

# Macroeconomic Effects of Secondary Market Trading\*

Daniel Neuhann<sup>†</sup>

November 18, 2016

## Abstract

This paper develops a theory of credit cycles to account for recent evidence that capital is increasingly allocated to inefficiently risky projects during credit booms. The model features lenders who sell risky assets to less informed investors in order to relax collateral constraints. When asset prices are high, however, these lenders begin producing and selling inefficiently risky assets. Asset prices rise during booms because the buyers of risky assets grow wealthy when their risk-taking pay offs, triggering a decline in investment efficiency and an increase in aggregate risk exposure. I study conditions that give rise to credit cycles and consider policy implications.

*JEL classification:* G01, E32, E44

*Keywords:* secondary markets, securitization, credit cycles, financial crisis, financial fragility, credit booms, saving gluts, risk-taking channel of monetary policy.

---

\*First version: October 15, 2015. This paper is based on Chapter 1 of my dissertation at the University of Pennsylvania. I am grateful for many helpful conversations with my advisors Harold Cole, Itay Goldstein, Dirk Krueger, and Guillermo Ordoñez. I thank Andres Almazan, Luigi Bocola, Murat Alp Celik, Selman Erol, John Geanakoplos, Urban Jermann, Michael J. Lee, Iourii Manovskii, Enrique Mendoza, Stephen Morris, Christian Opp, Francisco Silva, and seminar participants at Penn, Wharton, Penn State, Yale, Rochester, UT Austin, Johns Hopkins, NY Fed, Notre Dame, and the Fed Board for their feedback. I gratefully acknowledge financial support from the Robert Summers Dissertation Fellowship in Economics at the University of Pennsylvania. This paper has been prepared by the author under the Lamfalussy Fellowship Program sponsored by the ECB. Any views expressed are only those of the author and do not necessarily represent the views of the ECB or the Eurosystem.

<sup>†</sup>McCombs School of Business, University of Texas at Austin; daniel.neuhann@mcombs.utexas.edu.

# 1 Introduction

Credit booms are pervasive phenomena that frequently precipitate financial crises. A recent empirical literature suggests that an important mechanism linking booms to ensuing busts is that capital is increasingly allocated to inefficient and excessively risky investments during the boom phase.<sup>1</sup> To account for this mechanism, this paper presents a theory of financial intermediation in which the initial rise in credit volumes and the eventual collapse in investment efficiency both stem from the desire of financial intermediaries to sell off risk exposure in order to relax collateral constraints. In doing so, I provide an argument for why many prominent financial crises, such as the Great Depression, the Great Recession, and the 1997 Asian Financial Crisis, were preceded by sharp increases in the securitization and syndication of financial assets.<sup>2</sup> The theory also provides a characterization of the macroeconomic conditions that are most likely to give rise to credit booms and busts.

The model relies on three key ingredients. First, lenders to the real economy (“banks”) must provide collateral in order to borrow from risk-averse households (“savers”). Aggregate credit volumes are therefore partly determined by the wealth of banks. Second, banks must exert unobservable effort to originate high-quality, low-risk assets rather than low-quality, high-risk assets. Savers are thus concerned that banks may gamble on inefficiently risky projects at their expense, diluting the collateral capacity of bank wealth. Because this concern is particularly pressing when banks are already highly exposed to risk, banks’ collateral constraints introduce a motive to sell off risk exposure in order to borrow and lend more. To accommodate this motive, the third key ingredient is a market

---

<sup>1</sup> For example, Greenwood and Hanson (2013) show that the credit quality of corporate bond issuers falls during booms, while Piskorski, Seru, and Witkin (2015) document increased fraud in mortgage originations that resulted in unusually high default rates prior to the 2008 financial crisis. Similarly, the Report of the U.S. Financial Crisis Inquiry Commission (2011) show that sub-prime borrowers with high default rates accounted for a large fraction of increased household credit prior to 2008.

<sup>2</sup> Gorton and Metrick (2012), Brunnermeier (2009), Shin (2009), and the Report of the U.S. Financial Crisis Inquiry Commission (2011) survey the development of secondary markets and securitization in the United States prior to the 2008 crisis. While financial intermediaries issued less than \$100 billion in securitized assets in 1900, they issued more than \$3.5 trillion in 2006. Mian and Sufi (2009) and Ivashina and Sun (2011) provide evidence of a credit boom for households and firms during the same period. White (2009) and Kaminsky (2008) provide evidence of securitization and syndication booms for the Great Depression and Asian Financial Crisis, respectively.

for risk exposure in which banks sell assets exposed to aggregate risk to intermediaries who do not themselves provide funds to the real economy (“financiers”).

The fact that financiers do not themselves originate assets has a benefit and a cost. The benefit is that the collateral capacity of financier capital is not diluted by moral hazard in origination. As such, financiers can borrow more from savers than banks per unit of net worth, boosting the flow of funds into the financial system and raising credit volumes. The cost is that financiers are less informed about asset quality than the banks who originate these assets. For this reason, banks may sometimes find it in their interest to produce low-quality assets just to sell them. This leads to falling lending standards and deteriorating investment efficiency.

The extent to which asset sales boost credit volumes and hurt investment efficiency depends on the aggregate wealth distribution. On the one hand, asset sales sharply boost credit volumes without harming investment efficiency when intermediaries have little capital and collateral constraints are tight. This provides a strong positive rationale for asset sales as means of reallocating aggregate risk.

On the other hand, asset sales hurt investment efficiency without increasing credit volumes if financiers are very wealthy relative to banks. The mechanism operates through a feedback from asset prices to origination incentives. Since producing bad assets is inefficient, banks do so only when they expect to sell them to financiers at a sufficiently high price. This threshold price is strictly lower than the expected value of good assets because producing good assets is privately costly for banks. If all banks were to produce good assets, financiers would therefore earn strictly positive profits by buying at the threshold price. Yet because all banks would begin to shirk if the price were to increase, excess demand at the threshold price cannot be cleared through price adjustments. Instead, the only means of clearing the market at fixed prices is for a fraction of banks to shirk, since banks who shirk sell more assets than those who produce good assets. Excess financier wealth thus hampers investment efficiency by boosting demand for bank-originated assets.

Market-clearing via shirking is consistent with all agents making individually optimal decisions: the threshold price is chosen such that banks are indifferent between retaining good assets and selling bad assets, while financiers continue to earn profits as long as the fraction of shirking banks is not too large. Moreover, banks who produce good assets are willing to sell below par because collateral constraints introduce a shadow cost of holding risky assets. Collateral constraints thus provide the scope for asset sales to impact both credit volumes and investment efficiency. Yet private incentives are not aligned with social welfare, because financiers do not internalize that their asset purchases harm origination incentives. The welfare effects of this pecuniary externality can be severe, in that a partial destruction of financier wealth may generate a strict Pareto-improvement. Indeed, secondary market trading may be an important source of financial fragility because low-quality assets are disproportionately risky.

The adverse effects of asset sales thus stem from two imbalances: an excess of savings that tightens collateral constraints, and an excess of financier wealth that raises asset prices. The theory's dynamic implications, in turn, stem from the endogenous evolution of the wealth distribution. Because the role of financiers is to take on aggregate risk exposure, their wealth grows disproportionately during macroeconomic upturns. Initially, increased demand for risky assets boosts credit volumes by relaxing collateral constraints. Eventually, however, origination incentives deteriorate and investment efficiency falls. It is thus macroeconomic *upturns* that generate the wealth dynamics that lead to credit booms with falling asset quality. Because declining investment inefficiency generates excess risk exposure, booms also sow the seeds of an eventual bust. Financiers suffer disproportionately after bad shocks because they are particularly exposed to both aggregate risk and low-quality assets. Accordingly, recoveries from crises are slow because banks must retain all risk exposure when financiers are impaired, and longer booms precede sharper busts because investment efficiency falls gradually as wealth imbalances accumulate.

The law of the motion of the wealth distribution also determines the conditions that give rise to credit cycles in the first place. The two key factors are the initial wealth of

financiers and savers' supply of loanable funds. Financiers can buy only a small share of the stock of risky assets when they are poor, and so banks remain more exposed to risk and grow faster during upturns when this is the case. Yet financiers can make up for low initial wealth by exploiting their higher collateral capacity to take on more leverage than banks. This mechanism is particularly powerful when a large supply of savings pushes down interest rates. Credit booms with falling asset quality can therefore be triggered by saving gluts or capital inflows. Since the mechanism operates entirely through the risk-free interest rate, the model also suggests that monetary policy is an important determinant of the distribution of wealth across intermediaries. Notably, temporary shocks to the interest rate may be enough to set financiers on a path towards growth, because financiers can continue to buy a large share of risky assets at higher interest rates once they have accumulated wealth. This implies that a one-for-one reversal of the policy shock may not suffice to choke off a credit cycle once it is underway, and suggests a persistent and asymmetric risk-taking channel of monetary policy.

The model's narrative of credit booms is consistent with empirical evidence. For the pre-2008 credit boom in the U.S., Adrian and Shin (2010) estimate that the combined balance sheet size of non-bank intermediaries such as hedge funds and broker-dealers was smaller than that of bank holding companies before 1990 but almost twice as large by 2007. Greenlaw, Hatzius, Kashyap, and Shin (2008) show that non-bank intermediaries were more exposed to downside risk than loan-originating banks, while Coval, Jurek, and Stafford (2009) document that securitized assets were primarily exposed to systematic risk. Keys, Mukherjee, Seru, and Vig (2010) and Piskorski, Seru, and Witkin (2015) provide empirical evidence of falling credit standards and growing moral hazard over the course of the 2000-2007 U.S. credit boom. Griffin, Lowery, and Saretto (2014) show that, prior to 2008, issuers of complex securities produced and sold excessively risky products that under-perform during downturns. Importantly, these studies suggest that asset quality was inefficiently low, rather than just reflecting a lack of high-quality investment opportunities. Due to the endogenous upper bound on asset prices, the model is also consistent with the observation that asset prices did not reflect deteriorating in-

vestment efficiency prior to the crisis (Gennaioli, Shleifer, and Vishny (2012)). Schularick and Taylor (2012), Mendoza and Terrones (2012), and Reinhart and Rogoff (2009) document that longer credit booms predict sharper crises. Gorton and Metrick (2012) and Krishnamurthy, Nagel, and Orlov (2014) show that the fragility of leveraged secondary market traders was at the heart of the 2008 financial crisis. Caballero and Krishnamurthy (2009) argue that monetary policy was indeed expansionary during the early stages of the pre-2008 U.S. credit boom.

I derive two main policy implications in addition to the risk-taking channel of monetary policy. The first is that regulation which limits the accumulation of financier wealth or hampers financiers' ability to purchase excess amounts of loan-backed assets can be welfare-enhancing. Notably, this motive for regulation is independent of the financial structure of financiers. Indeed, it applies equally to zero-leverage financial institutions, such as asset managers, who have traditionally been outside the scope of financial regulation precisely because their lack of leverage was thought to eliminate financial fragility and agency frictions. The second is that leverage restrictions on banks may prematurely harm origination incentives because banks respond by selling fewer assets, raising asset prices.

*Related Literature.* In studying the impact of collateral constraints, my paper relates to Fostel and Geanakoplos (2015) who characterize leverage and collateral constraints in binomial economies with default. They show that the equilibrium collateral constraint is such that lending is risk-free. I take such a collateral-based borrowing constraint as given and study how the reallocation of aggregate risk exposure stretches scarce collateral and affects the efficiency of investment. Fostel and Geanakoplos (2008) study how collateral and leverage affect asset prices and generate spillovers across asset classes, while Garleanu and Pedersen (2011) do so using collateral-based margin constraints. Gromb and Vayanos (2002) study liquidity provision in segmented markets by collateral-constrained arbitrageurs. They find that arbitrageurs may be overexposed or underexposed to risky assets from a welfare perspective due to a pecuniary externality. The inefficiency here also

stems from a pecuniary externality, but it affects real investment rather than liquidity.

Kiyotaki and Moore (1997) is the seminal study of the macroeconomic effects of collateral constraints. More recently, Mendoza (2010) quantitatively studies collateral constraints and leverage over the business cycle, and shows that these constraints amplify the response to negative macroeconomic shocks. Lorenzoni (2008), Bianchi (2011), and Bianchi and Mendoza (2012) show that pecuniary externalities can trigger fire sales that amplify credit crunches, while Bigio (2014) and Kurlat (2013) study market shutdowns and adverse selection during downturns. Rather than focusing on the amplification of shocks in bad times, I study how the reallocation of risk harms credit quality and generates excessive risk-taking in good times.

Gorton and Ordoñez (2014) propose a dynamic model of credit booms and busts based on the desire of agents to trade information-insensitive assets. Booms and busts occur due to the evolution of beliefs. I emphasize the evolution of the wealth distribution and the deterioration of investment efficiency over the credit cycle. Gennaioli, Shleifer, and Vishny (2013) argue that securitization allows for improved sharing of idiosyncratic risk, and is efficient unless agents neglect aggregate risk. I study the re-allocation of aggregate risk, and show that excessive securitization can have deleterious effects even in a fully rational framework. Moreover, I explicitly model the dynamics of secondary markets and argue why booms can endogenously lead to financial fragility.

Parlour and Plantin (2008) and Vanasco (2014) study the effects of secondary market liquidity on moral hazard and information acquisition in primary markets in static partial equilibrium. I differ in that I study the macroeconomic dynamics of secondary markets and emphasize the credit cycle. Chari, Shourideh, and Zetlin-Jones (2014) show how secondary markets may collapse suddenly in the presence of adverse selection. I study how growing secondary markets can lead to falling asset quality. This concern is shared by Bolton, Santos, and Scheinkman (2016), who study how origination incentives vary with the demand for assets by informed and uninformed buyers.

The paper is structured as follows. I begin my analysis in Section 2, where I study

single-period model in which the wealth of all agents is fixed. I use this setting to show how risky asset sales can either relax collateral constraints or harm origination incentives, and argue that these two effects are shaped by the wealth distribution. In Section 3, I then embed the model in an overlapping generation setting to study the endogenous evolution of the wealth distribution. I study the model’s policy implications in Section 4, and conclude in Section 5. All proofs are in Appendix A.

## 2 Basic Model with Fixed Wealth

### 2.1 Setting

There is a single period divided into multiple stages to be described below. There are three types of agents, each forming a continuum of unit mass: *savers*, *banks*, and *financiers*. Types are indexed by subscripts  $S$ ,  $B$ , and  $F$ , respectively. Financiers and banks are risk neutral. I create a role for collateral by assuming that savers are infinitely risk averse as in Gennaioli, Shleifer, and Vishny (2013) and Caballero and Farhi (2014). As a result, all lending by savers to financiers and banks must be risk-free. The assumption thus yields the *zero-value-at-risk-collateral* constraint derived by Fostel and Geanakoplos (2015) and assumed in reduced form by e.g. Gromb and Vayanos (2002).

There is a single good that can be used for consumption and investment. Agents are born with endowment  $w_S$ ,  $w_B$  and  $w_F$  of this good, respectively. The only source of risk is an aggregate state  $z \in \{l, h\}$  whose realizations I refer to as the *low state* and the *high state*. The probability of state  $z$  is  $\pi_z \in (0, 1)$ . There are two investment technologies: the *risky technology* and the *storage technology*. The storage technology is available to all agents. It generates a certain rate of return  $R \in \{\underline{R}, \bar{R}\}$  per unit of capital invested. I assume that  $R = \underline{R} \leq \bar{R}$  for savers, while  $R = \bar{R} = 1$  for banks and financiers. While the majority of my analysis focuses on the benchmark case  $\underline{R} = \bar{R} = 1$ , I sometimes use  $\underline{R} < 1$  to explore the cross-sectional implications of increases in savers’ willingness to pay for financial services.

Markets are segmented in that the risky technology is available only to banks and requires bank monitoring to operate efficiently. The technology thus represents lending to households and firms that requires monitoring expertise or the acquisition of soft information. The technology generates a rate of return  $Y_z$  in state  $z$  if the bank monitors, and  $y_z$  if it shirks. Monitoring is unobservable and has a private utility cost  $m$  per unit of investment.

**Assumption 1 (Payoffs).**

*The payoffs of the risky technology satisfy  $\mathbb{E}_z Y_z > \mathbb{E}_z y_z + m$ ,  $y_l < Y_l$ , and  $\mathbb{E}_z Y_z > \bar{R}$ .*

The assumption states that (i) shirking is inefficient and induces additional downside risk, and (ii) the monitored risky technology delivers a higher return than safe the technology. Going forward, I use the shorthand  $\hat{Y} = \mathbb{E}_z Y_z$  and  $\hat{y} = \mathbb{E}_z y_z$ .

**Asset markets.** Markets are incomplete. There are two financial assets in zero net supply: a risk-free zero-coupon bond with face value one, and a risky asset described in the next section. Financiers and banks issue bonds to savers at price  $q$ , while banks use the risky asset to offload risk exposure to financiers. Risk-averse savers do not trade the risky asset.<sup>3</sup>

**The risky asset.** The risky assets represents infinitely divisible claims on the returns of the risky technology. I say that a risky asset is *good* if it represents a claim on monitored risky investment, and *bad* otherwise. Good assets thus yield a return of  $Y_z$  in state  $z$ , while bad assets yield a return of  $y_z$  in state  $z$ . In order to capture the notion that asset sales may harm origination incentives, I assume that information frictions prevent financiers from perfectly screening the quality of assets.<sup>4</sup> The assumption has two parts. First,

<sup>3</sup> There are two notable missing markets. First, banks do not issue bonds to financiers. This assumption is for expositional ease, and I show that it is immaterial to my results. Second, there is no equity market in which banks can raise funds by selling shares of *inside* equity to financiers. This segmentation of the equity market is common in the literature on financial intermediation with collateral constraints, and is consistent with the data. For example, Ivashina and Sun (2011) provide evidence that tranches of loans sold in secondary markets had lower yields than those held via direct claims on banks. Nevertheless, Section 3 shows that credit cycles with falling asset quality can arise even when banks are able to freely issue inside equity to financiers at no cost. Note also that risky asset sales are similar to *outside* equity, in that they allow banks to acquire risk-bearing capital.

<sup>4</sup> In the absence of such an assumption, no bad assets would ever be traded, and there would be no feedback to origination incentives, contradicting the empirical evidence in Piskorski, Seru, and Witkin (2015) and Griffin, Lowery, and Saretto (2014).

financiers cannot tell good assets apart from bad assets. Second, financiers cannot draw perfect inferences about asset quality by observing bank balance sheets, nor can they screen the quality of assets by offering sufficiently rich price-quantity menus. In practice, bank balance sheets are opaque and difficult to assess in real time. Moreover, financial institutions frequently trade with many other institutions simultaneously, so that asset markets are non-exclusive and buyers may not be able to efficiently screen by quantity (Attar, Mariotti, and Salanié (2011)).

