that divides the two regimes. We can see that the risky steady state is inside the regime of aggregate net worth booms and recessions, and for most shocks the economy lives inside this region. However, large enough negative shocks $z^Y_t, z^B_t$ can take the economy to the banking crises region. The following section presents the periods for which this was the case.

\(^{34}\)If the model were linear, the Kalman Filter representation would lead to a structural VAR that includes all possible lags of output and bank net worth. So, we could identify the elasticity from a predictive regression of output on lagged bank net worth controlling for lagged values of output. Appendix F.1 shows that such regression would provide a similar estimate of $\alpha$. 

40
Figure 7: Stochastic steady state

Note: Stochastic steady state for the calibrated parameter values. Lighter colors correspond to a higher density value, while darker colors correspond to a lower density. The green dashed line defines the threshold that divides the two regimes.

6.3 Results

6.3.1 Identifying banking crises

I start by presenting the periods where the U.S. entered into the regime of banking crises according to my model. Through the lens of my model this regime is identified whenever the bank’s net worth experiences a severe fall: $\beta_t < \bar{\beta}_t$. Figure 8 shows the estimated value of the latent share of bank net worth $\beta_t$, along with its data counterpart. The estimation procedure identifies the recessions of the early 90s and 2008 as banking crises, as opposed to the 2001 crisis where the recession did not come along with a severe drop in bank net worth. In addition, this methodology differentiates the severity of banking crises. The severity of the crisis is associated with the size of the drop in banks’ net worth. While the recession of the 90s was a banking crisis, it was not as severe as the 2008 crisis.

6.3.2 Importance of the bank accelerator

In order to assess the quantitative relevance of my mechanism, I compare the implications of the full model that features the bank accelerator to a counterfactual model in which the bank accelerator is turned off. The model without the bank accelerator is characterized by the equations described in section 6.1 but fixing $\phi_t = \bar{\phi}$ at all times, which implies a constant multiplier $\lambda_t = \bar{\lambda}$. Shutting off the mechanism can be interpreted as using the equilibrium equations in the around-
steady-state regime to describe the whole dynamics. References to the full model mean the dynamics are implied by the full system (56).

Each model implies a different path for the estimated latent variables (endogenous states and exogenous shocks) and delivers different predicted dynamics for each series. In the first exercise, I fix the path of all latent variables to the ones recovered from the full model and used them for both models (with and without the mechanism). In the subsequent exercises, I separately estimate the paths of latent variables for each case.

**Contribution of the bank accelerator.** To compute the counterfactual without the bank accelerator, I use the same latent variables and shocks \((s_t, Z_t)\) estimated from the full model, but I turn off the mechanism setting \(\lambda_t = \bar{\lambda}\). Figure 9 displays the series of output and bank net worth implied by the full dynamics and by the dynamics when the mechanism is turned off along with the data counterpart, for the periods after the recession of 2008. All series are shown as deviations from their linear trends. By construction, the full model matches the data perfectly, as we are feeding in the estimated shocks. Using these shocks, we set \(\lambda_t = \bar{\lambda}\) to find the counterfactual series in the absence of the bank accelerator. The differences between these series represents the variation in the data that is accounted by the endogenous mechanism. Table 2 reports the average size of the deviations of each series from its linear trend and the average difference as a percentage from the total observed drop. I conclude from this exercise that the
mechanism explains on average 40% of the fall in output and 78% of the fall in net worth during the Great Recession.

Figure 9: Contribution of Intermediation Mechanism

(a) Banks’ Net Worth

(b) Output

Note: Output and banks’ net worth implied by the full model (green line) and model without the mechanism (orange line), along with the data counterpart (black dots). All series are shown as deviations from their linear trends. The model without the mechanism sets \( \lambda_t = \lambda(\bar{\omega}) \), but uses the estimated latent variables (endogenous states and exogenous shocks) from the full system.

Table 2: Average Drop from Trend

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full model (data)</th>
<th>Without mechanism</th>
<th>Contribution (%Δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banks’ Net Worth</td>
<td>35%</td>
<td>8%</td>
<td>77%</td>
</tr>
<tr>
<td>Output</td>
<td>4%</td>
<td>2.4%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Note: Average deviation from linear trend for output and banks’ net worth implied by the full model (matches observed data) and model without the mechanism, from 2008:Q1 to 2015:Q1. The model without the mechanism sets \( \lambda_t = \lambda(\bar{\omega}) \), but uses the estimated latent variables (endogenous states and exogenous shocks) from the full system. The mechanism contribution (%Δ) represents the average difference between the drop implied by the full model (observed drop) and the model without the mechanism as a percentage of the observed drop.

