Macroeconomic Effects of Secondary Market Trading*

Daniel Neuhann†

December 30, 2017

Abstract

I provide a theoretical framework to analyze the evolution of private investment incentives over the credit cycle. Endogenous movements in the wealth distribution generate run-ups in asset prices that lead to credit booms in which banks increasingly engage in moral hazard. Equilibrium outcomes are inefficient due to a pecuniary externality that leads to excess demand for financial assets, and credit cycles may be triggered by saving gluts or expansionary monetary policy. Bank-specific capital requirements may exacerbate the inefficiency, but regulating non-bank financial institutions can be welfare-enhancing.

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 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, Gregor Matvos, Enrique Mendoza, Stephen Morris, Christian Opp, Guillaume Plantin, Francisco Silva, and various seminar audiences 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.

†McCombs School of Business, University of Texas at Austin; daniel.neuhann@mccombs.utexas.edu.

1 Introduction

This article is a theoretical study of credit cycles. It is motivated by two empirical observations. First, credit growth is the best predictor of the likelihood of financial crises, and financial crises are typically preceded by credit booms (Jorda, Schularick, and Taylor (2011)). Second, recent evidence suggests that capital is increasingly allocated to inefficient investment opportunities during credit booms.[1] One reading of these observations is that financial crises are partly rooted in a deterioration of private investment incentives over the credit cycle. So far, however, there is little theoretical work on the underlying mechanisms that link credit booms to the subsequent decline in asset quality. This leaves a gap in our understanding of the optimal policy response to credit cycles.

To make progress in this direction, I propose a dynamic framework in which increasing credit volumes and falling credit quality are both driven by endogenous movements in the aggregate wealth distribution. My specific focus is on a pecuniary externality in financial markets that hampers origination incentives if and only if asset prices are too high. A distinguishing feature of the model is that banks engage in costly moral hazard in equilibrium even though all agents are fully rational. I use the framework to characterize the macroeconomic conditions that give rise to credit booms with falling investment efficiency, and argue that expansionary monetary policy and saving gluts are plausible triggers of credit cycles.

The basic view I take is that financial intermediaries ("banks") interact with both risk-averse savers and risk-neutral financial market investors ("financiers"). Savers supply funds to banks to invest on their behalf, but are wary of a moral-hazard problem on the part of banks. This leads to an endogenous borrowing constraint that limits the amount of funds banks can borrow as a function of their internal wealth and risk exposure. If internal wealth is scarce, banks can borrow more only by reducing risk exposure. This is accomplished by selling assets to financiers on a secondary market. Secondary market

[1] Greenwood and Hanson (2013) show that the credit quality of corporate bond issuers falls during booms. Mian and Sufi (2009) argue that there was a decline in lending standards prior to the 2008 crisis. Piskorski, Seru, and Witkin (2015) document increased fraud in mortgage originations over the same period.

trading thus allows for an efficient expansion of aggregate credit volumes as long as banks do not sell too many assets. Indeed, I show that a financial system featuring both banks and financiers may extend more credit than one with only banks, holding aggregate net worth fixed.

There is also a downside to secondary market trading, however. In practice, financial markets are typically opaque and bank balance sheets are hard to observe in real time. This paper's starting point thus is that financiers can neither observe the quality of individual assets nor the total number of assets sold by each individual bank. As a result, banks have limited commitment with respect to the degree of risk exposure they retain on their balance sheet. This leads to the possibility that some banks choose to originate and sell a large number of bad assets to financiers under the guise of efficient risk reallocation. When this channel is operational, there is a decline in credit quality and an inefficient increase in aggregate risk exposure. The pecuniary externality of interest is that higher secondary market prices strengthen this nefarious motive for asset sales, and that this effect is not internalized by atomistic financiers. Overall, asset sales may thus at times efficiently boost credit volumes, but at other times lead to inefficient declines in credit quality. The focus of the paper is on the dynamic interaction between these two channels.

In competitive markets with wealth constraints, market-clearing prices reflect the relative wealth of buyers and sellers. In the presence of moral hazard, moreover, the asset price simultaneously shapes incentives. In the present model, the feedback from prices to incentives follows a simple rule: banks shirk if the asset price is above an endogenous threshold, and exert effort otherwise. This leads to an endogeous *maximal price* that is consistent with incentives. So long as the asset price is below this upper bound, it adjusts freely to clear the asset market, and it is increasing in the ratio of financier wealth to bank wealth ("relative financier wealth"). It follows that there exists a cutoff level for relative financier wealth at which the asset price is equal to the maximal price. Beyond this point, the asset price cannot rise further (otherwise all banks would shirk). Accordingly, I show that the wealth distribution can be partitioned into two regions. Below the cutoff for relative wealth, the asset price is low, all banks exert effort, and asset sales boost credit volumes through an efficient reallocation of risk. Above the threshold, the asset price is at its upper bound, a strictly positive fraction of banks shirk, and asset sales lead to an inefficient decline in credit quality. This effect can be understood through a classical tradeoff between incentives and insurance: some risk transfer allows banks to efficiently scale up their investment (insurance), while excessive risk transfer harms incentives. The key difference to classical agency theory is that the equilibrium allocation of risk is implemented not by optimal contracts but by a market mechanism that is sensitive to the wealth distribution.