In particular, and to maintain a role for asset sales as collateral in the presence of such information frictions, I assume that banks balance sheets are partially observable. That is, I assume that banks can commit to selling *at least*  $\underline{a}$  risky assets, but can deviate to selling  $a_B \in [\underline{a}, k]$  ex-post. While  $\underline{a}$  is observable,  $a_B$  is unobservable. This structure effectively assumes that banks are able to provide verifiable documentation that they have sold a given number of loans, allowing these asset sales to serve as collateral, while retaining the ability to engage in hidden trades that affect monitoring incentives.<sup>5</sup>

Because the choice of  $\underline{a}$  may impact banks' monitoring incentives, financiers will use  $\underline{a}$  as well as aggregate trade volumes to form inferences about the average quality of assets traded by  $\underline{a}$ -banks. I therefore assume that there are submarkets indexed by  $\underline{a}$  in which risky assets issued by  $\underline{a}$ -banks trade competitively at marginal price  $p(\underline{a})$ . I denote the fraction of low-quality assets trading on submarket  $\underline{a}$  by  $\phi(\underline{a})$ , and the fraction of  $\underline{a}$ -banks who shirk by  $\Phi(\underline{a})$ . All financiers who purchase assets in a submarket are allocated an equal share of low-quality assets. Risky assets purchased on submarket  $\underline{a}$  thus generate the return  $x_z(\underline{a}) = \phi(\underline{a})y_z + (1 - \phi(\underline{a}))Y_z$  in state  $z$ , with the expected return denoted by  $\hat{x}(\underline{a})$ . Submarket  $\underline{a}$  is *active* if financier's demand and bank's supply of risky assets at  $\underline{a}$  is strictly positive. Not all submarkets will be active in equilibrium, and I impose market-clearing conditions only in active submarkets. Nevertheless, it will be important

---

<sup>5</sup> Note that banks would never want to *buy back* risky loans in equilibrium because they would have to pay at least the expected value of the loan once they are owned by financiers. As a result, any loan sale documentation that banks provide serves as a commitment to not re-buy the associated risk exposure later on. An alternative interpretation is that there are multiple stages of bond trading in which agents observe loan sales at every stage, but banks cannot commit to not selling assets again in the future. The partial commitment provided by  $\underline{a}$  captures these concerns while maintaining tractability.

to specify off-equilibrium prices in inactive submarkets. Throughout, I restrict attention to equilibria in which the pricing function  $p : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is differentiable. As will become clear, this is a natural restriction in my setting.

**Remark 1.**

*In assuming that the risky asset represents a direct claim on the output of the underlying technology, I am implicitly ruling out further securitization and tranching of risky claims. I do so because securitization frequently serves to eliminate idiosyncratic risk, resulting in securities that are highly exposed to aggregate risk (Coval, Jurek, and Stafford (2009)). This is not necessary in my setting because there is only aggregate risk.*

**Timing.** Figure 1 summarizes the timing of events within the period. In stage 1, all agents receive their endowments. In stage 2, banks commit to selling at least  $\underline{a}$  units of the risky asset in stage 4, and risk-free bonds are traded.<sup>6</sup> In stage 3, banks invest in the risky technology using their own wealth and the proceeds from bond issuances in the funding market, and decide whether to monitor or shirk, and all agents invest in the safe technology. In stage 4, risky assets are traded subject to the constraint that banks must sell at least  $\underline{a}$ . In stage 5, the productivity shock  $z$  is realized, returns on investment accrue, accounts are settled, and all agents consume.

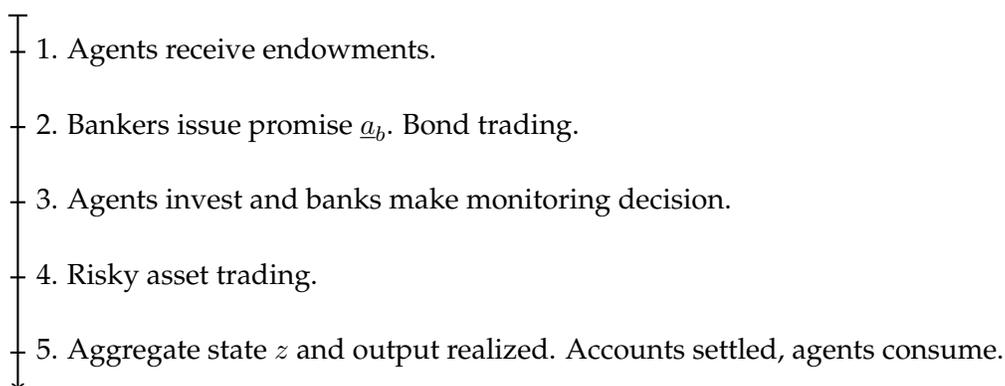


Figure 1: Timing of Events

---

<sup>6</sup> Specifying the partial commitment in this manner allows banks to rely on future asset sales as collateral in stage 1, while maintaining the scope for a deviation to excessive asset sales ex-post. It also allows me to straightforwardly embed the static model into a dynamic context in Section 3.

## 2.2 Decision Problems

I now characterize the supply and demand for bonds and risky assets by analyzing agents' decision problems. Given a bond price  $q$ , the gross risk-free rate is  $\frac{1}{q}$ . Financiers and banks issue bonds to borrow from savers. Banks sell risky assets to transfer risk exposure to financiers. All agents take prices as given.

### 2.2.1 Savers' Problem

Since savers are infinitely risk averse, they choose a portfolio consisting only of risk-free assets. Savers buy as many bonds as they can when  $\frac{1}{q} > \underline{R}$ , and use the storage technology otherwise:

$$b_S(w_S, q) = \begin{cases} \frac{w_S}{q} & \text{if } \frac{1}{q} > \underline{R} \\ [0, \frac{w_S}{q}] & \text{if } \frac{1}{q} = \underline{R} \\ 0 & \text{if } \frac{1}{q} < \underline{R}, \end{cases}$$

where  $b_S$  denotes savers' bond purchases. Observe that the bond price is bounded above by  $\frac{1}{\underline{R}}$ , and that bond demand is increasing in  $w_S$ .

### 2.2.2 Financiers' Problem

Financiers choose the amount of capital to invest in storage  $s_F$ , risky asset purchases  $a_F$ , and bond issuances  $b_F$ . Generically, exactly one active submarket will offer the highest return on risky assets. I denote this submarket by  $\underline{a}^*$  and solve a relaxed decision problem in which the financier purchases a non-negative quantity of assets in  $\underline{a}^*$  only, taking as given that financiers choose the optimal submarket. This decision problem is

$$\begin{aligned} \max_{s_F, a_F(\underline{a}^*), b_F} \quad & \mathbb{E}_z [s_F + x_z(\underline{a}^*)a_F(\underline{a}^*) - b_F] \\ \text{s.t.} \quad & s_F + p(\underline{a}^*)a_F(\underline{a}^*) \leq w_F + qb_F \end{aligned} \tag{1}$$

$$b_F \leq s_F + x_z(\underline{a}^*)a_F(\underline{a}^*) \text{ for all } z. \tag{2}$$

(1) is the budget constraint restricting asset purchases and storage to be weakly smaller than the sum of own wealth and bond issuances, while (2) is the collateral constraint that ensures that financiers are always able to fully honor their debts. The two pertinent choices are whether to buy risky assets, and whether to issue bonds to do so. A financier who does not issue any bonds can buy at most  $a_F = \frac{w_F}{p(\underline{a}^*)}$  risky assets, and earns an expected rate of return of  $r_a(\underline{a}^*) = \frac{\hat{x}(\underline{a}^*)}{p(\underline{a}^*)}$ . Financiers can increase their asset purchases by issuing bonds subject to (2). Conditional on not investing in storage ( $s_F = 0$ ), the collateral constraint can be rearranged to yield a maximum bond issuance of  $\bar{b}_F(\underline{a}^*) = \frac{x_l(\underline{a}^*)w_F}{p(\underline{a}^*) - qx_l(\underline{a}^*)}$ . If banks do not shirk, financiers can thus borrow against  $Y_l$  per unit of the risky asset. Issuing  $\bar{b}_F(\underline{a}^*)$  bonds allows the financier to purchase  $a_F = \frac{w_F}{p(\underline{a}^*) - qx_b(\underline{a}^*)}$  risky assets, generating an expected rate of return  $\tilde{r}_a(\underline{a}^*) = \frac{\hat{x}(\underline{a}^*) - x_l(\underline{a}^*)}{p(\underline{a}^*) - qx_l(\underline{a}^*)}$ .

Financiers are willing to invest in risky assets only if  $r_a^*(\underline{a}^*) = \max\{r_a(\underline{a}^*), \tilde{r}_a(\underline{a}^*)\} \geq 1$ , and strictly prefer to do so if  $r_a^*(\underline{a}^*) > 1$ . Moreover, it is straightforward to verify that financiers strictly prefer to invest in risky assets if  $r_a^*(\underline{a}^*) = 1$  but  $b_F > \bar{b}_F(\underline{a}^*)$  because buying the risky asset shifts risk exposure onto bondholders. The collateral constraint (2) is thus equivalent to  $b_F \leq \bar{b}_F(\underline{a}^*)$  no matter the financier's portfolio.

The leveraged return is higher than the unleveraged return if  $q\hat{x}(\underline{a}^*) > p(\underline{a}^*)$ . Financiers thus issue  $\bar{b}_F(\underline{a}^*)$  bonds if  $q\hat{x}(\underline{a}^*) > p(\underline{a}^*)$ , do not issue any bonds if  $q\hat{x}(\underline{a}^*) < p(\underline{a}^*)$ , and are indifferent when  $q\hat{x}(\underline{a}^*) = p(\underline{a}^*)$ . I summarize financier leverage by rewriting (2) as an equality constraint of the form  $b_F = \mu \cdot \bar{b}_F(\underline{a}^*)$ , where  $\mu \in [0, 1]$  denotes the degree to which financiers exhaust their borrowing capacity. Accordingly, the optimal portfolio is

$$a_F = \lambda_F(p, q)w_F \quad b_F = \mu(p, q)x_l(\underline{a}^*)\lambda_F(p, q)w_F \quad \mu(p, q) = \begin{cases} 1 & \text{if } q\hat{x}(\underline{a}^*) > p(\underline{a}^*) \\ [0, 1] & \text{if } q\hat{x}(\underline{a}^*) = p(\underline{a}^*) \\ 0 & \text{otherwise,} \end{cases}$$

where  $\lambda_F(p, q) = \frac{1}{p(\underline{a}^*) - \mu(p, q)qx_l(\underline{a}^*)}$  is financier leverage. The demand for risky asset is strictly increasing in  $w_F$ , strictly decreasing in  $p(\underline{a}^*)$ , and weakly increasing in  $q$ .

### 2.2.3 Banks' problem

I study the banks' problem under the presumption that banks earn intermediation rents  $(\hat{Y} > \frac{1}{q})$  and thus want to issue as many bonds as possible. Since banks may not exert monitoring effort in equilibrium, let  $e \in \{0, 1\}$  denote the banks' monitoring action, with  $e = 1$  if the bank exerts effort.<sup>7</sup> The private benefit associated with  $e$  is  $m^*(e) = (1 - e)m$ , and the associated return of the risky technology is  $Y_z^*(e) = eY_z + (1 - e)y_z$ . Bank utility in state  $z$  is

$$u_B(z, \underline{a}, a_B, k, s_B, b_B, e) \equiv \max \{s_B + Y_z^*(e)(k - a_B) - b_B + p(\underline{a})a_B, 0\} + m^*(e)k.$$

where  $s_B$ ,  $k$ ,  $a_B$  and  $b_B$  denote the bank's storage, risky investment, asset sales and bond issuance, respectively, and consumption is non-negative because bondholders cannot extract more than the bank's end-of-period assets. The bank's optimal monitoring choice conditional on  $(a_B, \underline{a}, k, s_B, b_B)$  is:

$$e^*(a_B, \underline{a}, k, s_B, b_B) = \arg \max_{e' \in \{0, 1\}} \mathbb{E}_z u_B(z, \underline{a}, a_B, k, s_B, b_B, e') \quad (3)$$

Asset sales are chosen once  $(\underline{a}, k, s_B, b_B)$  is determined. Hence,  $a_B^*$  must be optimal conditional on  $(\underline{a}, k, s_B, b_B)$  given  $e = e^*(a_B, \underline{a}, k, s_B, b_B)$ . That is,

$$a_B^*(\underline{a}, k, s_B, b_B) = \arg \max_{k \geq a_B \geq \underline{a}} \mathbb{E}_z [\max \{s_B + Y_z^*(e^*)(k - a_B) - b_B + p(\underline{a})a_B, 0\}] + m^*(e^*)k. \quad (4)$$

---

<sup>7</sup> In my baseline specification, I assume that the bank either monitors its entire portfolio (i.e.  $k$ ) or does not monitor at all. More generally, one might imagine that the bank is able to make monitoring decisions at the more granular level of individual assets, exposing financiers to the risk of banks *always* producing low-quality assets whenever they know they will sell them. In practice, markets for securitized assets have evolved to circumvent this strong form of adverse selection by allowing financiers to choose the individual assets they want to purchase from the bank's portfolio. By employing a random selection rule, financiers can then guarantee themselves the average quality of the portfolio even if they are uninformed about the quality of individual assets. In Appendix C I show that a model in which banks can shirk at the level of the individual asset generates the same conclusions as my baseline specification if financiers do indeed use a random selection rule.

The bank's decision problem can then be summarized as

$$\max_{s_B, k, b_B, \underline{a}} \mathbb{E}_z [\max \{s_B + Y_z^*(e^*) (k - a_B^*) - b_B + p(\underline{a})a_B, 0\}] + m^*(e^*)k$$

$$\text{s.t. } s_B + k \leq w_B + qb_B, \quad (5)$$

$$b_B \leq s_B + Y_z^*(e^*) (k - a_B^*) + p(\underline{a})a_B \text{ for all } z, \quad (6)$$

where  $a_B^*$  and  $e^*$  denote optimal choices in accordance with (3) and (4), and (5) and (6) are the bank's budget and solvency constraint, respectively. I analyze this problem in steps. I first derive the decision rule determining whether monitoring is consistent with the bank's ex-post optimal choice of asset sales. Second, I derive the optimal asset sale promise  $\underline{a}$ , investment  $k$  and bond issuances  $b_B$ . Banks will never invest in storage because the risky technology offers a strictly higher return ( $\hat{Y} > \bar{R}$ ) and the bank is the residual claimant. I thus take  $s_B = 0$  as given throughout.

**Ex-post optimal asset sales.** Recall that the bank chooses  $e = e^*(a_B, \underline{a}, k, s_B, b_B)$  when it sells  $a_B$  risky claims. Monitoring effort is thus conditional on asset sales, and we must worry about "double deviations" in which the bank simultaneously decides to sell more assets than promised and shirks. The objective function is linear because banks are risk neutral and (6) imposes solvency in every state of the world. The bank thus either sells everything ( $a_B = k$ ) or just as much as initially promised ( $a_B = \underline{a}$ ). Moreover, it shirks if it sells everything,  $e^*(k, \underline{a}, k, s_B, b_B) = 0$ .

Monitoring thus takes place if and only if banks monitor given  $a_B^* = \underline{a}$ . Taking this as given, the payoff to *exerting effort and holding* is  $\hat{Y}(k - \underline{a}) - b_B + p(\underline{a})\underline{a}$ , while the payoff to *shirking and selling* is  $p(\underline{a})k - b_B + mk$ . This leads to the following decision rule.

**Proposition 1** (Ex-Post Optimal Asset Sales and Monitoring).

*Assume that monitoring is optimal at  $\underline{a}$ . Then the bank sells  $\underline{a}$  assets and monitors only if*

$$p \leq \bar{p}(k, \underline{a}) \equiv \hat{Y} - m \left( \frac{k}{k - \underline{a}} \right) \quad (7)$$

Banks thus sell all their assets and shirk when the asset price is too high. Notably,  $\bar{p}(k, \underline{a}) < \hat{Y}$ , which implies that the upper bound is low enough that it is profitable to buy good assets at  $\bar{p}(k, \underline{a})$ . Sufficiently rich financiers will therefore bid up the asset price until it reaches the shirking threshold.

*Collateral constraints.* The next step is to characterize the bank's optimal choice of bonds  $b_B$  and investment  $k$ , taking as given the asset-sale promises  $\underline{a}$  and assuming that banks must monitor (which requires  $a_B^* = \underline{a}$ ). The banks' ability to issue bonds is limited by the solvency constraint (6) and by an incentive constraint that ensures banks prefer monitoring over shirking,

$$\sum_z \pi_z [Y_z (k - \underline{a}) - b_B + p(\underline{a})\underline{a}] \geq \sum_z \pi_z [\max \{y_z (k - \underline{a}) - b_B + p(\underline{a})\underline{a}, 0\}] + mk. \quad (8)$$

The incentive constraint binds before the solvency constraint because shirking generates more downside risk than monitoring ( $y_l < Y_l$ ). Let  $\Omega_z \equiv y_z (k - \underline{a}) + p(\underline{a})\underline{a}$  denote the bank's cash-on-hand in state  $z$  conditional on shirking. If  $\underline{a}$  is such that  $\Omega_z < b_B$ , then the incentive constraint is equivalent to the borrowing constraint

$$b_B \leq \bar{b}_B(k, \underline{a}) = \left[ \frac{\pi_h}{\pi_l} (Y_h - y_h) + Y_l - \frac{m}{\pi_l} \right] k + \left[ p(\underline{a}) - Y_l - \frac{\pi_h}{\pi_l} (Y_h - y_h) \right] \underline{a}$$

Banks can relax this constraint by selling assets only if shirking represents a risk-shifting problem (Jensen and Meckling (1976)).

**Observation 1** (Risk-shifting Problem).

*Asset sales promises ( $\underline{a} > 0$ ) can increase bank borrowing capacity only if  $y_h > \hat{Y}$ .*

The observation follows from noting that the asset price is bounded above by  $\hat{Y}$  by financier's demand. This implies that there exists a  $p(\underline{a})$  such that the coefficient on  $\underline{a}$  in the borrowing constraint is positive only if  $y_h > \hat{Y}$ . The losses from shirking must therefore be sufficiently concentrated in the low state. The intuition is that the insurance provided by asset sales is valuable only if the bank is constrained by a lack of capital in the low state. I assume that this condition is satisfied from now on. In order to obtain

easily interpretable closed-form solutions for equilibrium prices and trading behavior I use the following special case.<sup>8</sup>

**Assumption 2.**

*The returns of the risky technology conditional on shirking satisfy  $y_h = Y_h$  and  $y_l = 0$ .*

This assumption allows me summarize the severity of the moral hazard problem by the moral-hazard discount factor

$$\tilde{m} \equiv 1 - \frac{m}{\pi_l Y_l} \in (0, 1).$$

This statistic is close to one when the moral hazard problem is not severe ( $m$  is close to zero) and close to zero when the moral hazard problem is severe ( $m$  is close to the expected output loss from shirking  $\pi_l Y_l$ ). High values of  $\tilde{m}$  therefore indicate a loose bank moral hazard problem. The bank's borrowing constraint can then be stated as the collateral constraint

$$\bar{b}_B(k, \underline{a}) = \tilde{m} Y_l k + (p(\underline{a}) - Y_l) \underline{a}. \quad (9)$$

The first term is the collateral capacity of the risky asset itself. It is composed of the worst-case return scaled by the moral-hazard discount factor  $\tilde{m}$ . Note that the collateral capacity of the risky asset in the hands of financiers is  $Y_l$ , because financiers are not subject to moral hazard in origination. This difference in collateral capacity is the fundamental source of gains from trade between banks and financiers. The second term is the collateral capacity provided by asset sales. It exceeds that of the risky asset to the extent that  $p(\underline{a})$  is larger than the worst-case return  $Y_l$ . Using the budget constraint, the investment opportunity set of banks is  $k \leq \bar{k}(\underline{a}) = \lambda_B(q) [w_B + q(p(\underline{a}) - Y_l)\underline{a}]$ , where

$$\lambda_B(q) = \frac{1}{1 - q\tilde{m}Y_l} \quad (10)$$

---

<sup>8</sup> Note that Assumption 2 would not be innocuous if banks tranching risky assets, since then banks could sell off the "equity tranche" to obtain insurance without harming incentives. Maintaining the notion that asset sales may lead to shirking would then require that  $Y_h \neq y_h$ . I abstract from this issue in order to derive a simple collateral constraint that can be easily embedded into a dynamic setting.

is bank leverage. Asset sales thus substitute for bank equity by replacing risky returns on investment with safe asset-market returns.