Long horizon forecasting analysis. The following exercise assesses the prediction power of the bank accelerator for longer horizons. To do so, I compare the models’ forecasting performance for short and long horizons, with and without the mechanism, for the periods after the Great Recession.

Let \( \hat{obs}_{T_0} \) be the predicted value of the observables at time \( t \) conditioning on the information set up to time \( T_0 \) delivered by system (56), and let \( \hat{\epsilon}_{T_0}(t) = \log(\hat{obs}_{T_0}) - \log(obs_t) \) be the corresponding forecasting error. Table 3 reports the root mean square of the forecasting error \( RMSE(h) = \sqrt{\sum_{T_0} (\hat{\epsilon}_{T_0}(T_0 + h))^2} \) for different horizons. Sub-table 3a shows that the average forecasting error RMSE (4 quarter) for the full model is 17% and increases to 28% when estimating
the model without the mechanism. For longer horizons the full model delivers a RMSE of 22% and increases to 31% without the mechanism. The main message of this table is that the predictive power over banks’ net worth improves at short- and longer-term horizons when the bank accelerator is considered. Sub-table 3b shows the RMSE for output which is less volatile and presents smaller forecasting errors than banks’ net worth. Still, the full model delivers a lower RMSE for longer horizons.

As an example, figure 10 displays the actual values of the observables as a ratio of their linear trends (black dots) and their correspondent forecasts generated by both the full model (green line) and the model without the mechanism (orange line). In the figure, the model uses information up to the fourth quarter of 2008, and delivers forecasts for the next four quarters. The model with the mechanism is more capable of replicating both the level and the persistence of the fall that we see in the data.

The Great Recession featured a slow recovery compared to other recessions; both output and banks’ net worth dropped in 2008 and remained below their trends for many periods. The bank accelerator implies a reduction of the net worth multiplier $\lambda_t$ which leads to lower values of output per unit of net worth. Moreover, the mechanism implies a slow recovery: the severe drop in bank lending affects bank revenues, leading to slower recovery of bank net worth and banking crises to be persistent (as described in Corollary 1). Hence, the mechanism induces persistence and improves the forecasts of net worth, which in turn helps in explaining the slow recovery of output relative to its trend.

Table 3: Root Mean Square Error by horizons

(a) RMSE, Banks Net Worth

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>without mechanism</th>
<th>full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 quarter</td>
<td>0.28</td>
<td>0.17</td>
</tr>
<tr>
<td>8 quarter</td>
<td>0.31</td>
<td>0.22</td>
</tr>
</tbody>
</table>

(b) RMSE, Output

<table>
<thead>
<tr>
<th>Horizon</th>
<th>without mechanism</th>
<th>full model</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 quarter</td>
<td>0.025</td>
<td>0.019</td>
</tr>
<tr>
<td>8 quarter</td>
<td>0.029</td>
<td>0.027</td>
</tr>
</tbody>
</table>

*Note:* Root mean square of forecasting errors from 2008:Q1 to 2015:Q1 for different horizons ($RMSE(h)$). $RMSE(h) = \sqrt{\sum_{T_0} (\hat{e}_{T_0}(T_0 + h))^2}$ with $\hat{e}_{T_0}(t) = \log (\hat{obs}_{T_0/T_0}) - \log (obs_t)$, where $\hat{obs}_{T_0/T_0}$ is the predicted value of the observables at time $t$ conditioning on the information set up to time $T_0$ delivered by system (56). Model without mechanism (green line) sets $\lambda_t = \overline{\lambda}$ in system (56). Each model separately estimates the latent variables (endogenous states and exogenous shocks).
Figure 10: Forecast Analysis: Linear vs Non-linear dynamics

Note: Forecasts delivered by the system using information up to 2008:Q4 for the next four quarters, for banks’ net worth and output, respectively, along with the realized observations (black diamonds). All series are shown as deviations from their linear trend. Model without mechanism (green line) sets $\lambda_t = \bar{\lambda}$ in system (56). Each model separately estimates the latent variables (endogenous states and exogenous shocks).

6.3.3 Bank lending market

The model has also implications for the dynamics of the composition of credit and the cost of credit during different recessions. In this section, I study the model implied behavior of these variables for the recessions of the early 1990s, 2001, and 2008, and compare them to their aggregate data counterpart. Let me remark that the data analyzed in this section was not used in the estimation so it serves as external validation of the model.