The discussion above implicitly assumes that financiers continue to buy assets even if some banks shirk. The model makes explicit the conditions under which this is the case. The key observation is that the maximal price is strictly below the expected return of good assets. Financiers thus earn profits even if they buy at the maximal price as long as the fraction of good assets is sufficiently high. To see why there must be shirking in equilibrium when relative financier is above the threshold, note that price adjustments cannot clear the asset market once the maximal price is reached. Excess demand at the maximal price must therefore be matched by an increase in supply given the maximal price. This is accomplished through an increase in the fraction of shirking banks, since banks who shirk sell more assets than those who exert effort. Due to increasing demand, moreover, asset quality falls monotically with financier wealth. This is consistent with individual optimality because financiers earn rents on all infra-marginal asset purchases. Yet it is socially inefficient because banks shirk on their marginal investments. The pecuniary externality is central to this result: there would be no excess demand at the threshold price if financiers internalized their impact on credit quality. A social planner can therefore generate a strict Pareto improvement by restricting financiers' asset purchases. This points to a novel motive for regulation that is independent of the leverage of financial institutions.

The model generates dynamic predictions because the wealth distribution evolves endogenously. The critical state variable is the relative wealth of financiers. It is typically pro-cyclical because financiers endogenously hold more aggregate risk exposure than banks. This is for two reasons. First, the fundamental motivation for asset sales is to transfer risk exposure from banks to financiers. Second, financiers are able to take on more leverage because they are unencumbered by moral hazard in origination. The result is that a sequence of good aggregate shocks (a macroeconomic expansion) leads to a credit boom in which credit quality steadily declines, with longer booms ending in sharper busts. Due to the central role of the wealth distribution, such credit cycles arise even in the absence of exogenous shocks to the production possibility frontier.

The nature and likelihood of credit booms depends on initial conditions. Financiers who are very poor to begin with can only afford to buy a small share of the stock of risky assets, and may therefore fail to grow in relative terms. As a result, asset prices remain low, credit quality does not decline, and credit growth is muted. However, I also show that relative financier wealth grows *no matter* the initial condition as long as the risk-free rate is sufficiently low. The reason is that financiers can leverage disproportionately when borrowing is cheap. This points to a role for saving gluts (an increase in saver wealth) or expansionary monetary policy in triggering credit cycles. Importantly, temporary reductions in the risk-free rate can have persistent effects. The reason is that wealth is a substitute for leverage. Specifically, financiers do not need to rely on cheap borrowing once they have grown sufficiently wealthy. Short-lived shocks to the interest rate may therefore set in motion credit booms with falling credit quality. This points to a dynamic risk-taking channel of monetary policy. I also show that capital requirements on banks may exacerbate the inefficiency because they force banks who exert effort to sell fewer assets, thereby increasing the asset price.

This paper is organized as follows. Section 2 introduces the basic mechanisms in a static model in which the wealth distribution is fixed. Section 3 characterizes the competitive equilibrium. Section 4 embeds the static model in an infinite-horizon economy to study the evolution of the wealth distribution. Section 5 collects the model's empirical predictions and discusses evidence for the key mechanisms. Section 6 studies policy implications, and the conclusion is in Section 7. All proofs are in Appendix A. Extensions are in Appendix B and Appendix C.

Related Literature. An influential literature studies credit booms through the lens of fire-sale externalities in the presence of price-sensitive borrowing constraints (e.g. Kiyotaki and Moore (1997), Bernanke and Gertler (1989), Gromb and Vayanos (2002), Lorenzoni (2008), Mendoza (2010), Bianchi (2011), Gertler and Karadi (2011), He and Krishnamurthy (2011), Brunnermeier and Sannikov (2014) and Davila and Korinek (2018)). These papers have in common the notion that individual agents do not take into account that increasing leverage leads to sharper price drops after bad shocks, generating overinvestment ex-ante. Inefficiencies are therefore tied to the notion that the natural holders of risky assets have too little wealth during bad times, generating inefficient cash-in-the-market pricing and low investment. The opposite is true here: asset quality declines when financiers have too much wealth and asset prices are too high. This allows the model to generate increasingly inefficient outcomes during booms. Moreover, cash-in-the-market pricing is a necessary condition for investment efficiency rather than a manifestation of excessive borrowing ex-ante.

Previous work has studied the relationship between moral hazard and credit booms and busts. Corsetti and Roubini (1999), Bianchi (2016), Bianchi and Mendoza (2013), and Schneider and Tornell (2004) argue that the anticipation of government bailout guarantees leads private agents to make inefficient choices. This paper's moral hazard channel focuses on excessive private demand for financial assets rather than government policy. Myerson (2012) studies a model in which credit cycles stem from the optimal life-cycle incentive contract for bankers. Actions are always efficient since the optimal contract deters moral hazard. I study a model with incomplete markets in which actions are inefficient during booms. Martinez-Miera and Repullo (2017) and Bolton, Santos, and Scheinkman (2016) study models in which banks exert less effort when risk-free rates are low. Here, banks exert less effort during booms with increasing risk-free rates and asset prices.

Gorton and Ordoñez (2014) and Moreira and Savov (forthcoming) study credit booms focusing on the information sensitivity of financial assets. Eisfeldt (2004), Kurlat (2013) and Bigio (2015) study macroeconomic models in which adverse selection amplifies aggregate productivity shocks. These papers share the notion that negative shocks to aggregate productivty or increased uncertainty can lead to sharp drops in credit volumes and investment. In contrast, I study the dynamic feedback from endogenous market liquidity to private incentives and find that high liquidity may lead to a decline in asset quality. Parlour and Plantin (2008) and Vanasco (2017) study the feedback from liquidity to asset quality in static models with fixed secondary market demand. I study the endogenous evolution of secondary market demand via the aggregate wealth distribution and find that secondary markets may grow to be inefficiently large during booms.

2 Static Model

I begin the analysis with a static model in which the wealth distribution is fixed. I study the endogenous evolution of the wealth distribution in Section 4, where I embed the static model in a dynamic framework.