Yet asset sales boost collateral only if banks do not sell too many assets. Specifically, if a bank sells so many assets that its cash-on-hand after shirking exceeds its debts ( $\Omega_z > b_B$  for all  $z$ ), then the incentive constraint can be restated as the skin-in-the-game constraint

$$\underline{a} \leq \tilde{m}k. \quad (11)$$

Banks thus shirk when they sell too many assets, and the moral hazard discount factor  $\tilde{m}$  summarizes the bank's required exposure to its investments.

**Optimal promises.** The previous section showed that banks can increase borrowing capacity by issuing a promise  $\underline{a} \in [0, \tilde{m}k]$ . I now show that banks find it optimal to do so only if the asset price is sufficiently high. Issuing bonds to finance investment allows banks to earn the leveraged intermediation premium

$$\rho(q) = \lambda_B(q)(\hat{Y}q - 1) \quad (12)$$

The benefit of selling assets at price  $p$  is that doing so generates  $p - Y_l$  units of additional collateral that can be levered to earn  $\rho(q)$ . The cost is that risky assets trade at the discount  $\hat{Y} - p$ . The bank's indirect utility function conditional on  $\underline{a}$  is

$$u_B(q, p, \underline{a}) = \lambda_B(q)(\hat{Y} - \tilde{m}Y_l)w_B + \underline{a} \left[ \rho(q)(p(\underline{a}) - Y_l) - (\hat{Y} - p(\underline{a})) \right].$$

The first-order condition with respect to  $\underline{a}$  is

$$u'_B(q, p, \underline{a}) = \left[ \rho(q)(p(\underline{a}) - Y_l) - (\hat{Y} - p(\underline{a})) \right] + p'(\underline{a})\underline{a} \left[ \lambda(\hat{Y}q - 1) + 1 \right]. \quad (13)$$

The second term reflects the price impact of changes in the promise  $\underline{a}$ , because promises may be a signal of asset quality. An instructive special case is when the price is constant

on the relevant interval  $[0, \tilde{m}k]$ . Banks are then willing to sell assets ( $u' \geq 0$ ) if

$$p(\underline{a}) \geq \underline{p}(q) = \frac{\hat{Y} + \rho(q)Y_l}{1 + \rho(q)}. \quad (14)$$

Because the same lower bound obtains when evaluating (13) at  $\underline{a} = 0$ , trade occurs only if prices exceed the lower bound  $\underline{p}(q)$ . This lower bound is strictly decreasing in  $q$  since intermediation rents  $\rho(q)$  are strictly increasing in  $q$ . Moreover,  $\underline{p}\left(\frac{1}{\hat{Y}}\right) = \hat{Y}$  because banks are not willing to sell assets below par if there are no intermediation rents to be earned. Intermediation rents thus generate the scope for asset trade and shape the pass-through of profits to financiers. Given  $q$  and a promise  $\underline{a}$ , the asset price moves within the interval  $[\underline{p}(q), \bar{p}(k(q), \underline{a})]$  as a function of the relative wealth of financiers and banks. The bond price in turn lives on the interval  $\left[\frac{1}{\underline{R}}, \frac{1}{\hat{Y}}\right]$ , and it is large when savers are wealthy relative to intermediaries. I therefore classify equilibrium outcomes as follows.

**Definition 1.**

Given  $q$  and  $\underline{a}$ , the asset market is **slack** if  $p(\underline{a}) = \underline{p}(q)$ , and **tight** if  $\underline{p}(q) \in (\underline{p}(q), \bar{p}(k(q), \underline{a}))$ . Financial intermediaries are **highly constrained** if  $q = \frac{1}{\underline{R}}$  and **constrained** if  $q \in \left(\frac{1}{\underline{R}}, \frac{1}{\hat{Y}}\right)$ .

In a slack asset market, all trading rents accrue to financiers, while banks retain some rents when asset markets are tight. Similarly, all bond market rents accrue to intermediaries when they are highly constrained, but some returns are passed on to savers when they are merely constrained. Note that there may be “excessive” financier demand for risky assets even if the asset market is slack: the former condition relates to aggregate quantities and market clearing when prices are bounded, the latter pertains to the pass-through of rents among intermediaries.

### 2.3 Equilibrium

I now turn to characterizing the competitive equilibrium. Whether or not the asset price is below its upper bound cannot be verified ex-ante. I therefore use a guess-and-verify approach. I first consider an *efficient monitoring equilibrium* in which the constraint  $p \leq \bar{p}$

is presumed to hold and all banks exert effort. I then verify whether or not the asset price violates the upper bound. If the upper bound is violated, I consider an *excessive trading equilibrium* in which a fraction of banks shirks, and show why shirking serves to clear asset markets when the asset price is bounded.

### 2.3.1 Efficient monitoring equilibrium

Asset prices in efficient monitoring equilibrium must be determined by financier wealth (“cash-in-the-market pricing”). Else, financiers would bid up prices until  $p = \hat{Y}$ , violating the price bound (7). Given that all firms monitor and pricing is cash-in-the-market, prices must be constant across all active submarkets. I thus restrict attentions to price schedules such that  $p(\underline{a}) = p$  for all  $\underline{a} \in [0, \tilde{m}k]$ . Since banks who issue a promise greater than  $\tilde{m}k$  shirk for sure, let  $p(\underline{a}) = 0$  for all  $\underline{a} > \tilde{m}k$ . Conditional on the asset price, the banks’ decision problem is simple. If the asset market is tight ( $p > \underline{p}(q)$ ), banks earn strictly positive rents by increasing borrowing capacity through asset sales. As a result, they maximize borrowing capacity by choosing  $\underline{a}^* = \tilde{m}k^*$ , yielding the portfolio

$$k^* = \frac{w_B}{1 - q\tilde{m}p}, \quad b_B^* = \tilde{m}pk^* = \frac{\tilde{m}pw_B}{1 - q\tilde{m}p} \quad \text{and} \quad \underline{a}_B^* = \tilde{m}k^* = \frac{\tilde{m}w_B}{1 - q\tilde{m}p}.$$

Observe that asset sales boost bank’s effective leverage by substituting the safe cash flow  $p$  for the risky asset’s worst-case return  $Y_l$ . The degree to which banks rely on outside collateral is determined by  $\tilde{m}$ , since banks require little skin-in-the-game when  $\tilde{m}$  is large.

The policy functions of all agents are linear in wealth, permitting straightforward aggregation. The market-clearing conditions are

$$\frac{\tilde{m}pw_B}{1 - q\tilde{m}p} + \frac{\mu(q, p)Y_l w_F}{p - \mu(q, p)qY_l} = \frac{w_S}{q} \quad \text{and} \quad \frac{\tilde{m}w_B}{1 - q\tilde{m}p} = \frac{w_F}{p - qY_l},$$

where  $x_z = Y_z$  because all banks exert effort. Conditional on  $q$ , the asset price thus is a function of relative wealth  $\omega = \frac{w_F}{w_B}$  only:  $p^*(q) = \frac{1}{\tilde{m}} \left( \frac{\tilde{m}qY_l + \omega}{1 + q\omega} \right)$ . It follows that asset markets are tight if  $p^*(q^*) \geq \underline{p}(q^*)$ , where  $q^*$  is the price that clears the bond market conditional on

$p = \underline{p}(q)$ . This is the case if

$$\omega \geq \tilde{m} \left[ \frac{\underline{p}(q^*) - \mu(q^*), \underline{p}(q^*)}{1 - q^* \tilde{m} \underline{p}(q^*)} q^* Y_l \right] \quad (15)$$

That is, asset markets are tight if financiers are wealthy relative to banks, or if financiers can take on sufficiently more leverage than banks. Note also that increases in  $\tilde{m}$  favor market slackness by boosting asset supply.

If asset markets are slack ( $p = \underline{p}(q)$ ), banks are indifferent toward asset sales and trade volumes are demand-determined within the interval  $[0, \tilde{m}k]$ . Accordingly, the bank portfolio is given by  $k^* = \frac{w_B + q(\underline{p}(q) - Y_l)a_F}{1 - q\tilde{m}Y_l}$  and  $b_B^* = \frac{Y_l \tilde{m} w_B + (\underline{p}(q) - Y_l)a_F}{1 - q\tilde{m}Y_l}$ , while the bond market clearing condition is

$$\frac{Y_l \tilde{m} w_B + (\underline{p}(q) - Y_l)a_F}{1 - q\tilde{m}Y_l} + \frac{\mu(q, \underline{p}(q))Y_l w_F}{\underline{p}(q) - \mu(q, \underline{p}(q))qY_l} = \frac{w_S}{q}, \text{ where } a_F = \frac{w_F}{\underline{p}(q) - \mu(q, \underline{p}(q))qY_l}.$$

**Equilibrium outcomes.** The key comparative statics of prices with respect to the wealth distribution are straightforward:  $q^*$  is increasing in  $w_S$  and decreasing in  $w_B$  and  $w_F$ , while  $p^*$  is increasing in  $w_F$ . These observations have immediate implications for the existence of efficient monitoring equilibrium with asset market trading.

**Proposition 2.**

1. For any  $w_S$ , there exists a cutoff  $\bar{w}_B(w_S) \geq 0$  such that no efficient monitoring equilibrium with asset trade exists if  $w_B > \bar{w}_B(w_S)$ .
2. For any  $w_S$  and  $w_B \leq \bar{w}_B(w_S)$ , there exists a cutoff  $\bar{w}_F(w_S, w_B) > 0$  such that no efficient monitoring equilibrium exists if  $w_F > \bar{w}_F(w_S, w_B)$ .

The first result follows because  $q$  is decreasing in  $w_B$  and  $\underline{p}(q)$  is decreasing in  $q$ . That is, increases in bank wealth drive down intermediation rents and push up the minimum price at which banks are willing to sell assets. Hence, there exists a  $w_B$  large enough such that the lower bound  $\underline{p}(q^*)$  violates the upper bound  $\bar{p}(k^*, 0)$  even when no assets are traded. Banks are thus willing to sell only at prices which, if financiers paid them, would lead banks to shirk. This means that asset sales are a valuable means of reallocating risk

exposure only if banks are sufficiently constrained. The second result follows because  $p$  is increasing in financier wealth. As a result, the asset price breaches its upper bound if financiers are sufficiently wealthy. As the next section will show, increases in  $w_F$  can therefore lead to shirking by banks. An implication is that cash-in-the-market pricing is a necessary condition for efficient monitoring.

When instead  $w_F$  is *not* excessively large, increases in financier wealth typically boost investment, and may do so more effectively than increases in  $w_B$ .

**Proposition 3.**

*Assume that  $w_B < \bar{w}_B(w_S)$  and  $w_F < \bar{w}_F(w_S, w_B)$ . Then  $k$  is strictly increasing in  $w_F$  if asset markets are tight, or if asset markets are slack and intermediaries are highly constrained. Moreover,  $\frac{\partial k^*}{\partial w_F} > \frac{\partial k^*}{\partial w_B}$  if asset markets are slack, intermediaries are highly constrained, and  $\underline{R} < 1$ .*

The intuition behind the first part of the proposition is straightforward. As financier wealth increases, so does the demand for risky assets. Rising asset prices mean that banks receive more collateral per risky asset sold, allowing for more investment. Since all banks monitor, expected aggregate output also increases. A caveat applies when asset markets are slack. In this case, financiers receive all asset market rents, and increases in financier wealth may lead to a drop in bond prices that crowds out bank borrowing at intermediate levels of  $w_F$ . This channel is not operational if intermediaries are highly constrained, however.

The third part of the proposition shows that the impact of increased financier wealth can be large. Specifically, increases in financier funding spur investment more than increases in bank wealth when the financial sector is highly constrained ( $q = \frac{1}{\underline{R}}$ ) and savers pay a premium for intermediation services ( $\underline{R} < 1$ ). The reason is that financiers can take on more leverage than banks when bond prices are high and asset prices are low. The separation of asset origination and the holding of the resulting risk may thus allow for higher aggregate volumes than a financial system consisting of banks only.

### 2.3.2 Excessive trading equilibrium.

I now turn to characterizing the competitive equilibrium given that financiers are too wealthy to sustain monitoring by all banks,  $w_F > \bar{w}_F(w_S, w_B)$ . The fundamental problem is that the asset price cannot increase beyond  $\bar{p}(k, a)$ , since all banks would shirk otherwise. This means that the price mechanism fails to clear the market. This section shows that equilibrium shirking will serve to clear the market when prices are fixed.

The argument is in three parts. First, banks are indifferent between shirking and effort at  $\bar{p}$ . Second, banks who shirk do so because they sell more assets than those who exert effort, raising asset supply at fixed prices. Third, financiers are willing to tolerate some shirking because they buy at  $\bar{p} < \hat{Y}$ . The only subtlety is that  $\bar{p}(\bar{k}(\underline{a}), \underline{a})$  is a function of the banks promise  $\underline{a}$ . We must therefore check whether banks have an incentive to use  $\underline{a}$  to signal the quality of their assets. Note that banks who shirk and banks who exert effort must choose the same asset-sale promises, bonds, and investment, since financiers could otherwise distinguish asset quality by conditioning on observables.

The first part of the argument holds by construction: given  $\underline{a}$ ,  $\bar{p}$  is defined to be the price at which banks are indifferent between (sell  $\underline{a}$ , effort) and (sell  $\bar{k}(\underline{a})$ , shirk). The second part follows because banks who shirk sell  $k^*$  assets, while those who exert effort sell  $\tilde{m}k^* < k^*$ . Formally, let  $\Phi \in [0, 1]$  denote the fraction of shirking banks (or, alternatively, the probability of choosing to shirk and sell). Then excess asset market demand is

$$\eta(\underline{a}, \Phi) = \frac{w_F}{\bar{p}(\bar{k}(\underline{a}), \underline{a}) - qx_b(\underline{a})} - [\Phi \bar{k}(\underline{a}) + (1 - \Phi)\underline{a}],$$

which is decreasing in  $\Phi$  but increasing in  $w_F$ , given prices. More financier wealth thus requires more shirking to clear the asset market. The third part requires that financiers weakly prefer to buy risky assets rather than invest in storage even when  $\Phi$  banks shirk. Given  $\underline{a}$ , the fraction of low-quality assets traded in the asset market is  $\phi = \frac{\Phi \bar{k}(\underline{a})}{\Phi \bar{k}(\underline{a}) + (1 - \Phi)\underline{a}} \geq \Phi$ . The risky asset's expected rate of return is  $\frac{\hat{x}}{\bar{p}(\bar{k}(\underline{a}), \underline{a})} = \frac{\phi \hat{y} + (1 - \phi)\hat{Y}}{\bar{p}(\bar{k}(\underline{a}), \underline{a})}$ . Investing in risky assets is preferable to storage if  $\phi \leq \bar{\phi}(\underline{a}) = \frac{\hat{Y} - \bar{p}(\bar{k}(\underline{a}), \underline{a})}{\hat{Y} - \hat{y}}$ , where  $\bar{\phi}(\underline{a}) > 0$  for all  $\underline{a}$  since

$\bar{p}(\bar{k}(\underline{a}), \underline{a}) < \hat{Y}$  for all  $\underline{a}$ . Financiers are thus always willing to tolerate some shirking, and the maximum sustainable  $\phi$  is such that financiers are indifferent between risky assets and storage.

Finally, we must check whether banks want to deviate from the previously optimal promise  $\underline{a}^* = \tilde{m}k$ . The key concern is that prices need no longer be constant in  $\underline{a}$  on the relevant interval  $[0, \tilde{m}k]$  because the upper bound  $p = \bar{p}(k(\underline{a}), \underline{a}) = \hat{Y} - m \frac{\bar{k}(\underline{a})}{k(\underline{a}) - \underline{a}}$  is itself a function of  $\underline{a}$ . As a result, banks may adjust quantities in order to affect prices. Naturally, banks' incentives to shade their promises depend on the price schedule  $p(\cdot)$ , which encodes prices in both active and inactive submarkets.

In the presence of wealth constraints and a feedback from prices to incentives, fully specifying this price schedule would require a theory of counterfactual off-the-equilibrium-path market tightnesses for all  $\underline{a} \in [0, \tilde{m}k]$ . In order to arrive at robust predictions that do not rely on the particulars of such a theory, I assume that  $p^*(\underline{a}) = \bar{p}(\bar{k}(\underline{a}), \underline{a})$  for all  $\underline{a} \in [0, \tilde{m}k]$  if asset markets are tight, while  $p^*(\underline{a}) = \underline{p}(q^*)$  for all  $\underline{a} \in [0, \tilde{m}k]$  if asset markets are slack. Consistent with the notion that rents accrue to the short side of the market, prices are thus constant on the relevant interval only if markets are slack, while banks receive the highest possible price after any deviation when asset markets are tight. Underlying this assumption is the notion that asset-market deviations by individual financiers do not affect bond-market clearing, so that  $q^*$  is invariant to such deviations.

Banks thus choose the same portfolio as in efficient monitoring equilibrium when asset markets are slack. The next result provides a sufficient condition that ensures banks do not want to locally deviate from the benchmark portfolio  $\underline{a}^* = \tilde{m}k^*$  when asset markets are tight.<sup>9</sup> The key feature of the model that delivers this result is that the price bound  $\bar{p}(\cdot)$  is continuous in  $\underline{a}$ . Marginal changes in  $\underline{a}$  thus cannot result in discrete price increases even if investors' beliefs about asset quality were to jump discontinuously. Indeed, the price might *fall* upon a deviation: because the price bound  $\bar{p}(k, \underline{a}) = \hat{Y} - m \frac{k}{k - \underline{a}}$  is increasing in  $k$ , reductions in  $\underline{a}$  that lead  $k$  to fall may also depress the price banks receive upon

<sup>9</sup> In Appendix D, I verify numerically that there are typically no profitable global deviations either.

a deviation. The proposition makes use of this intuition to derive the stated sufficient condition.