As discussed in section 3.2 the amount of safe debt $b_{i,t}$ that is held by households or banks is undetermined in the model, and part of the firm payoffs that back-up the riskless debt $b_{i,t}$ could actually represent the recovery value of loans $l_{i,t}$. Thus, for this section, I assume that banks hold a mix of the loans $L_t$ and a fraction $\Delta$ of the aggregate safe debt $B_t$, so that total bank lending to firms equals $L_t + \Delta B_t$. I calibrate $\Delta = 0.22$ so that the share of aggregate bank loans on total firm debt $\frac{L_t + \Delta B_t}{L_t + B_t}$ matches 47%, which is the average value of the loan share data explained next.

Bank loans’ share. Figure 11 displays the dynamics of the share of bank credit relative to total credit implied by the model along with the data. All series are normalized to 1 at the beginning of each recession. The loan share data comes from the U.S. Financial Accounts and corresponds to the ratio of non-financial corporate sector loans relative to total non-financial corporate sector debt securities and loans.

Notice that, in the model, the difference between direct finance and intermediated finance regards the riskiness of the loan. In the model, bank loans are risky so they can only be provided
by banks which are willing to accept those risky contracts due to their monitoring ability, while
direct finance is assumed to be perfectly safe. In the data, of course, both types of firm financing
(bank loans and corporate bonds) have some risk associated. However, on aggregate, the vast
majority of corporate bonds are investment grade, while bank loans are given the highest risk
weight by regulators. So, I use these series as an approximation of data counterparts to the
model. The main hypothesis is that the supply of intermediated finance requires bank specialness
so it depends on bank net worth, while this is not the case for the supply of direct finance.

The data shows a considerable decline in the bank loans share during the crises of 1990 and
2008. For these particular periods the full model is able to produce a decrease in the loan share
of about 15% and 30% respectively, as opposed to small decreases when the mechanism is turned
off. In the model, the supply of credit is determined by banks’ ability to intermediate. During
banking crises, bank net worth falls and reduces banks’ capacity to bear aggregate risks. The bank
accelerator exacerbates this effect, as the reduction of bank loans implies less diversification and
collateral, leading to a further shrinkage of banks’ balance sheets and credit supply. Consequently,
the bank loan share remarkably drops during these type of recessions. Notice that the pattern
generated by the full model is followed in the data with a certain lag. This lagged behavior
might be explained by the difference in maturities on the different types of loans in the data.
Instead, the model considers only one-period contracts. For the recession of 2001 the nonlinear
mechanism of the full model is not active and hence both models (with and without mechanism)
display the same dynamics, in which the share is unaffected. This pattern is followed in the data.

The bank loan spread. Figure 12 plots the evolution of the cost of credit implied by the full
model and the observed evolution of two measures of credit spreads for the three recessions. In
the model, variation in the loan rate is driven by the shocks to bank and firm net worth, which
determine credit supply and demand conditions. Thus, in addition to presenting data of loan
spreads from the US syndicated loan market from LPC DealScan, I also show the excess bond
premium (EBP) constructed by Gilchrist and Zakrajšek [2011], which represents variation in
corporate spreads that is unrelated to variation in expected defaults. As suggested by Gilchrist
and Zakrajšek [2011], the EBP provides a timely and useful gauge of credit supply conditions.
My simple binomial model of bank loans is not intended to fit the lending data. Still, it does a
good job in explaining the patterns in spreads observed for the three recessions.

35 According to S&P Global Fixed Income Research more than 70% of rated corporate debt in the U.S. are
investment grade. 36 Another explanation might be that at the beginning of financial crises firms with pre-signed credit lines
borrow up to their limits as a precautionary behavior to the expected contraction in credit, this is stressed for
example in Ivashina and Scharfstein [2010].
Figure 11: Bank Loans Share during Recessions

Note: Share of bank loans in total firm debt implied by the model and data during the recessions of 1990, 2001, and 2008. The loan share data corresponds to the ratio of non-financial corporate business loans relative to total non-financial corporate sector debt securities and loans. Variables are scaled by their values at the beginning of each recession. Model without mechanism sets $\lambda_t = \overline{X}$ in system (56) and re-estimates all latent variables.