2.1 Environment

Consider a one-period economy populated by a unit mass each of risk-neutral banks and financiers and a unit mass of infinitely risk-averse savers. Savers cannot access investment opportunities directly. Hence they lend to banks and financiers to invest on their behalf. Since savers are infinitely risk averse, they lend only against risk-free debt.

Banks use their own net worth w_B and the debt issued to savers to originate *pools* of risky assets. The payoff generated by a pool depends on unobservable bank effort and an aggregate state $z \in \{l, h\}$, but is deterministic conditional on effort and z. Hence the aggregate state is the only source of risk. The effort choice is binary: the bank either exerts effort or shirks.[2] The probability of state z is $\pi_z \in (0,1)$. The output generated by an investment into a pool of size k in state z is $Y_z k$ if banks exert effort and $y_z k$ if banks shirk. Returns are higher in the high aggregate state, $Y_h > Y_l$ and $y_h > y_l$, and expected returns are denoted $\hat{Y} = \mathbb{E}_z Y_z$ and $\hat{y} = \mathbb{E}_z y_z$, respectively. Shirking generates a private non-pecuniary benefit of mk. Banks who exert effort are said to be *good banks* who produce *good assets*. Banks who shirk are said to be *bad banks* who produce *bad assets*. The payoff structure is assumed to lead to a risk-shifting problem (Jensen and Meckling (1976)).

Assumption 1 (Risk-shifting Moral Hazard)**.** *Shirking is inefficient and disproportionately increases downside risk:* $\mathbb{E}_z Y > E_z y_z + m$*,* $y_l < Y_l$ *and* $y_h > E_z Y_z$*.*

Financiers are endowed with net worth w_F and issue debt to savers. They use these funds to purchase risky assets originated by banks on secondary markets, and to invest in a safe constant-returns-to-scale technology that generates a risk-free return normalized to one. They cannot access the risky technology, and thus cannot originate assets.

Risky assets trade in a competitive secondary market. I follow Bigio (2015) and Kurlat (2013) in assuming that this market is *anonymous* and *non-exclusive*: assets are traded individually, banks and financiers interact with multiple counterparties, and financiers neither observe the quality of assets sold (since effort is not observed) nor can an individual bank commit to the total quantity of assets it sells (since markets are non-exclusive). Banks can pledge the proceeds from asset sales as collateral in order to ensure that the debt they issue to savers is risk-free. (That is, debt can be *secured* using the proceeds from asset sales. Gorton and Metrick (2012a) show that this is the typical motive for securitization.) If a bank pledges the proceeds from $\underline{a} \in [0,k]$ asset sales as collateral, it publicly reveals that it has sold at least $\underline{a} \in [0,k]$ assets. Anonymity and non-exclusivity ensures that it does not also reveal its *total sales* $a_B \in [\underline{a}, k]$. The number of bonds issued and the total investment of each bank is publicly observable.

I show below that bank incentives to sell more than $\underline{a}$ assets are determined by the

[2] Appendix D provides a simple argument for why binary effort is a sensible modeling assumption if financiers have some discretion over the specific assets they buy from a given pool.

asset price and $\chi \equiv \frac{a}{k}$, the ratio of pledged assets to total investment. The number of assets sold in turn affects effort incentives. Financiers will therefore condition their bids on χ. To account for this, I assume that the secondary market consists of a set of perfectly competitve submarkets indexed by $\chi \in [0, 1]$. Banks who pledge fraction χ are restricted to trading in submarket χ, and are price-takers in this market. Financiers can trade in any submarket, and are price-takers in every submarket they participate in. The market-clearing price in each submarket is $P(\chi)$. A submarket is *active* if banks offer a strictly positive number of assets for sale in this submarket. The set of active submarkets is $\mathcal{A}$. All assets sold in a given submarket are randomly allocated to buyers, so that each buyer receives a representative slice of all assets sold. The fraction of good assets in submarket χ is denoted by $f(\chi)$. The return generated by an asset in state z is $x_z(\chi) = f(\chi)Y_z + (1 - f(\chi))y_z$, and the expected return is $\hat{x}(\chi) = E_z x_z(\chi)$.

To isolate the key mechanisms, I assume for now that savers are deep-pocketed and elastically demand risk-free claims at the exogenous interest rate $1 + r_f$. I represent these claims as zero-coupon bounds with face value one that trade at price $q = \frac{1}{1+r_f}$. The risk-free rate is assumed to be weakly positive for simplicity, $q \leq 1$, but the model can also accommodate negative interest rates. The dynamic model in Section 4 endogenizes the risk-free rate by incorporating a wealth constraint for savers.

2.2 Discussion

The model makes three central assumptions. The first is that neither the quality nor the total quantity of assets sold by banks are observable, and that individual banks can trade with multiple financiers. This implies that financiers cannot screen for asset quality by offering price-quantity menus or by requiring banks to retain a given fraction of their assets. The motivation for this assumption is fourfold. First, originators of financial securities have wide discretion over asset quality. Second, the vast majority of asset-backed securities were traded in opaque over-the-counter markets prior to the financial crisis. Since these markets are decentralized, traders can neither observe the trading partners of their counterparties, nor the quanity of assets sold in other bilateral negotiations. Hence it is reasonable to model secondary markets as non-exlusive. Attar, Mariotti, and Salanié (2011) study such settings theoretically, and show that it may indeed not be possible to screen by quantity if sellers trade with multiple buyers. Third, the balance sheets of financial insitutions are hard to observe in real time. Hence financiers cannot properly assess a given banks exposure to a particular asset pool. Four, banks have access to a wide variety of financial securities with which to reduce exposure to a given risk factor.