**Proposition 4.**

*There is no profitable local deviation to  $\underline{a} < \underline{a}^* = \tilde{m}k^*$  if  $\frac{\tilde{m}}{1-\tilde{m}} > \frac{1}{q[\pi_l(\pi_l Y_l) + \pi_h m]}$ . This condition holds for all  $q$  if  $\frac{\tilde{m}}{1-\tilde{m}} > \frac{\hat{Y}}{[\pi_l(\pi_l Y_l) + \pi_h m]}$ . Fixing all other parameters, there exists  $m$  sufficiently close to zero such that there is no profitable local deviation for all  $q$ .*

That is, banks do not want to decrease their reliance on outside funding when intermediation rents are large ( $q$  is high) and when they can easily collateralize outside funding in order to borrow more ( $\tilde{m}$  is large). As a result, individually rational bank portfolios may not provide the market discipline required to prevent shirking. Going forward, I therefore focus on the excessive trading equilibrium in which banks choose the same portfolio as in efficient monitoring equilibrium, and verify that banks do not want to deviate. The next result shows that increases in  $w_F$  lead to more shirking by boosting excess demand.

**Proposition 5.**

*The share of shirking banks  $\Phi$  is strictly increasing in  $w_F$ . If financiers do not borrow or if intermediaries are highly constrained, then a partial destruction of financier wealth from  $w_F$  to any  $w_F \geq \bar{w}_F(w_S, w_B)$  is Pareto-improving.*

The inefficiency driving the second part of the proposition stems from a pecuniary externality: individual financiers do not internalize that their demand harms bank incentives. Under the stated conditions, the resulting decline in investment efficiency is severe enough that a destruction of financier wealth can make all agents better off. This externality also contributes to excess aggregate risk exposure because bad assets are riskier than good assets.

Figure 2 provides an illustration of the equilibrium effects of  $w_F$ . The gray line with circular markers corresponds to efficient monitoring equilibrium, while the black line with diamond markers depicts the excess trading equilibrium. The top left panel shows

asset prices, with the upper line representing  $p$  and the lower line showing  $q$ .  $p$  is increasing in  $w_F$  while  $q$  is decreasing.<sup>10</sup> Initially, increases in  $w_F$  boost investment (top middle panel) and expected output (top right panel). Eventually, the asset price reaches its upper bound. Investment no longer increases, but expected output declines because a growing fraction of banks shirk (bottom left panel). The top right panel shows the increase in aggregate risk, with the dotted lines depicting aggregate output after a good and a bad shock, respectively. Given that low-quality assets are more exposed to downside risk, increases in financier wealth lead to poorer worst-case outcomes. The two last figures in the bottom row show the risk exposure of both classes of financial intermediary. The solid line depicts the *expected* wealth of an intermediary at the end of the period, with the dotted lines corresponding to a high and low aggregate shock, respectively. Financiers take on a greater fraction of total risk exposure as  $w_F$  increases.<sup>11</sup>

### 3 Dynamics

I now incorporate the static model into an overlapping generations setting to study the endogenous evolution of the wealth distribution. In doing so, I show that the model can generate credit booms with falling asset quality. Time is discrete and runs from 0 to  $T$ . A generic period is indexed by  $t$ . There are overlapping generations of financiers and banks, each of whom lives for two periods. Savers live for one period only. Intermediaries in the first period of their life are called the *young*, while those in the second period are called the *old*. There are two goods: a consumption good and an intermediary wealth good. The former is used for consumption, the latter is essential as collateral for financial intermediation. Every generation of agents is born with an endowment of the consumption good. Only the initial generation of intermediaries is born with an endowment of wealth.

---

<sup>10</sup> Equilibrium outcomes are kinked because the economy transitions from slack to tight asset markets, and from regions in which financiers want to borrow to those in which it does not. Since financiers benefit disproportionately, all results are robust to requiring that financiers always find it optimal to issue bonds. This can be achieved by choosing  $w_S$  to be suitably large.

<sup>11</sup> Note that bank utility is constant in the excessive trading equilibrium because the private benefit shirking compensates for the decline in expected end-of-life wealth.

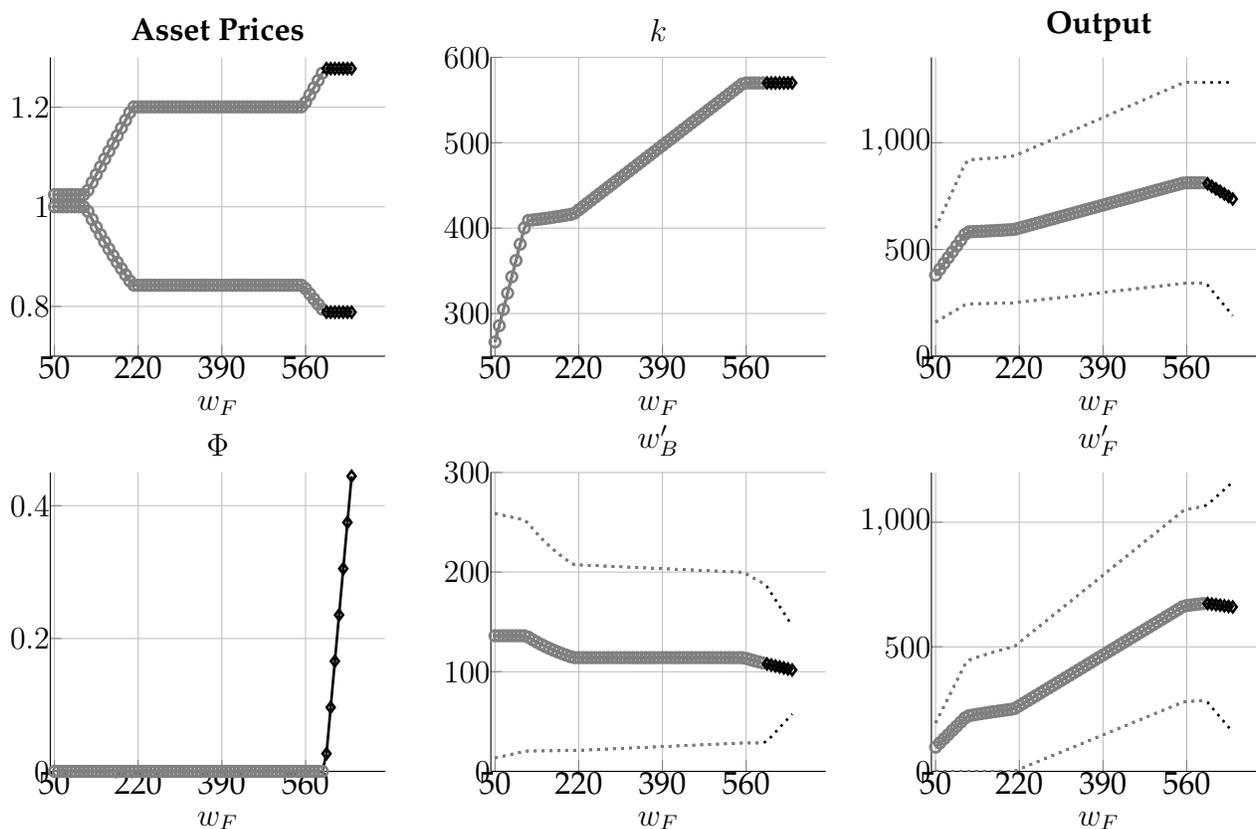


Figure 2: Equilibrium Outcomes as a Function of Financier Net Worth  $w_F$ .

Young intermediaries must therefore purchase their predecessors' wealth to be able to intermediate funds. In the first period of their life, intermediaries engage in intermediation as in the static model. Before consuming their end-of-life wealth in the second period, they can sell this wealth to young intermediaries in an intergenerational market.<sup>12</sup>

I proceed in two steps. First, I assume that the young intermediaries have all the bargaining power in the intergenerational market. This means that the old always receive the dollar value of their end-of-period wealth, and that the objective function of a young intermediary is to maximize expected end-of-life wealth. The advantage of this setting is that there are no endogenous hedging motives, so that the dynamic model is equivalent to repeating the static model period-by-period, with the evolution of the wealth distribution

<sup>12</sup> I assume that each young intermediary of a given type receives an equal share of the capital passed down by the previous generation, so as to not have to keep track of a wealth distribution within each type of intermediary. All results go through without this assumption. If a generation of intermediaries has zero wealth at the end of their life, then the new generation receives start-up funds of  $\epsilon$ . This ensures that both types of intermediaries are always active. I also assume that the young's endowment of the consumption is always large enough to purchase all old wealth.

linking periods. This setting transparently illustrates how risky asset sales generate credit booms with falling asset quality. Second, I consider the case where the old have all the bargaining power (see Appendix B). The objective function of young intermediaries then is to maximize the state-contingent *value* of end-of-life wealth. This leads to endogenous hedging motives. I show that bank risk aversion strengthens their incentives to shirk and sell off risky assets because doing so reduces risk exposure, while financier risk aversion tempers their desire to purchase risky assets. However, financier willingness-to-pay is not typically a binding constraint because pricing is cash-in-the-market. As a result, the model with risk aversion admits the same qualitative dynamics as the baseline model because risk-averse financiers are still willing to buy risky assets at  $\bar{p}$ .

### 3.1 All bargaining power to the young

When the young have all the bargaining power, the evolution of wealth follows directly from the statically optimal asset portfolios derived in Section 2. Naturally,  $w_F$  and  $w_B$  increase or decrease jointly because all risk is aggregate. While aggregate credit volumes and investment must increase after a good shock, the evolution of relative financier wealth  $\omega = \frac{w_F}{w_B}$  will determine the evolution of asset prices and monitoring incentives.

#### Proposition 6.

Let  $\omega_{t+1}^*(z_t)$ ,  $q_{t+1}^*(z_t)$  and  $p_{t+1}^*(z_t)$  denote the relative wealth of financiers, the bond price, and the asset price in period  $t + 1$  conditional on shock  $z_t$  in period  $t$ . If the asset market is tight in period  $t$ , then

(i) if financiers prefer to not issue bonds, then  $\omega_{t+1}^*(z_t) = \bar{\omega} \equiv \frac{\tilde{m}}{1-\tilde{m}}$ . Moreover,  $q_{t+1}^*(h) \leq q_t^*$  and  $p_{t+1}^*(h) \geq p_t^*$  if  $\omega_t \leq \bar{\omega}$ .

(ii) if financiers strictly prefer issue bonds, then  $\omega_{t+1}^*(z_t) = \underline{\omega}(z_t) \equiv \frac{\tilde{m}}{1-\tilde{m}} \frac{Y_{z_t} - Y_l}{Y_{z_t}}$  and  $q_{t+1}^*(h) \leq q_t^*$ . If  $q_{t+1}^*(h)$  is such that financiers strictly prefer to borrow in  $t + 1$ , then  $p_{t+1}^*(h) \geq p_t^*$  if  $\omega_t \leq \underline{\omega}(h)$  and  $\omega_t \geq (\tilde{m}Y_l)^{\frac{1}{2}}$ . Fixing all other parameters, there exists an  $m$  sufficiently close to zero such that  $\underline{\omega}(h) > (\tilde{m}Y_l)^{\frac{1}{2}}$ .

The equilibrium wealth distribution is thus shaped by  $\tilde{m}$  when asset markets are tight. Financiers grow to be large when the collateral value of safe cash-flows is large ( $\tilde{m}$  is large) because banks respond by selling more risk exposure. Moreover, the endogenous evolution of  $\omega$  is such that the asset price must rise during a sequence on good shock as long as  $\tilde{m}$  satisfies the given parametric condition. Since credit volumes also increase after good shocks, the model thus allows for credit booms with declining investment efficiency.

The previous result relied on  $\omega$  being large enough for asset markets to be tight. The next result shows that  $\omega$  and  $p$  may grow after good shocks even if  $\omega$  is initially small.

**Proposition 7.**

*If asset markets are slack and financiers issue bonds, then  $p_{t+1}^* \geq p_t^*$  and  $\omega_{t+1}^*(h) > \omega_t$  if*

$$r_B(h) - r_F(h) < (1 - \tilde{m})\lambda_F(\underline{p}(q_t), q_t)Y_l \left( \frac{Y_h - \hat{Y}}{\hat{Y} - Y_l\tilde{m}} \right) \omega_t. \quad (16)$$

*where  $r_B(h) = \lambda_B(q_t)(Y_h - \tilde{m}Y_l)$  and  $r_F(h) = \lambda_F(\underline{p}(q_t), q_t)(Y_h - Y_l)$ . This inequality is satisfied for any  $\omega_t > 0$  if  $q_t \geq 1$ .*

Financiers thus begin to grow relative to banks as long as the the interest rate is sufficiently low. The right-hand side of (16) is the risk transfer from banks to financiers. It is increasing in  $\omega_t$  because asset market volumes are demand-determined when the market is slack. The left-hand side is the difference in realized returns on equity conditional on  $z_t = h$ . Since financiers are not subject to moral hazard, they can leverage disproportionately at low interest rates. This advantage allows them to earn higher returns on equity when  $q \geq 1$ . Caballero, Farhi, and Gourinchas (2008) and Krishnamurthy and Vissing-Jorgensen (2015) suggest that strong demand for safe assets led to a decrease in the risk-free rate prior to the 2008 financial crisis, while Adrian and Shin (2010) document that non-bank intermediaries grew disproportionately during this period. Credit booms with falling investment efficiency can thus be triggered by saving gluts and capital inflows even when financiers are very small too begin with.

Equation (16) also shows that financier wealth may substitute for low interest rates,

since wealthy financiers can take on sufficient risk exposure with little leverage. Financiers thus grow either because they can borrow cheaply or because they are rich enough to begin with. Figures 3 and 4 shows this reliance on initial conditions. I plot the evolution of financier and bank wealth after a sequence of positive aggregate shocks. In both figures, the left panel depicts a baseline scenario in which financier wealth is smaller than bank wealth initially, but grows to be larger over time. The right panel depicts deviations from this baseline. Figure 3 shows that financiers fail to catch up to banks if  $w_F$  is too small initially, while Figure 4 shows that reductions in saver wealth that lower the bond price have the same effect.

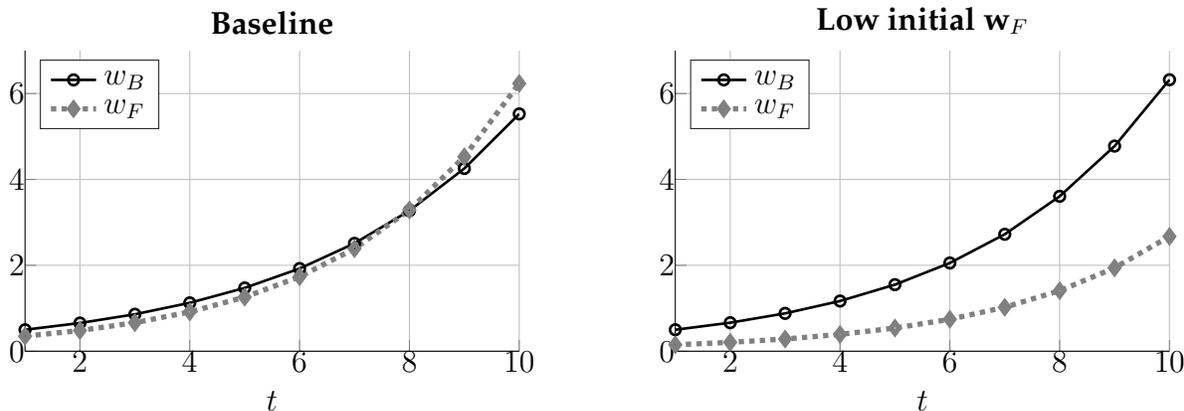


Figure 3: The effects of initial conditions – reduction in initial financier wealth  $w_F^0$ . Baseline parameter values:  $\pi_h = 0.8, Y_l = 0.5, Y_h = 1.2, \tilde{m} = 0.82$ . Initial wealth distribution:  $(w_S, w_B^0, w_F^0) = (25, 0.5, 0.35)$ . Comparative static:  $w_F^0$  from 0.3 to 0.15.

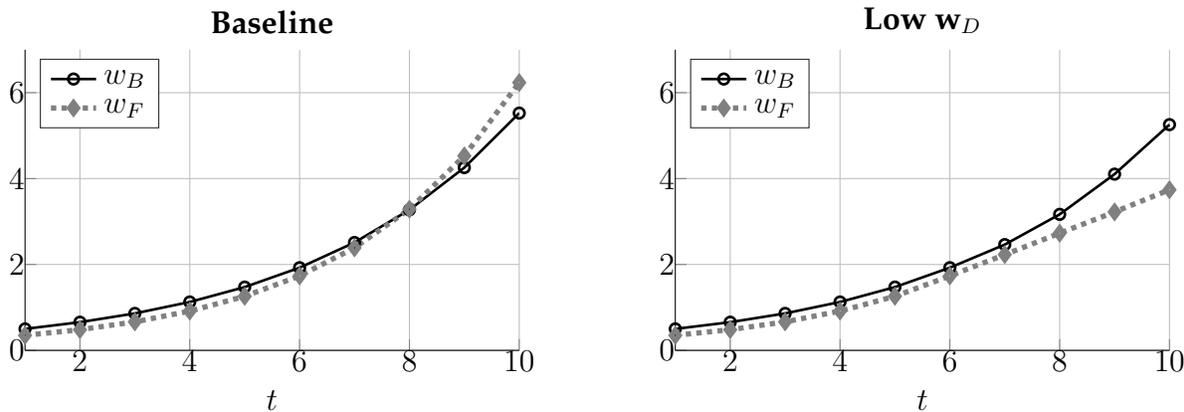


Figure 4: The effects of initial conditions – reduction in depositor wealth  $w_S$ . Baseline parameter values:  $\pi_h = 0.8, Y_l = 0.5, Y_h = 1.2, \tilde{m} = 0.82$ . Initial wealth distribution:  $(w_S, w_B^0, w_F^0) = (25, 0.5, 0.35)$ . Comparative static:  $w_S$  from 25 to 5.

*Characteristics of credit booms.* I now study the properties of credit booms by computing equilibrium outcomes given an initial wealth distribution  $\mathbf{w}^0 = (w_S^0, w_B^0, w_F^0)$  and a time path for the exogenous shock  $z$ . I simulate the economy for  $T$  periods. The initial  $T_{\text{boom}}$  shocks are good shocks. The next  $T_{\text{crisis}}$  shocks are negative. The remaining shocks are good.

Figure 5 depicts a typical credit boom with falling asset quality. I simulate the economy for 11 periods. There is an initial period, 8 positive shocks, a single negative shock, and then another positive shock. Financiers and banks each start out with 0.5 units of wealth. Initial conditions are such that the economy starts out in a efficient monitoring equilibrium. The left panel plots the evolution of wealth over time. There is a build-up of intermediary wealth, with financiers growing faster than banks. Financier wealth collapses sharply when the bad shock hits because they are disproportionately exposed to risk. Bank wealth drops only moderately because financiers provide partial insurance to banks. The middle panel plots the evolution of investment over time. The black line with circular markers depicts total investment, while the dashed gray line depicts the fraction of investment flowing to low-quality projects. Initially all banks exert effort. Over time, continued financier growth pushes the economy into an excessive trading equilibrium, and investment efficiency declines. The right panel plots the evolution of output. The solid black line depicts actual output, while the dashed gray line depicts output in the counterfactual economy in which capital accumulation is unaltered but all banks are forced to exert effort. During the boom phase, output increases steadily, but it collapses once the bad shock hits. The comparison between the solid and dashed lines shows that almost one third of the drop is accounted for by falling investment efficiency. Excessive demand for risky assets can therefore generate credit booms that end in sharp crises.