As illustrated in Figure 12b, a distinct feature of the recession of 2008 is that the economy experienced a remarkable increase in the cost of credit. The crisis of 2001 also came along with an increase in spreads but not as pronounced as the one of 2008. In contrast, the recession of 1990 did not present a substantial increase in spreads. Figure 12a shows the dynamics of the loan rate implied by the full model in each of the recessions. The model produces similar patterns to the ones observed in the data. Interestingly, the estimated model can produce such differential response in both identified banking crises: the large increase in the 2008 recession and the small change in the early 90s. The model-implied evolution of spreads is different in these banking crises because of two reasons. First, the severity of the recession is different, the drop in bank net worth is much larger in 2008. Second, in the model, what matters to activate the bank accelerator is the loan rate relative to the return on assets of firms. The estimation finds that the early 90s recession is associated with a larger productivity shock that reduced firm expected returns, whereas the Great Recession is mainly triggered by a specific shock to banks. Thus, in the early 90s, a small increase in the loan rate was sufficient to make firms contract their demand of bank loans below their limits, which triggered the mechanism that can simultaneously explain the drop in bank loans and the persistence in output and net worth described in the previous section.
Figure 12: Spread on Bank Loans during Recessions

(a) Model Dynamics: Loan Premium $R_{t+1}^L - R$

(b) Data: Credit Spreads

Note: Spread on bank loans implied by the model and data for the recessions of 1990, 2001, and 2008. The model implied series corresponds to the loan premium which is the difference between the loan rate and the risk-free rate $R_{t+1}^L - R$ estimated using the full model. The second panel shows two measures of credit spreads: (i) The loan spread is from DealScan database of loan originations, and corresponds to the weighted average of the cost of new loan issuances (in bps) extended for general corporate purposes and liquidity management. Cost is defined as all-in-drawn spread, which is total (interest plus recurring fees) spread paid over 6 month LIBOR. (ii) The EBP is the excess bond premium constructed by Gilchrist and Zakrajšek [2011].

7 Conclusion

This paper provides a dynamic theory of financial intermediation where the distribution of net worth across different constrained borrowers (banks and firms) plays a role in real activity. In particular, firms are constrained borrowers with access to real investment opportunities, while banks are constrained borrowers that use their funds to finance firms. Both bank and firm net worth matter for real activity and can have a differential impact on investment, output, and interest rates. In particular, during banking crises, the net worth of banks is special and has a stronger impact on investment and real activity.

The model assumes that banks are special because, due to their specific monitoring skills, they are willing to share risks with firms that other lenders (households) are not. By lending to multiple firms, banks pool idiosyncratic risk and use their diversified portfolios as collateral to borrow from households. Bank intermediation allows additional funds to flow from households to firms. The net worth of firms determines their financing capacity, both from households and banks. The net worth of banks determines their debt capacity and, in turn, the amount of funds intermediated to firms.

The model leads to different types of financial recessions. During aggregate net worth re-
cessions, shocks to the net worth of both banks and firms have the same effect on real activity. This happens even when there are frictions between banks and firms. The insight is that, in this regime, banks can pass on to firms any increase in borrowing costs, generated by a drop in their net worth, by increasing their loan rates. Instead, during banking crises, when the net worth of banks is critically low, loan rates are already so high that banks have a limited ability to pass on to firms the increase in their borrowing costs. Therefore, a drop in bank net worth implies a stronger reduction in lending volumes, which in turn impacts banking intermediation by contracting banks’ assets and the collateral provided to households. This further increases banks’ borrowing costs and amplifies the initial shock. This positive feedback, generated by this new bank accelerator, is a key innovation of the paper. The bank accelerator implies banking crises are longer and more severe. A transfer of resources to banks or firms has a positive impact on investment, but recapitalizing banks is more effective. An increase in bank net worth has an associated multiplier effect that allows more funds to be channeled to firms, because of their diversification ability, than what an increase in firms net worth would induce.

This new bank accelerator helps in explaining the observed non-linear relation between bank net worth and real economic activity. The estimation of the model finds that the bank accelerator is quantitatively important and improves the forecasting performance of the model during the recessions of the early 90s and 2008, and it can help explain the differential dynamics of the share of bank loans in total firm debt and credit spreads during the recessions of 1990, 2001, and 2008.

The paper main focus is on understanding the importance of the condition of firm and bank balance sheets for macroeconomic activity. With this in mind, the model simplifies as much as possible other sectors and frictions in the economy, and the quantitative analysis is based on such a model in which only financial friction forces (and exogenous shocks) drive booms and recessions with the intention to cleanly identify the mechanism contribution. The bank accelerator though could be embedded into a more standard DSGE framework in which one could evaluate how different frictions and shocks compete on their ability to explain business fluctuations. I leave these important issues for future work.
References


Jesús Fernández-Villaverde and Juan F Rubio-Ramírez. Estimating macroeconomic models: A likelihood