The second key assumption is that banks can pledge proceeds from asset sales as collateral when issuing debt. This assumption is motivated by the practice of selling pools of risky assets into bankruptcy-remote special purpose vehicles during securitization (Gorton and Metrick (2012a)). Since special-purpose vehicles are legally separated from bank balance sheets by construction, they serve as credible mechanisms by which banks can demonstrate that they have sold a given asset, even if they cannot commit to not selling additional assets in the future. Proceeds from asset sales can thus serve as collateral for issuances of new debt.[3] Indeed, this is the major benefit of securitization (Gorton and Metrick (2012a)).

Third, the contract space is such that financiers cannot claw back losses from banks in the event that financiers were sold low-quality assets. This can be motivated by the following example in which loan pools consist of individual assets whose payoffs have identical support. Hence effort is not ex-post observable.

Example 1. *Let asset pools consist of a continuum of individual assets all of which either succeed and yield Y^* or fail and yield nothing. Let the probability of success given aggregate state z and effort decision e be α_z^e. Let $\alpha_h^1 = \frac{Y_h}{Y^*}$, $\alpha_h^0 = \frac{y_h}{Y^*}$, $\alpha_l^1 = \frac{Y_l}{Y^*}$ and $\alpha_l^0 = \frac{y_l}{Y^*}$, and let project returns be independently distributed. Then effort is not ex-post observable from individual project returns. However, by the law of large numbers, the returns generated by a pool are as stated above.*

Individual assets can be interpreted as claims on real economy that either pay a fixed

[3] Appendix C provides a detailed microfoundation using a variant of the model in which there are multiple rounds of lending and investing in every period. Since asset sales allow banks to issue more debt in the next round, commitment to a minimum level of asset sales arises endogenously.

return or default. Examples include household mortgages or corporate loans. The i.i.d assumption conditional on the aggregate state ensures that returns are deterministic at the pool level given z. This is a convenient assumption given that savers lend only against risk-free debt. The downside is that shirking is observable ex-post at the level of the portfolio. Hence the form of market incompleteness assumed here is equivalent to the non-contractibility of bank portfolio returns. This is plausible given the complications involved in monitoring bank balance sheets in real time. One straightforward way of relaxing the assumption is to model savers' preferences using a generalized value-at-risk constraint that permits some default (Adrian and Shin (2014)), and to let the portfolio probability of default vary with effort. In that case, effort is not observable from portfolio returns ex-post, and yet there is a motive for selling assets to relax borrowing constraints.

2.3 Borrowing and retention constraints

Before turning to the characterization of equilibrium, it is instructive to study the conditions under which banks are willing to exert effort. Banks derive utility from returns generated by retained risky assets, revenues from asset sales $P(\chi)a_B$, and private benefits from shirking mk, if any. They suffer a disutility from debt payments b_B. The returns to retained risky assets are $Y_z(k-a)$ if the bank exerts effort and $y_z(k-a_B)$ if the bank shirks. So banks prefer to exert effort if

$$\sum_z \pi_z \left[Y_z\left(k-a_B\right)-b_B+P(\chi)a_B\right] \geq \sum_z \pi_z \left[\max\left\{y_z\left(k-a_B\right)-b_B+P(\chi)a_B,0\right\}\right]+mk. \tag{IC}$$

The max-operator imposes the *limited-liability constraint* that banks cannot pay back more than their net worth. (Given that banks are assumed to exert effort and debt is risk-free, banks must be solvent for all z conditional on effort.)

There are two reasons to shirk. The first is that the bank has issued too much debt. This leads to a borrowing constraint. The second is that the bank has sold too many assets. This leads to a retention constraint specifying that the bank must retain a minimum fraction of its assets. Both consraints can be succinctly described using the following reduced-form parameters.

Definition 1. *The* ***pledgable return*** *is* $\rho = \frac{\mathbb{E}_z Y_z - \pi_h y_h}{\pi_l} > 0$.

Definition 2. *The* ***moral hazard discount*** *is* $\tilde{m} = 1 - \frac{m}{\mathbb{E}_z Y_z - \pi_h y_h} \in (0, 1)$.

To derive the borrowing constraint, recall that $y_h > Y_l > y_l$ by Assumption 1, and guess and verify that the limited-liability constraint binds only in the low state conditional on shirking. Given a_B and k, the incentive-compatibility constraint (IC) can then be rearranged to give

$$b_B \leq \tilde{m}\rho k + (P(\chi) - \rho)\, a_B. \tag{1}$$

This *borrowing constraint* states that banks can borrow up to the pledgeable return on their investments ρk (adjusted by the moral hazard discount factor $\tilde{m}$), plus the pledgeable income generated by asset sales above and beyond ρ. Asset sales thus increase borrowing capacity by serving as a subsitute for own net worth. The following assumption ensures that the bank is solvent conditional on effort if the borrowing constraint is satisfied.[4]

Assumption 2. $y_h \geq Y_h - \frac{m}{\pi_h}$.

Now turn to the retention constraint. For any k and b_B satisfying the borrowing constraint, the bank is solvent (the limited-liability constraint does not bind) in all states of the world conditional on shirking if a_B is sufficiently close to k. If the bank is always solvent, then (IC) is satisfied only if

$$a_B \leq \bar{\chi} k \qquad \text{where} \qquad \bar{\chi} = \left(\frac{\tilde{m}\rho - y_l}{\rho - y_l}\right) \in (0, \tilde{m}). \tag{2}$$

This *retention constraint* states that banks are willing to exert effort only if they retain no less than a fraction $1 - \bar{\chi}$ of their assets.[5]

[4] If the assumption was not satisfied, the solvency constraint in state l conditional on effort may be the binding constraint given some a_B. This would complicate the analysis because multiple borrowing constraints would have to be checked. The basic mechanism that asset sales boost borrowing capacity would not be affected, however.