Figure 6 shows the importance of the moral hazard parameter  $\tilde{m}$ . While I set  $\tilde{m} = 0.82$  in Figure 5, I now set  $\tilde{m} = 0.85$ . Initial conditions and all aggregate shocks are the same for both simulations. Four observations stand out. First, financier wealth grows *faster* when  $\tilde{m}$  is large, even though increases in  $\tilde{m}$  allow banks to lever more. The reason is

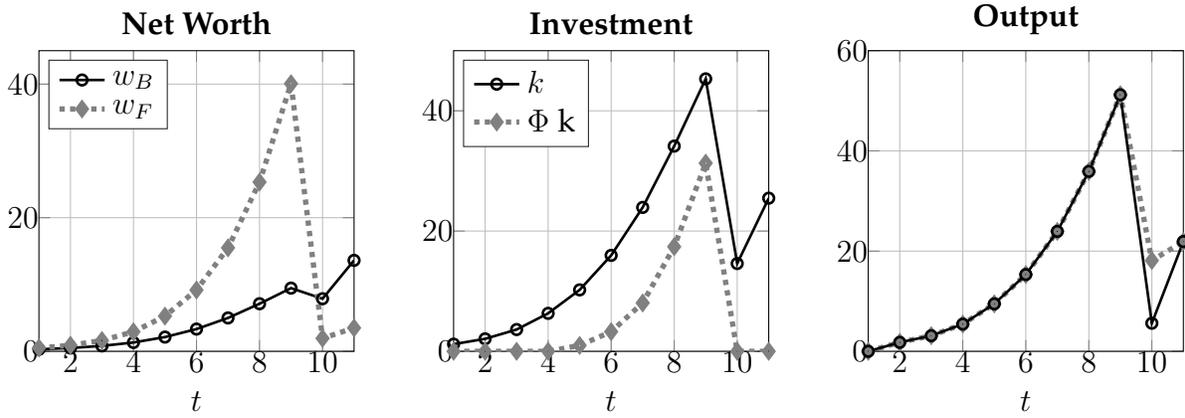


Figure 5: Credit Boom - low  $\tilde{m}$ . Parameter values:  $\pi_h = 0.65, Y_l = 0.4, Y_h = 1.5, \tilde{m} = 0.82$ . Initial wealth distribution:  $(w_B^0, w_F^0) = (0.5, 0.5)$ . Depositor wealth:  $w_S = 450$ .

that the shadow value of collateral is increasing in  $\tilde{m}$ , and banks use asset sales to acquire collateral. Increases in  $\tilde{m}$  thus boost asset supply, reducing asset prices and allowing financiers to purchase more risk exposure per unit of wealth. Increases in *potential* bank leverage thus lead to faster financier growth.

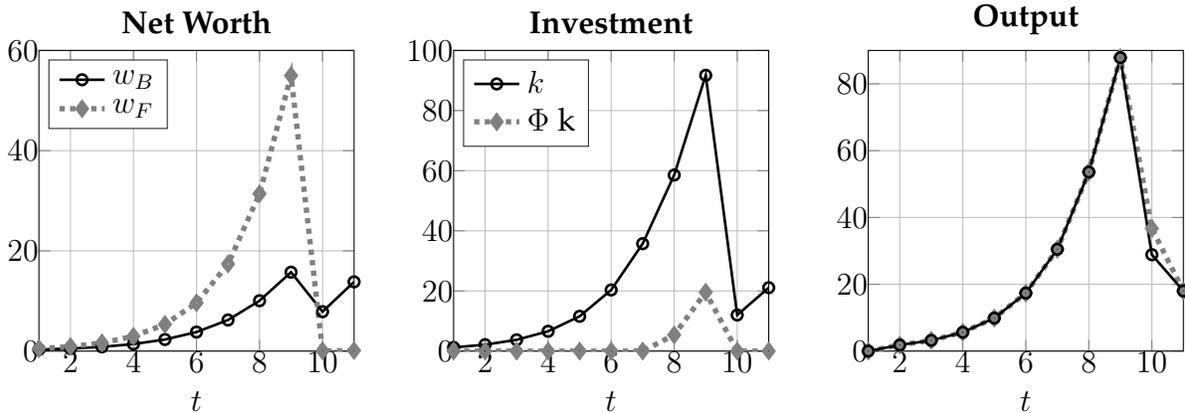


Figure 6: Credit Boom - high  $\tilde{m}$ . Parameter values:  $\pi_h = 0.65, Y_l = 0.4, Y_h = 1.5, \tilde{m} = 0.85$ . Initial wealth distribution:  $(w_S^0, w_B^0, w_F^0) = (450, 0.5, 0.5)$ .

Second, aggregate investment also grows faster because banks can take on more leverage. Third, investment efficiency is higher because an increase in asset supply lowers excess demand conditional on the wealth distribution. Fourth, higher asset supply increases financiers risk exposure and lowers bank risk exposure. Banks thus suffer less after a bad shock.

Next, I turn to the effects of boom duration. Figure 7 plots two simulated time paths

for identical parameters and initial conditions. The only difference is the timing of the negative shock. Solid lines depict the case where the negative shock hits in period 9, while dashed lines depict the case where the negative shock hits in period 8.

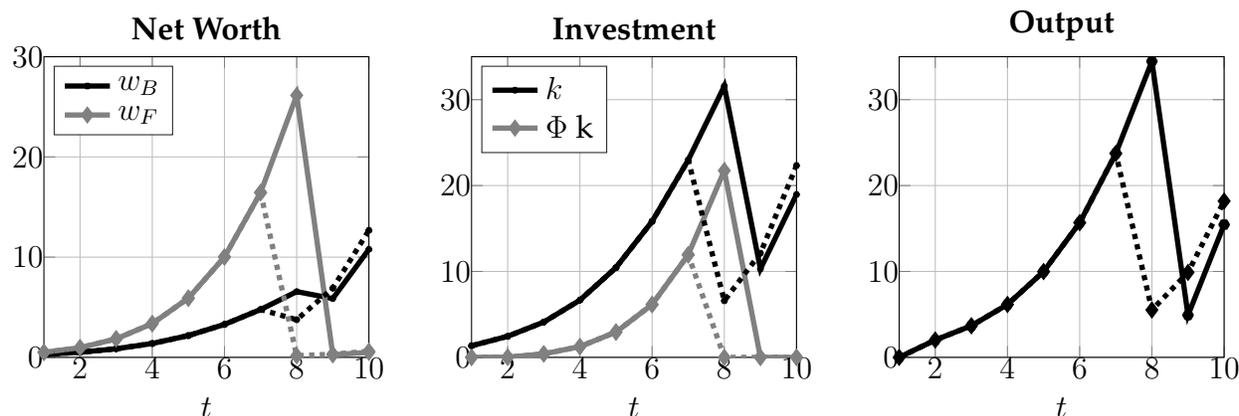


Figure 7: **Effects of increased boom length** - negative shock in period 9 (solid) vs. negative shock in period 8 (dashed).

while dashed lines depict the case where the negative shock hits in period 8. The decline in investment efficiency is more pronounced in the longer boom, as is the distance from peak to trough. Longer booms thus sow deeper crises by gradually eroding monitoring incentives.

Figure 8 shows how the duration of a crisis shapes the recovery. The solid line depicts a simulation in which a single negative shock hits in period 8. The dashed line depicts a simulation in which there are negative shocks in period 8 and 9. The left panel

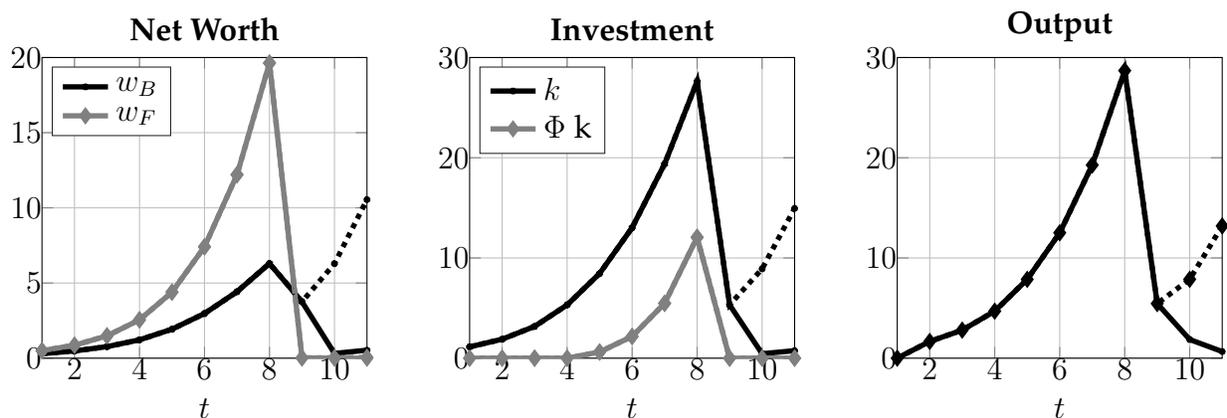


Figure 8: **Effects of increased crisis duration** - negative shock in period 8 (solid) vs. negative shock in periods 8 and 9 (dashed).

depicts the evolution of wealth, and shows that the model generates a migration of risk

exposure back onto bank balance sheets once the initial negative shock has depleted financier wealth. As a result, the second negative shock leads to a dramatic fall in bank wealth. This is consistent with the evidence in Krishnamurthy, Nagel, and Orlov (2014) that credit conditions were poor in the aftermath of the 2008 financial crisis because risk migrated back onto bank balance sheets.

*Relative returns on equity during booms.* The adverse effects of asset sales stem from an imbalance between banks and financiers. This raises the question of whether intermediaries have incentives to correct this imbalance by reallocating equity. While issuing equity is considered to be costly in practice, the next result shows that financiers may not want to purchase equity in banks even if issuing equity is costless and financiers are excessively large. The intuition is that the rents earned by financiers and banks are partly determined by depositor's bond demand, with financiers in particular benefiting from low borrowing costs. The model is thus robust to allowing for (costly) endogenous equity issuances.

**Proposition 8.**

*There exist parameters such that the model generates credit booms with falling asset quality in which relative financier wealth  $\omega$  grows, financiers earn higher expected returns on equity than banks, and financiers optimally choose not to buy bonds issued by banks.*

## 4 Policy

I now study the effects of policy by considering the positive implications of three policies: monetary policy as a determinant of short-term interest rates, restrictions on bank leverage, and macro-prudential tools to eliminate pecuniary externalities in asset markets.

*Monetary policy.* Krishnamurthy and Vissing-Jorgensen (2012) argue that safe assets produced by the financial system are a substitute for treasuries. In line with this mechanism, I study monetary policy in reduced form by assuming that the short-term rate affects savers' required return on bonds. That is,  $\underline{R} = M(\rho)$ , where  $\rho$  denotes the tightness

of monetary policy and  $M'(\rho) > 0$ . To assess the effects of monetary policy, I assume that the financial system is highly constrained. Tight monetary policy then serves to reduce the bond price. To evaluate whether loose monetary policy can trigger financier growth that would not have occurred in its absence, I assume that asset markets are slack initially. The combination of these two assumptions implies that all banks exert effort. Proposition 7 shows that looser monetary policy (leading to an increase in  $q$ ) must increase both investment and the growth rate of relative financier wealth after a good shock. This means that loose monetary policy may increase the risk of lower investment efficiency in the future by boosting the growth of financiers, giving rise to a *dynamic* risk-taking channel of monetary policy. Altunbas, Gambacorta, and Marques-Ibanez (forthcoming) provide evidence for this mechanism: extended periods of loose monetary policy are associated with increased risk-taking and higher default risk among financial institutions, but with a lag. Note also that short-lived monetary impulses may have persistent effects, because financiers who are already wealthy need to rely less on low interest rates to continue growing. This concern is depicted in Figure 9, which plots all combinations of  $(q, \omega)$  such that  $\omega$  is constant over time. A policy that lowers the risk-free rate from  $i^0$  to  $i^1$  may set financiers on a path towards growth that cannot be halted merely by reversing the policy action.

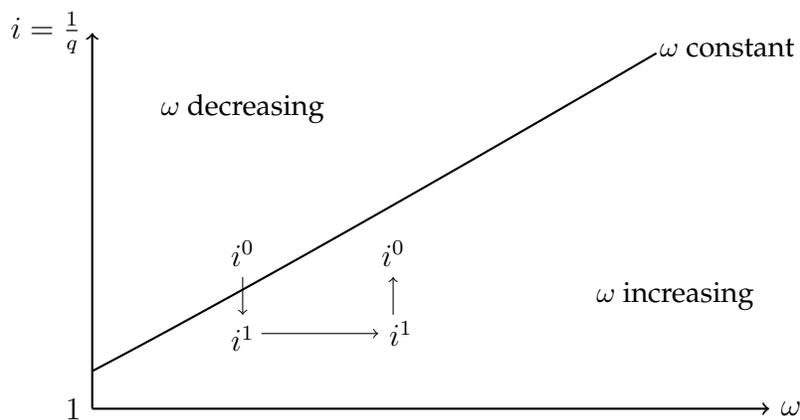


Figure 9: **Persistent effects of monetary policy** - iso- $\omega$  curve in  $(i, \omega)$  space. Monetary policy shift from  $i^0$  to  $i^1$  and back to  $i^0$ .

*Caps on bank leverage.* Next, consider bank leverage caps that restrict banks from investing more than a fixed multiple of their wealth,  $k \leq \bar{\lambda}_B w_B$ . For simplicity, I focus on regions of the state space where asset markets are tight in the absence of leverage caps. Recall that bank leverage in the absence of secondary markets and leverage caps is  $\lambda_B^0 = \frac{1}{1-q^* Y_I \bar{m}}$ , while equilibrium leverage in the absence of leverage caps is  $\lambda_B^* = \frac{1}{1-q^* \bar{m} p^*}$ . For leverage requirements to influence equilibrium outcomes without shutting down asset markets altogether, I assume that  $\lambda_B^0 < \bar{\lambda}_B < \lambda_B^*$ . In the presence of leverage caps, banks sell just enough assets to exactly hit the leverage constraint. A binding leverage constraint therefore acts as a negative supply shock in the asset market, boosting asset prices and harming origination incentives.

**Proposition 9.**

*Fix an efficient monitoring equilibrium with tight asset markets and highly constrained intermediaries. If  $\bar{\lambda}_B < \lambda_B^*$ , then  $\Phi^* > 0$  once leverage caps are introduced.*

Figure 10 plots equilibrium outcomes as a function of the leverage cap  $\bar{\lambda}_b$  when asset markets are tight in the absence of leverage caps. The top row shows that asset sales, investment, and bond supply all fall as the leverage cap shrinks. The left two panels on the bottom row show that  $p$  and  $\Phi$  both decrease in  $\bar{\lambda}_b$  because there is more excess demand when banks sell fewer assets. The bottom-right panel plots the evolution of relative financier wealth  $\omega$  conditional on a good aggregate shock. There are two countervailing forces. First, tighter leverage caps lead banks to reduce asset-sale promises and *decrease* the amount of risk transferred. Second, an increase in the fraction of shirking banks leads to an *increase* in risk transfer because shirking bank sell more assets. When leverage caps are not too tight, the first effect dominates and  $\omega$  grows more slowly in the constrained equilibrium. When leverage caps are tight, the second effect dominates and relative financier wealth grows faster. Note that the latter channel is likely to be particularly strong when leverage caps are risk-weighted and selling assets allows banks to circumvent regulations. The model's predictions are thus consistent with the regulatory-arbitrage view of securitization articulated in Acharya, Schnabl, and Suarez (2013).

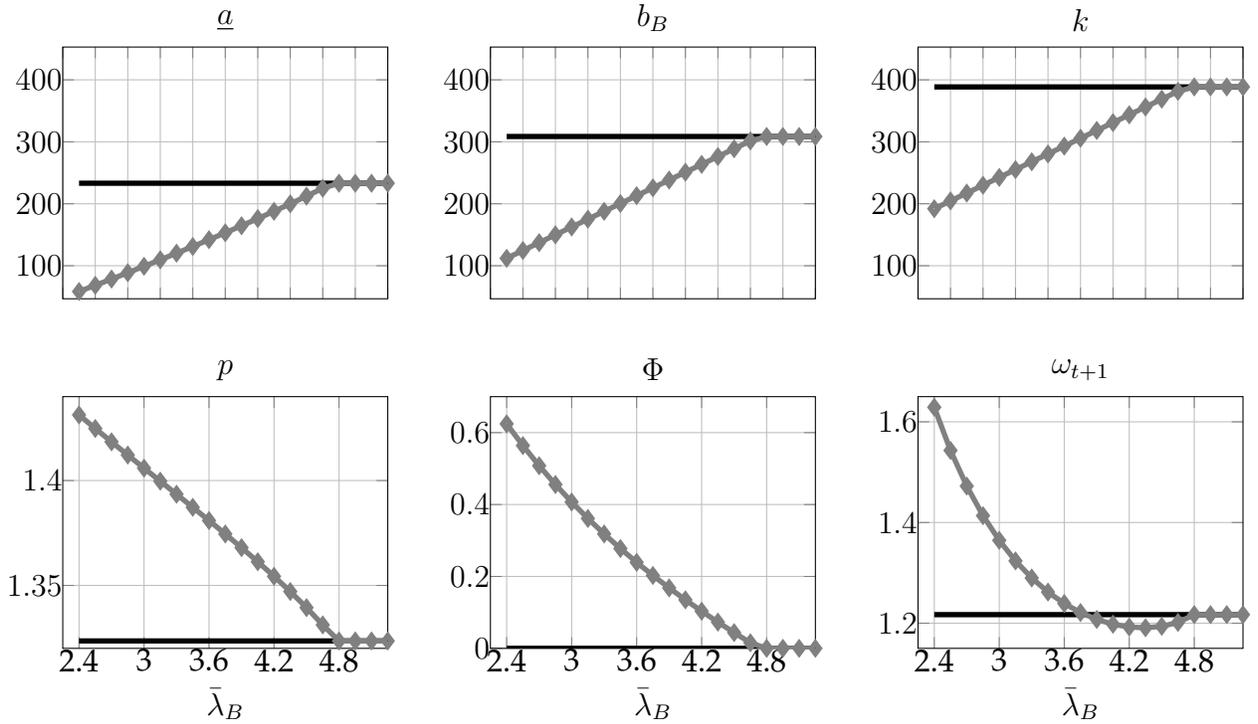


Figure 10: Equilibrium outcomes as a function of the bank leverage cap  $\bar{\lambda}_B$ . The unconstrained equilibrium is in solid black, the equilibrium with leverage caps is in gray. Leverage in the unconstrained equilibrium is equal to  $\lambda_B^* = 4.85$ . Parameter values:  $\pi_h = 0.5$ ,  $Y_l = 0.5$ ,  $\hat{Y} = 2.65$ ,  $\underline{R} = 1$ ,  $m = 0.1$ . Wealth distribution:  $w_S = 1000$ ,  $w_B = 80$ ,  $w_F = 190$ .

**Equity injections and macro-prudential market interventions.** Proposition 3 showed that aggregate lending volumes may increase more sharply in  $w_F$  than  $w_B$  when intermediaries are highly constrained and asset markets are slack. Because financiers take on disproportionate risk exposure during upturns, their wealth also falls disproportionately during crises. This suggests that providing capital to financiers may be more cost-effective than providing capital to banks if asset markets are impaired during crises.

On the other hand, excessively wealthy financiers cause investment inefficiencies. This suggests that macro-prudential policy should manage asset demand more generally. A tool that accomplishes this goal in the context of this model is a pro-cyclical cap on financiers' asset purchases, chosen to eliminate excessive demand when  $w_F$  is too large. Such a policy naturally eliminates within-period inefficiencies, and may also have dynamic benefits. Indeed, it is easy to see that aggregate intermediary wealth  $w_F + w_B$  is larger in any state of the world. The reason is that the elimination of shirking improves

the allocation of capital and boosts output. The motive underlying the policy is entirely independent of financier capital structure, and thus applies equally to zero-leverage intermediaries, such as asset managers, that were previously outside the scope of regulation precisely because their lack of leverage was thought to eliminate fragility and agency frictions. Finally, the aggregate size of the financier sector, rather than the systemic relevance of individual financial institutions, is the relevant concern.

## 5 Conclusion

This paper offers a theory of credit cycles in which the distribution of wealth and aggregate risk across financial intermediaries determines credit volumes and investment efficiency. Some risk transfer from lenders to non-lender intermediaries boosts credit volumes by relaxing collateral constraints, but investment efficiency declines when lenders sell too much risk exposure. The latter channel dominates when the buyers of risky assets are wealthy relative to lenders. Because those who carry risk exposure grow wealthy during good times, macroeconomic upturns generate credit booms with falling investment efficiency. The model's empirical predictions are in line with empirical evidence on credit booms and the role of securitization in prominent financial crises.

Credit cycles can be triggered by low interest rates. The model thus provides a link from expansionary monetary policy and "saving gluts" to future investment inefficiency. I also show that restrictions on lender leverage may be harmful, and that pro-cyclical constraints on purchases of asset-backed securities may be welfare-enhancing. There are two main avenues for future research. The first is to study the optimal design of policy in the context of secondary market trading. The second is to undertake a quantitative evaluation of the mechanisms proposed in this paper.

## References

- ACHARYA, V. V., P. SCHNABL, AND G. SUAREZ (2013): "Securitization without Risk Transfer," *Journal of Financial Economics*, 107(3), 515–536.
- ADRIAN, T., AND H. S. SHIN (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), 418–437.
- ALTUNBAS, Y., L. GAMBACORTA, AND D. MARQUES-IBANEZ (forthcoming): "Does Monetary Policy Affect Bank Risk?," *International Journal of Central Banking*.
- ATTAR, A., T. MARIOTTI, AND F. SALANIÉ (2011): "Nonexclusive Competition in the Market for Lemons," *Econometrica*, 79(6), 1869–1918.
- BIANCHI, J. (2011): "Overborrowing and Systemic Externalities in the Business Cycle," *American Economic Review*, 101(7), 3400–3426.
- BIANCHI, J., AND E. G. MENDOZA (2012): "Overborrowing, Financial Crises and 'Macroprudential' Policy," *University of Pennsylvania Working Paper*.
- BIGIO, S. (2014): "Financial Risk Capacity," *Working Paper*.
- BOLTON, P., T. SANTOS, AND J. A. SCHEINKMAN (2016): "Savings Gluts and Financial Fragility," *Columbia University Working Paper*.
- BRUNNERMEIER, M. K. (2009): "Deciphering the Liquidity and Credit Crunch 2007–2008," *Journal of Economic Perspectives*, 23(1), 77–100.
- CABALLERO, R. J., AND E. FARHI (2014): "The Safety Trap," *NBER Working Paper 19927*.
- CABALLERO, R. J., E. FARHI, AND P.-O. GOURINCHAS (2008): "An Equilibrium Model of "Global Imbalances" and Low Interest Rates," *American Economic Review*, 98(1), 358–93.
- CABALLERO, R. J., AND A. KRISHNAMURTHY (2009): "Global Imbalances and Financial Fragility," *American Economic Review, Papers and Proceedings*, 99(2), 584–88.
- CHARI, V., A. SHOURIDEH, AND A. ZETLIN-JONES (2014): "Reputation and Persistence of Adverse Selection in Secondary Loan Markets," *American Economic Review*, 104(12), 4027–70.
- COVAL, J. D., J. W. JUREK, AND E. STAFFORD (2009): "Economic Catastrophe Bonds," *American Economic Review*, 99(3), 628–66.
- FOSTEL, A., AND J. GEANAKOPOLOS (2008): "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98(4), 1211–1244.
- (2015): "Leverage and Default in Binomial Economies: A Complete Characterization," *Econometrica*, 83(6), 2191–2229.
- GARLEANU, N., AND L. PEDERSEN (2011): "Margin-based Asset Pricing and the Law of One Price," *Review of Financial Studies*, 24(6), 1980–2022.

- GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2012): "Neglected risks, financial innovation, and financial fragility," *Journal of Financial Economics*, 104, 452–468.
- (2013): "A Model of Shadow Banking," *The Journal of Finance*, 68(4), 1331–1363.
- GORTON, G., AND A. METRICK (2012): "Securitized Banking and the Run on Repo," *Journal of Financial Economics*, 104(3), 425–451.
- GORTON, G., AND G. ORDOÑEZ (2014): "Collateral Crises," *American Economic Review*, 104(2), 343–78.
- GREENLAW, D., J. HATZIUS, A. K. KASHYAP, AND H. S. S. SHIN (2008): "Leveraged Losses: Lessons from the Mortgage Market Meltdown," in *Proceedings of the U.S. Monetary Policy Forum 2008*.
- GREENWOOD, R., AND S. G. HANSON (2013): "Issuer Quality and Corporate Bond Returns," *Review of Financial Studies*, 26(6), 1482–1525.
- GRIFFIN, J., R. LOWERY, AND A. SARETTO (2014): "Complex Securities and Underwriter Reputation: Do Reputable Underwriters," *Review of Financial Studies*, 27(10), 2872–2925.
- GROMB, D., AND D. VAYANOS (2002): "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics*, 66, 361–407.
- IVASHINA, V., AND Z. SUN (2011): "Institutional Demand Pressure and the Cost of Corporate Loans," *Journal of Financial Economics*, 99(3), 500–522.
- JENSEN, M. C., AND W. H. MECKLING (1976): "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3(4), 305–360.
- KAMINSKY, G. L. (2008): "Crises and Sudden Stops: Evidence from International Bond and Syndicated-Loan Markets," *NBER Working Paper 14249*.
- KEYS, B. J., T. MUKHERJEE, A. SERU, AND V. VIG (2010): "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *The Quarterly Journal of Economics*, 125(1), 307–362.
- KIYOTAKI, N., AND J. MOORE (1997): "Credit Cycles," *The Journal of Political Economy*, 105(2), 211–248.
- KRISHNAMURTHY, A., S. NAGEL, AND D. ORLOV (2014): "Sizing up Repo," *The Journal of Finance*, 69(6), 2381–2417.
- KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2012): "The Aggregate Demand for Treasury Debt," *Journal of Political Economy*, 120(2), 233–267.
- KRISHNAMURTHY, A., AND A. VISSING-JORGENSEN (2015): "The Impact of Treasury Supply on Financial Sector Lending and Stability," *Journal of Financial Economics*, forthcoming.
- KURLAT, P. (2013): "Lemons Markets and the Transmission of Aggregate Shocks," *American Economic Review*, 103(4), 1463–89.

- LORENZONI, G. (2008): “Inefficient Credit Booms,” *The Review of Economic Studies*, 75(3), 809–833.
- MENDOZA, E. G. (2010): “Sudden Stops, Financial Crises, and Leverage,” *American Economic Review*, 100, 1941–1966.
- MENDOZA, E. G., AND M. E. TERRONES (2012): “An Anatomy of Credit Booms and Their Demise,” *NBER Working Paper* 18379.
- MIAN, A., AND A. SUFI (2009): “The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis,” *The Quarterly Journal of Economics*, 124(4), 1449–1496.
- PARLOUR, C. A., AND G. PLANTIN (2008): “Loan Sales and Relationship Banking,” *The Journal of Finance*, 63(3), 1291–1314.
- PISKORSKI, T., A. SERU, AND J. WITKIN (2015): “Asset Quality Misrepresentation by Financial Intermediaries: Evidence from RMBS Market,” *Journal of Finance*, Forthcoming.
- REINHART, C., AND K. ROGOFF (2009): *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton.
- SCHULARICK, M., AND A. M. TAYLOR (2012): “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008,” *American Economic Review*, 102(2), 1029–61.
- SHIN, H. S. (2009): “Securitisation and Financial Stability,” *The Economic Journal*, 119, 309–332.
- VANASCO, V. (2014): “Information Acquisition vs. Liquidity in Financial Markets,” *Working Paper*.
- WHITE, E. N. (2009): “Lessons from the Great American Real Estate Boom and Bust of the 1920s,” *NBER Working Paper* 15573.

## A Proofs

### Proof of Proposition 3

Let assets markets be tight. First confirm that  $p$  is increasing in  $w_F$ . The optimal bank portfolio satisfies  $b_B = pa_B$ , while the asset market clearing condition is  $a_B = a_F$ . If financiers borrow, then  $b_F = Y_l a_F$ , and bond-market clearing requires  $pa_F + Y_l a_F = \frac{w_S}{q}$ . Hence  $q(p) = \min\left(\frac{pw_S}{(p+Y_l)w_F+Y_l w_S}, \frac{1}{R}\right)$ , while  $p = \frac{w_F+w_B\tilde{m}Y_l q}{\tilde{m}w_B+\tilde{m}qw_F}$  by asset-market clearing. If  $q = \frac{1}{R}$ , then  $p^* = \frac{w_F+w_B\tilde{m}Y_l \frac{1}{R}}{\tilde{m}w_B+\tilde{m}\frac{1}{R}w_F}$ , and is increasing in  $w_F$  if and only if  $w_B > \tilde{m}w_B \frac{Y_l}{R}$ . This condition always holds because  $\tilde{m} \in (0, 1)$  and  $Y_l < R$ . Moreover,  $k = \frac{w_B}{1-\tilde{m}qp}$  is increasing because  $p$  is increasing and  $q$  is a constant. Since all banks monitor, expected output increases. Now

assume that  $q < \frac{1}{R}$ . Then  $p^* = \frac{w_F - \tilde{m}Y_l w_B + \sqrt{(w_F - \tilde{m}Y_l w_B)^2 + 4\tilde{m}(w_B + w_S)Y_l(w_F + w_S)}}{2\tilde{m}(w_B + w_S)}$ . Hence  $p^*$  is increasing in  $w_F$ . By the bank's portfolio,  $k = \frac{1}{\tilde{m}}a_B$ . By market clearing,  $k = \frac{1}{\tilde{m}}a_F$ . Since  $a_F$  is strictly increasing in  $w_F$ , the result follows. Moreover, expected output is increasing because all banks monitor. Next, assume that financiers do not borrow ( $\mu = 0$ ). Then  $q^* = \min\left(\frac{w_S}{p\tilde{m}(w_B + w_S)}, \frac{1}{R}\right)$  and  $p^* = \frac{w_F}{\tilde{m}(w_B + qw_F)}$ . If  $q^* = \frac{1}{R}$ , then  $p$  is increasing in  $w_F$ . If  $q < \frac{1}{R}$ , then  $p^* = \frac{w_F}{\tilde{m}(w_S + w_B)}$  which is again increasing in  $w_F$ . Next, note that  $q^*p^* = \frac{w_S}{\tilde{m}(w_B + w_S)}$ . Hence  $k = \frac{w_B}{1 - qp\tilde{m}}$  is increasing in  $w_F$ . Next, assume that financiers are indifferent between borrowing and lending ( $\mu \in (0, 1)$ ). This requires  $p^* = \hat{Y}q^*$ , and  $b_F = \mu Y_l a_F$ . Asset-market clearing is  $\frac{\tilde{m}w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}} = \frac{w_F}{p(1 - \frac{\mu Y_l}{\hat{Y}})}$ . Suppose for a contradiction that  $p$  is decreasing in  $w_F$ . Then  $a_B = \frac{\tilde{m}w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}}$  is also decreasing in  $w_F$ . By market-clearing,  $a_F$  must be decreasing in  $w_F$ , and hence  $\mu$ ,  $b_B$  and  $b_F$  must also be decreasing. But if  $b_B$  and  $b_F$  are decreasing in  $w_F$ , then  $q$  must be increasing in  $w_F$ . This is a contradiction  $q$  decreasing because  $p^* = \hat{Y}q^*$  and  $p$  was presumed to be decreasing. Hence  $k = \frac{w_B}{1 - p^2 \frac{\tilde{m}}{\hat{Y}}}$  is increasing in  $w_F$ . Because all banks monitor, expected output is increasing in  $w_F$ .

Now assume that asset markets are slack and intermediaries are highly constrained. Then  $q^* = \frac{1}{R}$ , and  $p^* = \underline{p}(q^*)$  are independent of  $w_F$ . Hence  $a_F$  is increasing in  $w_F$  and, as a result, so is  $k$ . Next, note that  $k$  is proportional to  $w_B + \frac{q^*[\underline{p}(q^*) - Y_l]w_F}{\underline{p}(q^*) - q^*Y_l}$ . Hence  $\frac{\partial k}{\partial w_B} = \frac{\partial k}{\partial w_F}$  if  $q = 1$  and  $\frac{\partial k}{\partial w_B} < \frac{\partial k}{\partial w_F}$  if  $q > 1$ . Since intermediaries are highly constrained, this is the case when  $\underline{R} < 1$ .  $\square$

## Proof of Proposition 4

Assume that asset markets are tight. Since there is a continuum of banks, individual deviations do not affect the bond price  $q$ . Fixing  $q$ , first-order condition (13) shows that a bank weakly prefers to increase its promise at  $\underline{a}$  given  $p(\underline{a}) = \bar{p}(\bar{k}(\underline{a}), \underline{a})$  if  $u'_B(q, \bar{p}, \underline{a}) \geq 0 \Leftrightarrow -\bar{p}'(\underline{a}) \cdot \underline{a} \leq \bar{p}(\bar{k}(\underline{a}), \underline{a}) - \underline{p}(q)$ , where  $\bar{p}'(\underline{a}) = \frac{\partial \bar{p}(\bar{k}(\underline{a}), \underline{a})}{\partial \underline{a}}$  and the right-hand is positive. Moreover, we have that  $\bar{k}(\underline{a}) = \lambda_B(q) [w_B + q(\bar{p}(\bar{k}(\underline{a}), \underline{a}) - Y_l)\underline{a}]$  and  $\bar{p}(\bar{k}(\underline{a}), \underline{a}) = \hat{Y} - m \frac{\bar{k}(\underline{a})}{\bar{k}(\underline{a}) - \underline{a}}$ . Totally differentiating the latter expression and rearranging yields  $\bar{p}'(\underline{a}) = \frac{-m\lambda_B(q)w_B}{(\bar{k}(\underline{a}) - \underline{a})^2 - qm\lambda_B(q)\underline{a}^2}$ . Evaluating this expression at  $\underline{a}^* = \tilde{m}k^*$  gives

$$-\bar{p}'(\underline{a}^*)\underline{a}^* = \frac{m\lambda_B(q)(\pi_l Y_l - m)}{\lambda_B^*(q, \underline{a}^*) [(1 - \tilde{m}) - \tilde{m}q\lambda_B(q)(\pi_l Y_l - m)]},$$

where  $\lambda_B^*(q, \underline{a}^*) = [1 - q\tilde{m}\bar{p}(\bar{k}(\underline{a}^*), \underline{a}^*)]^{-1}$  and  $\bar{p}(\bar{k}(\underline{a}^*), \underline{a}^*) = \hat{y}$ . The numerator is strictly positive. The stated condition ensures that the denominator is strictly negative. What remains to be shown is that, fixing all other parameters, there exists an  $m$  sufficiently close to zero such that  $\bar{p}'(\underline{a}^*)\underline{a}^* > 0$ . Rearranging gives  $-\bar{p}'(\underline{a}^*)\underline{a}^* = \frac{\lambda_B(q)(\pi_l Y_l - m)}{\lambda_B^*(q, \underline{a}^*) \left[ \frac{1}{\pi_l Y_l} - \frac{\tilde{m}}{m} q \lambda_B(q) (\pi_l Y_l - m) \right]}$ . We have that  $\lim_{m \rightarrow 0} \tilde{m} = 1$ ,  $\lim_{m \rightarrow 0} \lambda_B(q) = \frac{1}{1 - qY_l}$  and  $\lim_{m \rightarrow 0} \lambda_B^*(q, \underline{a}^*) = \frac{1}{1 - q\hat{y}}$ . Hence

$\lim_{m \rightarrow 0} \bar{p}'(\underline{a}^*) \underline{a}^* > 0$  for any  $q$ . □

### Proof of Proposition 5

Assume first that the asset market is tight. Since  $\underline{a}^* = \tilde{m}k^*$ , it directly follows that  $p^* = \bar{p}(k^*, \underline{a}^*) = \hat{y}$  is a constant. If intermediaries are highly constrained, then  $q^* = \frac{1}{R}$ . To clear the asset market at fixed prices,  $\Phi$  must be increasing in  $w_F$ . Given that  $k$  is constant because prices are, expected output must fall. If instead  $q < \frac{1}{R}$ , market clearing conditions are  $\frac{\hat{y}\tilde{m}w_B}{1-q\hat{y}\tilde{m}} + \frac{\mu(1-\phi)Y_l w_F}{\hat{y}-(1-\phi)q\mu Y_l} = \frac{w_d}{q}$  and  $\left(\frac{\Phi+(1-\Phi)\tilde{m}}{1-q\hat{y}\tilde{m}}\right) w_B = \frac{w_F}{\hat{y}-(1-\phi)q\mu Y_l}$ . Suppose first that financiers are indifferent toward leverage ( $\mu \in (0, 1)$ ). Then  $q^* = \frac{p^*}{\hat{y}}$ , which is a constant because  $p^* = \hat{y}$ . Given that prices are fixed,  $k$  is a constant and  $\Phi$  is increasing in  $w_F$ . Hence expected output must fall. Next, suppose financiers do not borrow ( $\mu = 0$ ). Then  $q$  is independent of  $w_F$ ,  $k$  is a constant, and expected output must decline. Finally, assume that financiers are fully levered ( $\mu = 1$ ). Imposing bond market clearing reveals that  $q$  must satisfy  $q_{BM}(\Phi, w_F) = \frac{w_S}{\tilde{m}[\hat{y}(w_B+w_S)+(1-\Phi)Y_l w_F]}$ , which is strictly increasing in  $\Phi$  and strictly decreasing in  $w_F$ . The bond price that clears the asset market is  $q_{AM}(\Phi, w_F) = \frac{w_F - w_B \hat{y}(\Phi + (1-\Phi)\tilde{m})}{\tilde{m}[(w_B + w_F)Y_l + w_B Y_l \Phi]}$ , which is strictly increasing in  $w_F$  but strictly decreasing in  $\Phi$ . Suppose for a contradiction that  $\Phi$  is decreasing in  $w_F$ . By bond market clearing, an increase in  $w_F$  and a decrease in  $\Phi$  leads to fall in  $q$ . But by asset market clearing, an increase in  $w_F$  and a decrease in  $\Phi$  leads to an *increase* in  $q$ . Hence both markets cannot not clear simultaneously.