[5] To verify that this is indeed the relevant retention constraint, note that the bank is indeed solvent in all states conditional on shirking if the borrowing constraint holds and $a_B \geq \bar{\chi} k$.

The analysis thus shows that banks exert effort only if the borrowing constraint and the retention constraint are jointly satisfied. There are important conceptual differences between the two constraints. The borrowing constraint is a saver-facing constraint, in that it determines the maximum debt that can be issued to savers without harming incentives given $k, \underline{a}$ and a_B. Since k, b_B, and $\underline{a} \leq a_B$ are observable, the constraint is never violated in equilibrium. (Banks would be insolvent on the equilibrium path if it were violated, and savers would cease to lend).

The retention constraint is a financier-facing constraint, in that it determines the maximum fraction of assets that can be sold on secondary markets without harming incentives. Since a_B is unobservable, it may be violated in equilibrium. This is the key mechanism by which secondary market trading leads to declines in asset quality.

Notably, savers are unaffected by violations of the retention constraint since banks continue to trade at the same asset price (and thus remain solvent) after a deviation to $a_B > \underline{a}$. The model can therefore be interpreted as an agency relationship in which the bank has two sets of principals (savers and financiers), an incentive-compatibility constraint for each, and only the financier-facing constraint can be violated in equilibrium.

2.4 Price bounds

In this section, I argue that the asset price is a key determinant of bank incentives. This is accomplished by showing that banks optimally choose to violate the retention constraint if and only if the asset price is too high. Consider a bank who has sold and pledged $\underline{a}$ assets as collateral, invested k, and issued b_B in debt. Since secondary markets are anonymous, the bank has the option to sell more than $\underline{a}$ assets without affecting the asset price. Since effort is unobservable, moreover, the bank can tailor its effort to the number of assets it sells. We must therefore worry about "double deviations" where the bank sells more assets than it promised and shirks as a result. The linear objective ensures that if it is optimal to deviate to some $a_B > \underline{a}$, then the optimal deviation is to sell *all* assets ($a_B = k$). By the retention constraint, a bank which sells all of its assets will find it optimal to shirk. So the payoff of a double deviation to *shirking and selling* is $P(\chi)k - b_B + mk$. The alternative is to sell just as many assets as promised ($a_B = \underline{a}$). Take as given that the bank exerts effort if $a_B = \underline{a}$.[6] The payoff to *retention and effort* is $\mathbb{E}_z Y_z(k - \underline{a}) - b_B + P(\chi)\underline{a}$. Comparing these payoffs leads to the following proposition.

Proposition 1 (Maximum Asset Price)**.** *Assume that effort is optimal given $a_B = \underline{a}$ and let $\chi = \underline{a}/k$. Then the bank optimally sells $\underline{a}$ assets and exerts effort only if $P(\chi) \leq \bar{P}(\chi) \equiv \hat{Y} - \frac{m}{1-\chi}$, and strictly prefers to sell k assets and shirk if $P(\chi) > \bar{P}(\chi)$. $\bar{P}(\chi)$ is strictly decreasing in χ and $\bar{P}(\bar{\chi}) = \hat{y}$.*

The intuition is as follows. If the bank sells all of its assets, then it receives a return of exactly $P(\chi)$ on each unit. If the bank retains part of its assets, it receives a return of $P(\chi)$ on χk of its assets, but an expected return of $\hat{Y}$ on the remaining $(1-\chi)k$ assets. So marginal increases in the asset price affect the returns of the sell and shirk strategy more than the retention and effort strategy. Banks therefore sell and shirk if the asset price is sufficiently high.

Notice that the maximal asset price is below the expected return of good assets, $\bar{P}(\chi) < \hat{Y}$. This has two implications. The first is that financiers earn a profit if banks are willing to sell good assets at $\bar{P}(\chi)$. Competition among financiers may therefore lead the price to reach its upper bound. If financiers are relatively poor, however, then the asset price will instead be determined by a binding wealth constraint ("cash-in-the-market pricing"). Since the upper bound is strictly decreasing, I restrict attention to downward-sloping price schedules.

Assumption 3. *$P(\chi)$ is decreasing in χ.*

The second implication is that banks must trade off the cost of selling assets below par against the value of additional borrowing capacity. This calculation leads to a lower bound on the asset price below which good banks are not willing to sell.

[6] If the bank were to shirk given $a_B = \underline{a}$, then it is strictly optimal to sell all assets, and the bank always shirks regardless.

Proposition 2 (Minimum Asset Price)**.** *Let $P(\chi)$ be differentiable on $\chi \in [0, \bar{\chi}]$ and let $\Pi(q) = \frac{q\hat{Y}-1}{1-q\tilde{m}\rho}$. Good banks are willing sell and pledge assets as collateral ($\underline{a} > 0$) only if $P(\chi) \geq \underline{P}(q) \equiv \frac{\hat{Y}+\rho\Pi(q)}{1+\Pi(q)}$. Moreover, $\underline{P}(q)$ is strictly decreasing and $\underline{P}\left(\frac{1}{\hat{Y}}\right) = \hat{Y}$.*

The minimum price $\underline{P}(q)$ is strictly decreasing in q because the ability to issue debt is more valuable when the cost of capital is low. This implies that good assets will not be traded in equilibrium if q is too low. Specifically, if $\underline{P}(q) > \bar{P}(\chi)$, then banks are willing to sell good assets only at prices that would lead them to shirk.

Corollary 1. *For each q there exists a $\underline{\chi}(q) < 1$ such that $\bar{P}(\chi) < \underline{P}(q)$ for all $\chi > \underline{\chi}(q)$. This cutoff value is given by $\underline{\chi}(q) = 1 - \frac{m}{\hat{Y}-\underline{P}(q)}$.*

This result implies that banks are not willing to exert effort if $\chi > \underline{\chi}(q)$ and the secondary market is active. We have already shown that banks do not exert effort if $\chi > \bar{\chi}$. This leads to the following definition.

Definition 3. *χ is **consistent with effort** given q if $\chi \leq \chi^{\max} \equiv \min\{\bar{\chi}, \underline{\chi}(q)\}$. The set of submarkets consistent with effort is $\mathcal{E} = [0, \chi^{\max}]$.*

Corollary 2. *$\underline{\chi}(q) \geq 0$ if and only if $q \geq \underline{q} \equiv \frac{\hat{Y}-\rho}{\hat{Y}\left(\hat{Y}-\rho\right)-m\left(\hat{Y}-\tilde{m}\rho\right)} > \frac{1}{\hat{Y}}$. Moreover, $\underline{\chi}(q) \geq \bar{\chi}$ if and only if $q \geq \bar{q} \equiv \frac{\hat{Y}-\rho}{\hat{Y}(\hat{y}-\rho)+\tilde{m}\rho\left(\hat{Y}-\hat{y}\right)} > \underline{q}$. Hence $\chi^{\max} = \bar{\chi}$ if and only if $q \geq \bar{q}$.*

It follows that there are no submarkets consistent with effort if $q < \underline{q}$. The reason is that banks are not particularly borrowing-constrained when q is small. So financiers anticipate that banks will shirk whenever they are offered a price at which they would be willing to sell good assets. Liquid secondary markets thus require that banks are sufficiently constrained. I will therefore assume that $q > \underline{q}$.

2.5 Definition of Equilibrium

I define a competitive equilibrium for this economy as follows.

Definition 4 (Competitive Equilibrium)**.** *A* competitive equilibrium *for this economy is a weakly decreasing price schedule $P : [0, 1] \to \mathbb{R}_+$, an effort decision $e \in \{0, 1\}$ and portfolio*

$\{k, b_{,B}, \underline{a}, a_B\}$ *for each bank, and a portfolio* $\{a_F, b_F, s_F\}$ *for each financier such that (i) bank portfolios and effort decisions are jointly optimal for each bank given* P*, (ii) financier portfolios are optimal given* P *and bank effort decisions, and (iii) all active submarkets clear.*

Similar to Dubey and Geanakoplos (2002), this definition is not sufficiently restrictive. Since all agents are price-takers and markets must clear only in active submarkets, the price schedule can be chosen to render any ad-hoc set of submarkets inactive (if $P(\chi) = 0$, for example, then there is no incentive to sell assets in submarket χ). So it is possible to sustain competitive equilibria in which there is trade only in some arbitrarily chosen submarkets. To address this concern, I use an equilibrium refinement in the spirit of Dubey and Geanakoplos (2002). The idea is to introduce a fictitious buyer who purchases an infinitesimal number of assets ϵ at price p in all inactive submarkets consistent with effort. This price is chosen to be the best price currently offered in active submarkets consistent with effort. That is, $p = \max_{\chi' \in \mathcal{E} \cap \mathcal{A}} P(\chi')$. This ensures that banks are not punished for deviating to some χ if doing so is consistent with effort, and rules out equilibria in which certain submarkets are inactive purely by means of an exogenously imposed price schedule. Off-equilibrium prices are thus linked to market-clearing prices so long as χ is consistent with effort. In order to minimize the scope for inefficient shirking, I impose that markets inconsistent with effort are inactive.[7]

Definition 5 (Refined Competitive Equilibrium)**.** *A* refined competitive equilibrium *is a competitive equilibrium in which* $P(\chi) = \max_{\chi' \in \mathcal{E} \cap \mathcal{A}} P(\chi')$ *for all* $\chi \in \mathcal{E} \cap \mathsf{C}(\mathcal{A})$ *and* $P(\chi) = 0$ *for all* $\chi \notin \mathcal{E}$

Going forward, I refer to the refined competitive equilibrium simply as the competitive equilibrium. Two observations are in order. First, there always exists a trivial competitive equilibrium in which all submarkets are inactive and $P(\chi) = 0$ for all $\chi \in [0, 1]$. I restrict attention to equilibria with active submarkets whenever they exist. Second, there

[7] If w_F is large and parameters satisfy certain conditions, it is possible to construct equilibria in which only submarkets inconsistent with effort are active. These necessarily involve all banks shirking, dramatically exacerbating the inefficiency studied here. Moreover, these equilibria can be ruled out by a simple form of coordination among financiers. I therefore do not consider them here.

may be instances in which individual portfolios are indeterminate at equilibrium prices. Whenever this is the case, I restrict attention to symmetric portfolios. Since all optimal policy functions will turn out to be linear in own net worth, this is without loss of generality with respect to aggregate outcomes.

Remark 1. *The refined equilibrium price schedule is constant across all submarkets consistent with effort. This implies that banks cannot affect the asset price deviating to some χ that is also consistent with effort. This price-taking assumption is natural the general equilibrium framework emphasized here. If the upper bound $\bar{P}(\chi)$ binds, however, it may nevertheless be reasonable to consider an alternative approach in which individual banks* can *affect the asset price by altering χ. The reason is that $\bar{P}(\chi)$ is strictly decreasing. Reducing χ may therefore allow the bank to signal that it will continue to exert effort even if it is offered a slightly higher asset price. Accordingly, Appendix B considers an alternative refinement in which this "signaling" channel is permitted, and shows that doing so strengthens the key mechanisms. Specifically, allowing banks to signal by retaining a larger fraction of their assets typically leads to more excess demand and thus more shirking.*

3 Equilibrium

I now chacterize competitive equilibrium. Let p denote the constant asset price for all submarkets $\chi \in [0, \chi^{\max}]$. To ensure that there is scope for trade on secondary markets, let the set of submarkets consistent with effort be non-empty ($\chi^{\max} > 0$). This requires that $q > \underline{q} > \frac{1}{\bar{Y}}$, and so banks will always borrow as much as possible.