Now assume that the asset market is slack, and note that the  $\underline{p}(q)$  can be expressed as  $\underline{p}(q) = \frac{\hat{Y} - Y_l + \hat{Y}(1-\tilde{m})Y_l q}{q(\hat{Y} - Y_l \tilde{m})}$ . Using this expression reveals that  $k = k^*(\underline{a}) = \frac{w_B}{1-qY_l \tilde{m}} + \left(\frac{\hat{Y} - Y_l}{\hat{Y} - Y_l \tilde{m}}\right) \underline{a}$ . I will first show that whenever the financial system is highly constrained, or financiers weakly prefer to not borrow, then  $q^*$ ,  $p^*$ ,  $k$  and  $\underline{a}$  are all invariant to  $w_F$ . Suppose first that intermediaries are highly constrained. Then  $q^* = \frac{1}{R}$  and  $p^* = \underline{p}(q^*)$  are constants. Moreover,  $p^* = \bar{p}(k^*(\underline{a}), \underline{a}^*)$  and  $k^*(\underline{a}) = \frac{w_B}{1-qY_l \tilde{m}} + \left(\frac{\hat{Y} - Y_l}{\hat{Y} - Y_l \tilde{m}}\right) \underline{a}$ . Hence  $k$  and  $\underline{a}$  are invariant to  $w_F$ . Now suppose that financiers do not borrow. The bond market-clearing condition is  $b_B = \frac{Y_l \tilde{m} w_B + (p - Y_l) \underline{a}}{1 - Y_l q^* \tilde{m}} = \frac{w_S}{q}$ . Hence  $\underline{a}$  is fixed conditional on  $q$ . Since  $p^* = \bar{p} = \hat{Y} - m \frac{k}{k - \underline{a}}$  and  $k$  is fixed once  $q$  is determined, the condition  $p = \underline{p}(q)$  suffices to pin down  $q$ ,  $p$ ,  $k$  and  $\underline{a}$  independently of  $w_F$ . In either case,  $q^*$ ,  $p^*$ ,  $k$  and  $\underline{a}$  are invariant to  $w_F$ , and  $\Phi$  must increase in  $w_F$  to clear asset markets. Since  $k$  is invariant to  $w_F$ , expected output must decline.

Finally, assume that financiers are fully levered and that the financial system is not highly constrained. Then the market clearing conditions are  $\frac{Y_l \tilde{m} w_B + (p - Y_l) \underline{a}}{1 - q Y_l \tilde{m}} + \frac{Y_l w_F}{p - (1 - \phi) q Y_l} = \frac{w_S}{q}$  and  $\Phi \left(\frac{w_B + q(p - Y_l) \underline{a}}{1 - q Y_l \tilde{m}}\right) + (1 - \Phi) \underline{a} = \frac{w_F}{p - (1 - \phi) q Y_l}$ , where  $p^* = \underline{p}(q)$ . Recall from above that  $k = k^*(\underline{a}) = \frac{w_B}{1 - q Y_l \tilde{m}} + \left(\frac{\hat{Y} - Y_l}{\hat{Y} - Y_l \tilde{m}}\right) \underline{a}$ . It follows that  $k$  is a function of  $\underline{a}$  and  $q$  only, while  $\frac{\partial k}{\partial \underline{a}}$  is

independent of  $q$ . Moreover, it is easy to check that  $p$  and  $q$  are fixed for a given  $\underline{a}$ , while  $\bar{p}(k^*(\underline{a}), \underline{a})$  is strictly *decreasing* in  $\underline{a}$  (note that this is not the case when asset markets are tight, since the marginal effect of  $\underline{a}$  on  $k$  is then not independent of market prices). We can then show that  $\Phi$  must be strictly increasing in  $w_F$ . Suppose for a contradiction that  $\Phi$  is weakly decreasing. For the bond market to clear, either  $q$  and/or  $\underline{a}$  must decrease. Since  $p = \underline{p}$ ,  $p$  must increase if  $q$  falls. Since  $p = \bar{p}$  and  $\bar{p}$  is strictly decreasing in  $\underline{a}$ , it follows that  $q$  and  $\underline{a}$  must both decrease. Since  $b_B$  and  $q$  are decreasing, it follows that  $k$  must decrease. Since  $k$ ,  $\underline{a}$ , and  $\Phi$  are all decreasing, total asset market supply must decrease. Yet  $a_F$  is weakly increasing. Hence asset markets cannot clear, yielding a contradiction.

Finally, we want to show that there are Pareto-improving financier wealth reductions. By the preceding arguments,  $q^*$  and  $k^*$  are invariant to  $w_F$  under the stated conditions, and so saver utility must be as well. If asset markets are tight, then bank utility is  $(1 - \tilde{m})k^*$ , which is constant. Financier utility is  $v_F = (\Phi\hat{y} + (1 - \Phi)\tilde{m}\hat{Y})k^*$  if financiers issue bonds, and  $v_F = (\Phi\hat{y} + (1 - \Phi)\tilde{m}(\hat{Y} - Y_l))k^*$  if they do not issue bonds. Hence  $v_F$  is strictly decreasing in  $w_F$ . If asset markets are slack, then bank utility is  $v_B = \frac{\hat{Y} - \tilde{m}Y_l}{1 - q^*Y_l\tilde{m}}$ . Since  $q^*$  is invariant to  $w_F$ , so is  $v_B$ . Financier utility is  $v_F = (\phi\hat{Y}' + (1 - \phi)(\hat{Y} - Y_l))a_B^*$  if financiers issue bonds and  $v_F = (\phi\hat{Y}' + (1 - \phi)\hat{Y})a_B^*$  when they do not. By market clearing,  $\Phi a_B^* + (1 - \Phi)k^* = a_F = \frac{w_F}{p^* - (1 - \phi)Y_l q'}$ , where  $a_B^*$  is a constant. Hence  $v_F = \Phi\hat{y}k^* + (1 - \Phi)\hat{Y}a_B^*$  when they do not borrow, and  $v_F = \Phi\hat{y}k^* + (1 - \Phi)(\hat{Y} - Y_l)a_B^*$  when they do. Since  $k^*$  and  $a_B^*$  are constants and  $\Phi$  is strictly increasing in  $w_F$ ,  $v_F$  is strictly decreasing in  $w_F$ .  $\square$

### Proof of Proposition 6

If  $q_t$  is such that financiers do not issue bonds, then  $a_{F,t} = \frac{w_{F,t}}{p_t}$  and  $a_{B,t} = \tilde{m}\frac{w_{B,t}}{1 - q_t\tilde{m}p_t}$ , while  $w_{F,t+1} = Y_z a_{F,t}$  and  $w_{B,t+1} = (1 - \tilde{m})Y_z k_t = \frac{1 - \tilde{m}}{\tilde{m}}Y_z a_{F,t}$ . By market clearing,  $a_{F,t} = a_{B,t}$ . Hence  $\omega_{t+1} = \bar{\omega} = \frac{\tilde{m}}{1 - \tilde{m}}$ .  $q_{t+1}^*(h) \leq q_t$  because  $w_F$  and  $w_B$  increase if  $z = h$ . Hence, financiers also do not issue bonds in period  $t + 1$ . By market-clearing,  $p_\tau = \frac{\omega_\tau}{\tilde{m}(1 + q_\tau\omega_\tau)}$  for  $\tau \in \{t, t + 1\}$ , and  $p_\tau$  is strictly increasing in  $\omega_\tau$  and strictly decreasing in  $q_\tau$ . Since  $q_{t+1} \leq q_t$  and  $\omega_{t+1} \geq \omega_t$  after a good shock given that  $\omega_t \leq \bar{\omega}$  by assumption, we must have  $p_{t+1}^*(h) \geq p_t^*$ .

If  $q_t$  is such that financiers issue bonds, then  $a_{F,t} = \frac{w_{F,t}}{p_t - q_t Y_l}$  and  $a_{B,t} = \tilde{m}\frac{w_{B,t}}{1 - q_t\tilde{m}p_t}$ , while  $w_{F,t+1} = (Y_z - Y_l)a_{F,t}$  and  $w_{B,t+1} = (1 - \tilde{m})Y_z k_t = \frac{1 - \tilde{m}}{\tilde{m}}Y_z a_{F,t}$ . By market clearing,  $a_{F,t} = a_{B,t}$  and thus  $\omega_{t+1} = \bar{\omega} = \frac{\tilde{m}}{1 - \tilde{m}}\frac{Y_z - Y_l}{Y_z}$ .  $q_{t+1}^*(h) \leq q_t$  because  $w_F$  and  $w_B$  increase if  $z = h$ . Because financiers are assumed to issue bonds in period  $t + 1$ ,  $p_\tau = \frac{\omega_\tau + \tilde{m}q_\tau Y_l}{\tilde{m}(1 + q_\tau\omega_\tau)}$  for  $\tau \in \{t, t + 1\}$ . Hence  $p_\tau$  is strictly increasing in  $\omega_\tau$ , and is strictly decreasing in  $q_\tau$  if  $\omega_\tau > \sqrt{\pi_l Y_l}$ . Since  $q_{t+1}^*(h) \leq q_t$  and  $\omega_{t+1}(h) \geq \omega_t$  given that  $\omega_t \leq \omega_{t+1}(h) = \underline{\omega}(h)$  by assumption,  $p_{t+1}^* \geq p_t$  if  $\underline{\omega}(h) > \sqrt{\pi_l Y_l}$ . Since  $\lim_{m \rightarrow 0} \underline{\omega}(h) = \infty$  because  $\lim_{m \rightarrow 0} \tilde{m}_{m \rightarrow 0}$ , there always exists an  $m$  sufficiently close to zero such that  $\underline{\omega}(h) > \sqrt{\pi_l Y_l}$ .

## Proof of Proposition 7

The law of motion for  $\omega$  follows directly from intermediary portfolios, noting that  $k = \frac{w_B}{1-q\hat{Y}_i\tilde{m}} + \left(\frac{\hat{Y}-Y_i}{\hat{Y}-Y_i\tilde{m}}\right)\underline{a}$  if  $p = \underline{p}(q)$  (see proof of Proposition 5). The difference in *expected* returns on equity is  $\mathbb{E}[r_F(z) - r_B(z)] = \frac{\hat{Y}-Y_i}{\underline{p}(q)-qY_i} - \frac{\hat{Y}-\tilde{m}Y_i}{1-q\tilde{m}Y_i}$ . Hence  $\mathbb{E}[r_F(z) - r_B(z)] \geq 0$  if  $\underline{p}(q) \leq \frac{\hat{Y}-Y_i+Y_i(1-\tilde{m})\hat{Y}q}{\hat{Y}-\tilde{m}Y_i} = q\underline{p}(q)$ . This implies that  $\frac{\hat{Y}-Y_i}{\underline{p}(q)-qY_i} = \frac{\hat{Y}-\tilde{m}Y_i}{1-q\tilde{m}Y_i}$  when  $q = 1$ , and so  $\frac{Y_h-Y_i}{\underline{p}(q)-qY_i} > \frac{Y_h-\tilde{m}Y_i}{1-q\tilde{m}Y_i}$  when  $q = 1$ . Moreover,  $p_{t+1}(h) \geq p_t$  because  $\underline{p}(q)$  is strictly decreasing in  $q$  and  $q$  is decreasing in  $w_B$  and  $w_F$ .

## Proof of Proposition 8

By construction. Choose  $w_S$  large enough and  $w_F$  small enough such that asset markets are slack and intermediaries are highly constrained, and note that such parameters always exist. Then  $q^* = \frac{1}{\underline{R}}$  and  $p^* = \underline{p}(q^*)$ . Let  $\lambda_F = \frac{1}{p^*-q^*Y_i}$  and  $\lambda_B = \frac{1}{1-q^*\tilde{m}Y_i}$  denotes the equilibrium leverage of financiers and banks in a efficient monitoring equilibrium, respectively. By the proof of Proposition 5, we then have that  $k^* = \lambda_B w_B + \chi a_F^*$  where  $\chi = \left(\frac{\hat{Y}-Y_i}{\hat{Y}-\tilde{m}Y_i}\right) \in (0, 1)$  and  $a_F^* = \lambda_F w_F$ . Hence, the upper bound on the asset price is  $\bar{p}(\omega) = \hat{Y} - m \left(\frac{\lambda_B + \chi \lambda_F \omega}{\lambda_B - (1-\chi)\lambda_F \omega}\right)$ . Note that  $\bar{p}(0) = \hat{Y} - m$  and  $\frac{\partial \bar{p}}{\partial \omega} < 0$ . It follows that as long as  $p^* < \hat{Y} - m$  there exists, for small enough  $w_F$ , an efficient monitoring equilibrium with slack asset markets in which the financial system is highly constrained, and  $p^* = \underline{p}(q^*) < \hat{Y} - m$  if  $\underline{R} < \frac{1}{\chi}(\hat{Y} - m) - \frac{(1-\tilde{m})Y_i\hat{Y}}{\hat{Y}-Y_i}$ . Next, note that  $\bar{p} \geq \hat{y}$  because  $a_B \leq \tilde{m}k$ . For an excessive trading equilibrium to exist for sufficiently large  $w_F$ , we require that  $p^* \geq \hat{y}$ , which is the case if  $\underline{R} > \frac{1}{\chi}\hat{y} - \frac{(1-\tilde{m})Y_i\hat{Y}}{\hat{Y}-Y_i}$ . Note that both parametric conditions are jointly satisfied if we set  $\hat{y} = \underline{R} - \epsilon$  for  $\epsilon$  and  $m$  sufficiently small. Hence there exist parameters such that  $p \in [\hat{y}, \hat{Y} - m)$ . Assume such parameters, and choose initial financier wealth such that the economy is initially in efficient monitoring equilibrium. To construct a credit boom with falling asset quality, note that, because intermediary wealth is bounded after any finite sequence of good aggregate shocks, there always exists a level of depositor wealth such that the financial system is highly constrained at all times. Assume this to be the case, so that  $q^* = \frac{1}{\underline{R}} \geq 1$  and  $p = \underline{p}(q^*)$  throughout. By Proposition 7,  $\omega_{t+1}^*(h) > \omega_t$  for any  $\omega_t$ . We have already shown above that a sufficiently long sequence of good aggregate shocks must trigger a excessive trading equilibrium under the given parametric conditions. Recall that in efficient monitoring equilibrium with slack asset markets,  $\mathbb{E}_z r_F(z) > \mathbb{E}_z r_B(z)$  for  $q > 1$ . Accordingly, there exists a  $\Phi > 0$  sufficiently small such that  $\mathbb{E}_z r_F(z) > \mathbb{E}_z r_B(z)$  in excessive trading equilibrium if  $q^* = \frac{1}{\underline{R}}$  and  $\underline{R} < 1$ .  $\square$

## Proof of Proposition 9

In the absence of leverage caps,  $a_B^* = \tilde{m}k^*$ . With a binding leverage cap, banks withdraw assets from secondary markets until the leverage cap just binds. Summarize the degree to which banks exhaust their borrowing capacity by  $\mu_B \in [0, 1]$ , so that  $a_B = \mu_B \left( \frac{b_B}{p} \right)$ . The optimal portfolio bank portfolio then is  $k = \left[ \frac{(1-\mu_B)p + \mu_B Y_l}{(1-\mu_B)p + \mu_B Y_l - pqY_l \tilde{m}} \right] w_B$  and  $b_b = \left[ \frac{pY_l \tilde{m}}{(1-\mu_B)p + \mu_B Y_l - pqY_l \tilde{m}} \right] w_B$ , and bank leverage fixing  $\mu_B$  is  $\lambda_B(\mu) = \frac{(1-\mu_B)p + \mu_B Y_l}{(1-\mu_B)p + \mu_B Y_l - pqY_l \tilde{m}}$ . Setting  $\lambda_B(\mu) = \bar{\lambda}_b$  shows that  $\mu_B^*(\bar{\lambda}_b) = \left( \frac{p}{p-Y_l} \right) \left[ 1 - \left( \frac{\bar{\lambda}_b}{\lambda_b - 1} \right) qY_l \tilde{m} \right] \in (0, 1)$ . Hence  $a_B$  is increasing in  $\bar{\lambda}_b$ , but *decreasing* in the asset price  $p$ . Now suppose for a contradiction that all banks monitor. Since all banks monitor,  $p^* < \hat{Y}$ . Since the financial system is highly constrained,  $q^* = 1$  with and without leverage constraints. Moreover,  $a_F = \frac{w_F}{p^* - Y_l}$  while  $a_B = \frac{w_B(\bar{\lambda}_b(1-Y_l \tilde{m}))}{p^* - Y_l}$ . Since secondary markets clear in the absence of leverage caps,  $a_F > a_B$  for any  $p^*$  if  $\bar{\lambda}_B < \lambda_B^*$ . Since  $p^* < \hat{Y}$ , there is excess demand if  $\Phi = 0$ .  $\square$

## B Dynamic Model with Endogenous Risk Aversion

I now consider a variant of dynamic model in which old intermediaries have full bargaining power, which implies that the objective of a young intermediary is to maximize the expected *value* of end-of-life wealth. The value of wealth is generically state-contingent, with intermediaries valuing a dollar of equity more highly in states of the world where aggregate intermediary wealth is low so that intermediation rents are large. Because intermediary wealth depends on the realization of aggregate risk, forward-looking behavior can thus lead to endogenous hedging motives.

The section has two goals. The first is to show how endogenous risk aversion shapes the supply and demand for risky assets and impacts banks' incentives to produce high-quality assets. The second is to show that the forces that drive credit booms with falling investment efficiency in the baseline dynamic model are present even with endogenous risk aversion. Both goals can be achieved most transparently in a finite horizon setting, since I can then characterize value of equity capital in the final period in closed form and study optimal portfolios in the next-to-final period.<sup>13</sup>

I denote the wealth distribution at the beginning of period  $t$  by  $\mathbf{w}_t = (w_{S,t}, w_{B,t}, w_{F,t})$ , and its endogenously determined law of motion by  $\Gamma(\cdot)$ . Given the aggregate shock  $z_t$ , we have that  $\mathbf{w}_{t+1} = \Gamma(z_t, \mathbf{w}_t)$ .