3.1 Optimal Portfolios

The first step is to characterize the optimal portfolios of banks and financiers. Optimal choices are denoted by asterisks.

Banks who exert effort (good banks) maximize expected utility subject to the budget and borrowing constraints, as well as the requirement that they trade in a submarket consistent with effort. Since the retention constraint must be satisfied, we can take as given that $a_B = \underline{a}$. Hence a good bank's decision problem is

$$\max_{k \geq 0, b_B \geq 0, \underline{a} \geq 0} \quad \hat{Y}\left(k - \underline{a}\right) - b_B + p\underline{a}$$

$$\text{s.t. } k \leq w_B + q b_B \qquad \text{(Budget constraint)}$$

$$b_B \leq \tilde{m}\rho k + \left(p - \rho\right)\underline{a} \qquad \text{(Borrowing constraint)}$$

$$\chi \leq \chi^{\max} \text{ where } \chi = \underline{a}/k. \qquad \text{(Effort consistency)}$$

Proposition 3 (Optimal Good Bank Portfolio)**.** *The optimal portfolio of a good bank is given by* $\{k^*(\chi^*), b_B^*(\chi^*), \underline{a}^*\}$*, where* $k^*(\chi) = \frac{w_B}{1 - q\tilde{m}\rho - \chi q(p-\rho)}$*,* $b_B^*(\chi) = \left(\tilde{m}\rho + \chi(p-\rho)\right)k^*(\chi)$*,* $\underline{a}^* = \chi^* k^*(\chi^*)$*, and* $\chi^* = \chi^{\max}$ *if* $p > \underline{P}(q)$ *and* $\chi^* \in [0, \chi^{\max}]$ *if* $p = \underline{P}(q)$*. Moreover,* $a_B^* = \underline{a}^*$*.*

The following assumption ensures that bank leverage is bounded.

Assumption 4 (Bounded Leverage)**.** $1 > q\tilde{m}\rho + \chi q(\bar{P}(\chi) - \rho)$ *for all* $\chi \in [0, \chi^{\max}]$

Next consider banks who shirk (bad banks). Since asset quality is a choice, not a type, banks shirk only if doing so generates at least as high a payoff as exerting effort. Given that only submarkets consistent with effort are active, it is optimal to shirk only if $p \geq \bar{P}(\chi)$ for some $\chi \in [0, \chi^{\max}]$. Since p is constant on $\chi \in [0, \chi^{\max}]$ and $\bar{P}(\chi)$ is strictly decreasing, financiers are willing to buy only if $p \leq \bar{P}(\chi^{\max})$. (Else all banks would shirk.) Hence we can focus without loss of generality on the case $p = \bar{P}(\chi^{\max})$.

Proposition 4 (Optimal Bad Bank Portfolio)**.** *Let* $p = \bar{P}(\chi^{\max})$*. Then the optimal portfolio of a bad bank is* $\{k^*(\chi^*), b_B^*(\chi^*), \underline{a}^*\}$*, where* $\chi^* = \chi^{\max}$*,* $k^*(\chi) = \frac{w_B}{1 - q\tilde{m}\rho - \chi q(p-\rho)}$*,* $b_B^*(\chi) = \left(\tilde{m}\rho + \chi(p-\rho)\right)k^*(\chi)$*, and* $\underline{a}^* = \chi^* k^*(\chi^*)$*. Moreover,* $a_B^* = k^*(\chi^*)$*.*

Bad banks thus choose the same observable portfolio $\{k, b_B, \underline{a}\}$ as good banks, and differ only in that they sell more assets ex-post ($a_B^* = k$). Hence asset quality cannot be inferred from observables. Let $\phi(\chi) \in [0, 1]$ denote the fraction of good banks in submarket χ. The total supply of risky assets in submarket χ then is $\mathbf{a}_B(\chi) = \phi(\chi)\underline{a} + (1 - \phi(\chi))k$, and the fraction of good assets is $f(\chi) = \frac{\phi(\chi)\underline{a}}{\mathbf{a}_B(\chi)} = \frac{\phi(\chi)\chi}{1 - \phi(\chi)(1-\chi)} \leq \phi(\chi)$.

Now consider the optimal financier portfolio. Financiers maximize expected utility by choosing how much debt to issue, how many risky assets to purchase, and how much to invest in the safe technology. Since financiers take prices as given and are never rationed, we can assume without loss of generality that each financier invests in at most one submarket, say χ. Let $a_F(\chi)$ denote the number of risky assets purchased, let s_F denote investment in the safe technology, and let b_F denote the amount of debt issued. Then the decision problem is

$$\max_{a_F \geq 0, s_F \geq 0, b_B \geq 0} \quad \hat{x}(\chi)a_F + s_F - b_F$$

$$\text{s.t. } pa_F + s_F \leq w_F + qb_F \qquad \text{(Budget constraint)}$$

$$b_F \leq x_l(\chi)a_F + s_F \qquad \text{(Borrowing constraint)}$$

The borrowing constraint requires financiers to be able to repay their debts in full in all states. The solution is in two steps. The first step is the optimal allocation of funds across risky assets and the safe technology.