<sup>13</sup> In an infinite-horizon economy, instead, the value of wealth at a given date would depend on the entire expected future path of the wealth distribution and on the expected transitions to and from regions of the state space where financiers are excessively wealthy and/or asset markets are slack. Both features of the model make it difficult to analytically characterize the impact of endogenous risk aversion on volume of credit and bank incentives.

**Proposition 10.**

The final-period value attained by an intermediary of type  $\theta$  with wealth  $w$  when the wealth distribution is  $\mathbf{w}$  can be written as  $V_\theta(w, \mathbf{w}) = v_\theta(\mathbf{w})w$ .

The result follows because all policy functions in the static game are linear in  $w$  and intermediaries are risk-neutral. Intermediaries thus exhibit endogenous risk aversion only to the extent that aggregate risk exposure generates variation in intermediation rents across states. Intermediaries thus evaluate cash flows using  $v_\theta(\cdot)$ . To simplify notation while highlighting the dependence on the aggregate shock, I write  $v_{\theta,z} = v_\theta(\Gamma(z, \mathbf{w}))$ . Accordingly, an intermediary of type  $\theta$  assigns the expected values

$$\tilde{Y}_\theta = \pi_h v_{\theta,h} Y_h + \pi_l v_{\theta,l} Y_l, \quad \tilde{y}_\theta = \pi_h v_{\theta,h} y_h + \pi_l v_{\theta,l} y_l \quad \text{and} \quad \tilde{v}_\theta = \pi_h v_{\theta,h} + \pi_l v_{\theta,l}$$

to a high-quality asset, a low-quality asset, and a safe cash flow, respectively. This means that intermediary  $\theta$  is willing to give up a unit of a safe cash flow in exchange for a unit of the risky asset if  $\tilde{Y}_\theta \geq \tilde{v}_\theta$ . Letting  $\gamma_\theta = \frac{v_{\theta,l}}{v_{\theta,h}}$  denote the endogenous risk aversion of intermediary type  $\theta$ , we have that  $\frac{\tilde{Y}_F}{\tilde{v}_F} \leq \frac{\tilde{Y}_B}{\tilde{v}_B}$  if and only if  $\gamma_F \leq \gamma_B$ . Endogenous risk aversion  $\gamma_\theta$  thus determines the efficient holder of the risky asset in the absence of collateral constraints. Note that individuals take the state-contingent value of wealth  $v_\theta(\cdot)$  as given because it depends only on the aggregate wealth distribution.

Given the linearity of the value functions, the arguments in the main text directly imply the following results.

**Observation 2** (Bank Portfolio with Endogenous Risk Aversion).

Given risk aversion, the moral hazard discount is  $\tilde{m}' = 1 - \frac{m}{\pi_l v_{B,l} Y_l'}$ , bank leverage is  $\tilde{\lambda}_B(q) = \frac{1}{1 - q\tilde{m}'Y_l'}$ , and the leveraged intermediation premium is  $\tilde{\rho}(q) = \tilde{\lambda}_B(q)(\tilde{Y}q - 1)$ . Moreover:

1. The bank's borrowing capacity is  $\bar{b}_B(k, \underline{a}) = \tilde{m}'Y_l k + (p(\underline{a}) - Y_l)\underline{a}$ .
2. The bank's skin-in-the-game constraint is  $\underline{a} \leq \bar{a}(k) = \tilde{m}'k$ .
3. The lower bound on the asset price is  $\underline{p}(q) = \frac{\tilde{Y} + \tilde{\rho}(q)Y_l}{1 + \tilde{\rho}(q)}$ .
4. The upper bound on the asset price is  $\bar{p}(k, \underline{a}) \equiv \frac{\tilde{Y}_B}{\tilde{v}_B} - \frac{m}{\tilde{v}_B} \left( \frac{k}{k - \underline{a}} \right)$ .

The next proposition shows that such increases in bank risk aversion, as proxied by mean-preserving spreads of  $v_{B,z}$  such that  $v_{B,l}$  increases but  $\tilde{v}_B$  remains constant, tempt banks into shirking and selling earlier, while increasing leverage and relaxing the skin-in-the-game constraint.

**Proposition 11.**

Consider a mean-preserving increase in bank risk aversion. Then  $\bar{b}(k, \underline{a})$  and  $\bar{a}(k)$  are strictly increasing, while  $\bar{p}(k, \underline{a})$  is strictly decreasing.

The proof is straightforward. Increases in  $v_{B,l}$  at the expense of  $v_{B,h}$  reduce the expected value of the risky asset, while  $\tilde{m}'$  is strictly increasing in  $v_{B,l}$  because banks who value wealth in the low state are less tempted to reduce their low-state cash-flows by shirking. Bank risk aversion thus *strengthens* the scope for risk transfer and shirking by banks. The next observation shows how risk aversion shapes the financier portfolio.

**Observation 3** (Financier Portfolio with Endogenous Risk Aversion).

*Financiers prefer to buy risky assets rather than invest in storage if  $p \leq \frac{\tilde{Y}_F}{\tilde{v}_F}$ . They strictly prefer to issue bonds to do so if  $p < q \frac{\tilde{Y}_F}{\tilde{v}_F}$  and are indifferent if the condition holds with equality.*

Naturally, increases in risk aversion temper financiers' desire to purchase risky assets. Importantly, however, this does not imply that financiers never demand enough assets to trigger the shirking channel. The reason is that, with cash-in-the-market pricing, the equilibrium asset price is typically *not* equal to financiers willingness-to-pay. Financier risk aversion thus prevents the equilibrium price from reaching the threshold  $\bar{p}$  only if  $\frac{\tilde{Y}_F}{\tilde{v}_F} \leq \bar{p}$ . As I now show, this condition fails to hold in under mild conditions even when intermediaries choose the same portfolios as in the static model. As a result, financiers continue to have a strict preference for purchasing risky assets even when they take on the same level of risk exposure as in the static model. This means that there is scope for equilibrium shirking even in the presence of endogenous risk aversion. For simplicity, I assume that  $\underline{R} = 1$ , so that  $q \leq 1$  and financiers and banks earn the same expected return on equity if assets market are slack and financial intermediaries are highly constrained.

**Proposition 12.**

*Let  $\underline{R} = 1$  and assume that  $\omega_{T-1}$  is such that asset markets are tight in period  $T - 1$  if all intermediaries choose the same portfolios as in the static model. Then  $\frac{\tilde{Y}_F}{\tilde{v}_F} > \bar{p}$  if intermediaries choose the same portfolio as in the static model.*

*Proof.* Given that all intermediaries are assumed to choose the same portfolio as in the static model and asset markets are tight,  $\underline{a}^* = \tilde{m}k^*$ . Then  $\bar{p}(k^*, \tilde{m}k^*) = \frac{\pi_h v_{B,h} Y_h}{\tilde{v}_B}$ , and  $\frac{\tilde{Y}_F}{\tilde{v}_F} > \bar{p}$  if and only if  $Y_l > \pi_h Y_h \left[ 1 - \frac{\gamma_B}{\gamma_F} \right]$ , where  $\frac{\gamma_B}{\gamma_F} = \frac{\frac{v_{B,l}}{v_{F,l}}}{\frac{v_{B,h}}{v_{F,h}}} = \frac{\frac{v_{B,h}}{v_{F,l}}}{\frac{v_{B,l}}{v_{F,h}}}$ . This condition is tightest when  $\frac{\gamma_B}{\gamma_F}$  is small. Hence, it is tightest when  $\frac{v_{F,h}}{v_{B,h}}$  is small (financiers earn relatively low returns following a good shock) and  $\frac{v_{F,l}}{v_{B,l}}$  is large (financiers earn relatively high returns following a bad shock). Financiers earn relatively low returns when asset markets are tight, and relatively high returns when asset markets are slack. Hence the condition is tightest when asset markets are tight after a good shock and slack after a bad shock. Assume this is the case. The next step is to derive the the relative returns on equity when asset markets are tight and slack, respectively. If asset markets are tight, then  $p^*(q, \omega) = \frac{1}{\tilde{m}} \left[ \frac{\tilde{m}qY_l + \omega}{1 + q\omega} \right]$ . Hence

the returns on equity are  $v_b = \frac{(1-\tilde{m})\hat{Y}}{1-p^*(q,\omega)\tilde{m}q} = (1-\tilde{m})\hat{Y} \left[ \frac{1+q\omega}{1-\tilde{m}q^2Y_l} \right]$  and  $v_F = \frac{\hat{Y}-Y_l}{p^*(q,\omega)-qY_l} = \frac{\tilde{m}(\hat{Y}-Y_l)}{\omega} \left[ \frac{1+q\omega}{1-\tilde{m}q^2Y_l} \right]$ . Accordingly, the relative return on equity when asset markets are tight is  $\frac{v_F}{v_B}(\omega) = \left(\frac{1}{\omega}\right) \frac{\tilde{m}(\hat{Y}-Y_l)}{(1-\tilde{m})\hat{Y}}$ . Next, turn to the case when asset markets are slack. Then  $v_b(q) = \frac{\hat{Y}-\tilde{m}Y_l}{1-q\tilde{m}Y_l}$  and  $v_F(q) = \frac{\hat{Y}-Y_l}{\underline{p}(q)-qY_l}$ . Accordingly, the relative return is  $\frac{v_F}{v_B} = \left(\frac{\hat{Y}-Y_l}{\hat{Y}-\tilde{m}Y_l}\right) \left(\frac{1-q\tilde{m}Y_l}{\underline{p}(q)-qY_l}\right)$ , and is strictly increasing in  $q$ . Since  $q \leq 1$  and  $v_F(1) = v_B(1)$  when asset markets are slack, then  $\frac{v_F}{v_B} \leq 1$  when asset markets are slack. It follows that  $\frac{\gamma_B}{\gamma_F} \geq \frac{v_{F,h}}{v_{B,h}} = \left(\frac{1}{\omega}\right) \frac{\tilde{m}(\hat{Y}-Y_l)}{(1-\tilde{m})\hat{Y}}$ . Using this lower bound yields the sufficient condition  $Y_l > \pi_h Y_h \left[ 1 - \left(\frac{1}{\omega_T(h)}\right) \frac{\tilde{m}(\hat{Y}-Y_l)}{(1-\tilde{m})\hat{Y}} \right]$ . If banks choose the same portfolio period as in the static model in period  $T-1$ , then Proposition 6 shows that  $\omega_T(h) = \frac{\tilde{m}}{1-\tilde{m}} \left(\frac{Y_h}{Y_h-Y_l}\right)$ . Using this expression yields the condition  $\hat{Y}(Y_h - Y_l) > \pi_h Y_h (Y_h - \hat{Y})$ , which always holds.  $\square$

Finally, I present an example where the economy transitions into the shirking region even with endogenous risk aversion. Note that both banks and financiers choose the same portfolio as in the static model if  $\gamma_B = \gamma_F = 1$ .

**Proposition 13.**

*Assume that  $\underline{R} = 1$ . If the competitive equilibrium in the final period is an efficient monitoring equilibrium with slack asset markets and a highly constrained financial system after any shock, then  $\gamma_F = \gamma_B = 1$ .*

*Proof.* In an equilibrium with slack asset markets in which the financial system is highly constrained we have  $q^* = \frac{1}{\underline{R}} = 1$  and  $p^* = \underline{p}(q^*)$ . By Proposition 12, the returns on equity earned by financiers and banks, respectively, are  $v_F = \frac{\hat{Y}-Y_l}{\underline{p}(1)-Y_l}$  and  $v_B = \frac{\hat{Y}-Y_l\tilde{m}}{1-Y_l\tilde{m}}$ , and  $v_B = v_F$ . Given that the financial system is highly constrained after any shock, the result follows.  $\square$

Proposition 8 provides an example of a credit boom with falling asset quality in which intermediaries are highly constrained and asset markets are slack. Proposition 13 shows that intermediaries must choose the same portfolio as in that example when the equilibrium is an efficient monitoring equilibrium. What remains to be shown is that the economy also transitions into a excessive trading equilibrium after a sequence of good shocks.

**Proposition 14.**

*Assume that  $\underline{R} = 1$ . If the competitive equilibrium in the final period is an efficient monitoring equilibrium with slack asset markets and a highly constrained financial system after a bad shock, and an excessive trading equilibrium with slack asset markets and a highly constrained financial system after a good shock, then  $\gamma_B = 1$  and  $\gamma_F(\phi) = \frac{\phi\hat{y}+(1-\phi)(\hat{Y}-Y_l)}{\underline{p}(1)-(1-\phi)Y_l} \frac{\hat{Y}-Y_l}{\underline{p}(1)-Y_l}$ , where where  $\phi$  denotes the*

*fraction of low-quality assets traded tomorrow conditional on a good shock. Moreover,  $\gamma_F(0) = 1$  and  $\frac{\partial \gamma_F}{\partial \phi} < 0$ .*

*Proof.* For banks, the result follows from the fact that return on equity is independent of  $\phi$  by construction. For financiers, the result follows from the optimal portfolio.  $\square$

The proposition implies that the economy with endogenous risk aversion must transition into an excessive trading equilibrium if the economy without endogenous risk aversion does. To see why, suppose that the economy with endogenous risk aversion does *not* transition into a excessive trading equilibrium after a good shock, while the economy without endogenous risk aversion does. Then  $\phi = 0$  after a good shock. By Proposition 14, then  $\gamma_F = \gamma_B = 1$ , and intermediaries choose the same portfolio as in the static model. But this implies economy must transition into a excessive trading equilibrium, yielding a contradiction. Note that  $\underline{p}$  is unaffected by risk aversion since  $\gamma_B = 1$  throughout. Moreover, financiers are willing to buy risky assets when  $\phi$  is sufficiently small tomorrow because they receive strictly positive rents from doing so when  $\gamma_F = 1$ . The intuition behind the result is that individual intermediaries take the evolution of the aggregate wealth distribution as given when making their portfolio choice.

**Online Appendix. Not for Publication**

## C Robustness to Alternative Moral Hazard Specification

Consider an alternative specification of the bank monitoring decision in which banks can choose to exert effort at the level of individual assets rather than at the level of the portfolio. Fixing  $k$  and  $\underline{a}$ , let  $L \leq k$  denote the number of low-quality assets, and let  $mL$  denote the associated private benefit. In line with the data, assume that secondary markets are organized such that the bank offers up its portfolio of assets  $k$  to financiers, and financiers can choose which  $\underline{a}$  out of  $k$  assets they want to purchase. Since financiers are uninformed about the quality of the assets, assume they employ a random selection rule. Since  $k$  is infinitely divisible into individual assets, financiers then receive a portfolio with a fraction  $\frac{L}{k}$  of low-quality assets. Similarly, a fraction  $\frac{L}{k}$  of the assets retained by the bank are low-quality, and the remainder is of high-quality. Given this structure, the bank's optimal shirking decision is

$$L^* = \arg \max_{0 \leq L \leq k} \sum_z \pi_z \left[ \max \left\{ Y_z(k - \underline{a}) - (Y_z - y_z)L \left( \frac{k - \underline{a}}{k} \right) - b_B + p(\underline{a})\underline{a}, 0 \right\} \right] + mL.$$

Assuming that the limited-liability constraint does not bind in any state of the world, the derivative with respect to  $L$  is  $-(\hat{Y} - \hat{y}) \left( \frac{k - \underline{a}}{k} \right) + m$ , and is independent of  $L$ . It follows that the bank does not shirk on any asset ( $L^* = 0$ ) if  $\underline{a} \leq \left( 1 - \frac{m}{\hat{Y} - \hat{y}} \right) k$ , and shirks on all assets ( $L^* = k$ ) otherwise. Under Assumption 2 and letting  $y_l = 0$  as in the main text, we have that  $\hat{Y} - \hat{y} = \pi_l Y_l$ , so that the condition can be equivalently stated as  $\underline{a} \leq \tilde{m}k$ , precisely as in the skin-in-the-game constraint (11). Hence the kink at which the limited-liability constraint begins binding when banks can shirk at the asset level is the same as when they must shirk at the portfolio level. Next, assume that the limited-liability constraint binds in the low state. Then the derivative with respect to  $L$  is  $-\pi_h(Y_h - y_h) \left( \frac{k - \underline{a}}{k} \right) + m$ , which is again independent of  $L$ . Banks will thus either choose to shirk on all assets or not at all, and they strictly prefer to shirk on all assets under assumption (2) whenever the limited-liability constraint binds, precisely as in the main text. Conditional on letting financiers use a random selection rule, there thus is no loss of generality in assuming that the bank either shirks on all assets or on none.

## D Global Deviations from $\underline{a}^*$ in excessive trading equilibrium.

Figure 11 depicts bank utility and the maximum asset price the bank can receive by deviating from an asset sale promise  $\underline{a}^* = \tilde{m}k^*$ , assuming that  $p = \bar{p}(k(\underline{a}), \underline{a})$  after any deviation. It shows numerically that there are no profitable global deviations. Parameter

values satisfy the sufficient condition for the absence of a profitable local deviation stated in Proposition 4. The flat part of the schedules depicts the region where the upper bound  $\bar{p}$  is equal to the lower bound  $\underline{p}(q)$ .

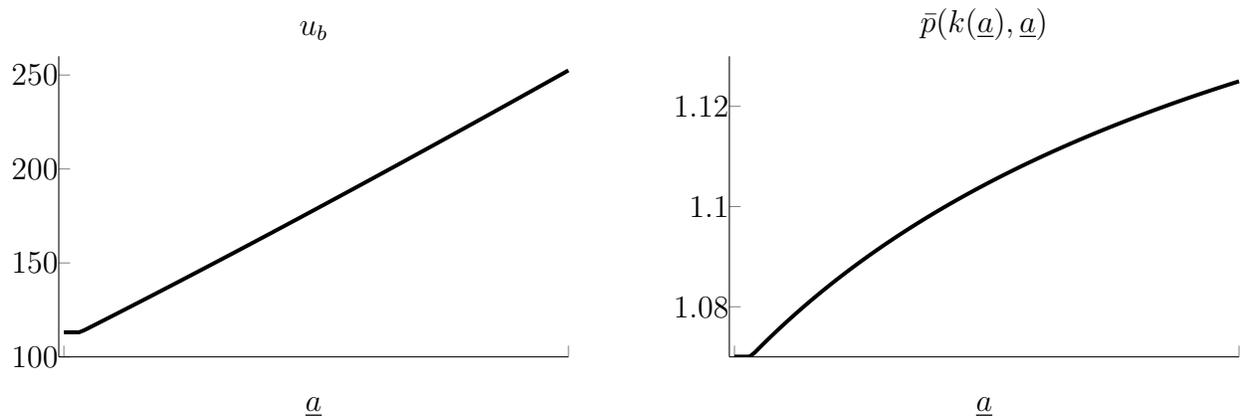


Figure 11: Bank utility and the price bound as the asset sale promise  $\underline{a}$  varies from  $\frac{\underline{a}^*}{2}$  to  $\underline{a}^*$ , with  $\underline{a}^* = \tilde{m}k^*$ . Parameter values:  $\pi_h = 0.5$ ,  $Y_l = 0.5$ ,  $\hat{Y} = 2.25$ ,  $\underline{R} = 1$ ,  $\tilde{m} = 0.025$ ,  $w_B = 70$ . Bond price fixed at  $q^* = 0.95$ .