Lemma 1 (Optimal Investment Rule). *Fix some $b_F \geq 0$ satisfying the borrowing constraint. If $p < \hat{x}(\chi)$, then $a_F^*(\chi|b_F) = \frac{w_F + qb_F}{p}$ and $s_F^* = 0$. If $p > \hat{x}(\chi)$, then $a_F^*(\chi|b_F) = 0$ and $s_F^* = w_F + qb_F$. If $p = \hat{x}(\chi)$, then $a_F^*(\chi|b_F) \in \left[0, \frac{w_F + qb_F}{p}\right]$ and $s_F^* = w_F + qb_F - pa_F^*(\chi|b_F)$.*

Observation 1. *Financiers earn the expected portfolio rate of return $\tilde{r}(\hat{x}, p) \equiv \max\left\{\frac{\hat{x}}{p}, 1\right\}$. Financiers are willing to buy risky assets only if $\hat{x}(\chi) \geq p$. This condition is equivalent to $\phi(\chi) \geq \underline{\phi}(p, \chi) \equiv \frac{p - \hat{y}}{p - \hat{y} + \chi(\hat{Y} - p)}$. Financiers thus buy risky assets only if the fraction of good banks is not too low.*

The second step is to characterize optimal leverage. To this end, it is convenient to rewrite the borrowing constraint as the equality constraint $b_F = \gamma\left[x_l(\chi)a_F + s_F\right]$, where $\gamma \in [0, 1]$ denotes the degree to which borrowing capacity is exhausted. I will refer to γ simply as the *borrowing decision*.

Lemma 2 (Optimal Borrowing Rule). *Let γ^* denote the optimal borrowing decision given $\{\hat{x}(\chi), p, q\}$. If $\tilde{r}(\hat{x}(\chi), p) > \frac{1}{q}$, then $\gamma^* = 1$. If $\tilde{r}(\hat{x}(\chi), p) < \frac{1}{q}$, then $\gamma^* = 0$. If $\tilde{r}(\hat{x}(\chi), p) = \frac{1}{q}$,* *then* $\gamma^* \in [0, 1]$.

The optimal investment and borrowing rules jointly determine the optimal portfolio. I provide a characterization for $\hat{x}(\chi) \geq p$ only. (Financiers do not buy assets otherwise.)

Proposition 5 (Optimal Financier Portfolio)**.** *Let $\hat{x}(\chi) \geq p$. The optimal portfolio is such that:*

(i) If $q\hat{x}(\chi) > p$, then $s_F^* = 0$, $a_F^*(\chi) = \frac{w_F}{p - qx_l(\chi)}$, and $b_F^* = x_l(\chi)a_F^*(\chi)$.

(ii) If $\hat{x}(\chi) > p = q\hat{x}(\chi)$, then $s_F^* = 0$, $a_F^*(\chi) = \frac{w_F}{p - \gamma^* qx_l(\chi)}$, $b_F^* = \gamma^* x_l(\chi)a_F^*(\chi)$, and $\gamma^* \in [0, 1]$.

(iii) If $\hat{x}(\chi) > p > q\hat{x}(\chi)$, then $s_F^* = 0$, $a_F^*(\chi) = \frac{w_F}{p}$, and $b_F^* = 0$.

(iv) If $\hat{x}(\chi) = p$, then $a_F \in [0, \frac{w_F}{p - \gamma^* qx_l(\chi)}]$, where $\gamma^* \in [0, 1]$ if $q = 1$ and $\gamma^* = 0$ if $q < 1$.

Demand for risky assets is strictly increasing in financier wealth w_F and, if financiers are leveraged, in the bond price q.

Putting the optimal portfolios together yields the competitive equilibrium. The market clearing condition that determines the asset price is $\mathbf{a}_B(\chi) = a_F(\chi)$. Since asset demand is increasing in w_F and asset supply is increasing in w_B, the equilibrium can be characterized as function of relative wealth $\omega \equiv \frac{w_F}{w_B}$. The key result is that there exists a threshold for relative financier wealth $\bar{\omega}$ such that a strictly positive fraction of banks shirks if $\omega > \bar{\omega}$. Throughout, I focus on symmetric portfolios to resolve indeterminacies in individual portfolios. This is sufficient if only one side of the market is indeterminate. If both bank and financier portfolios are indeterminate, I select the equilibrium with maximum secondary market volumes. I refer to this as as the *symmetric equilibrium*. Given ϕ and k, aggregate output in state z is defined to be $\mathbf{Y}(z) = (\phi Y_z + (1 - \phi)y_z)\, k$.

Proposition 6. *Let $\omega = \frac{w_F}{w_B}$ denote the relative wealth of financiers. In addition to the trivial equilibrium in which all submarkets are inactive, there exists a unique symmetric equilibrium with trade featuring a single active submarket χ^* and equilibrium price p^*. This equilibrium is characterized by two thresholds for relative wealth, $\underline{\omega}$ and $\bar{\omega} \geq \underline{\omega}$, and has the following structure.*

(i) If $p^* = q\hat{x}(\chi^*)$, then $\gamma^* = \min\left\{\max\left\{\frac{p^*\mathbf{a}_B(\chi^*) - w_F}{qx_l(\chi^*)\mathbf{a}_B(\chi^*)}, 0\right\}, 1\right\}$.

(ii) *If* $\omega \in (0, \underline{\omega}]$*, then* $\chi^* < \chi^{\max}$*,* $p^* = \underline{P}(q)$*, and* $\phi^*(\chi^*) = 1$*.* k^* *is increasing in* w_B *and* w_F*.*

(iii) *If* $\omega \in (\underline{\omega}, \bar{\omega}]$*, then* $\chi^* = \chi^{\max}$*,* $p^* = \frac{\omega(1-q\rho(\tilde{m}-\bar{\chi}))+\bar{\chi}\gamma^* qY_l}{\bar{\chi}(1+q\omega)}$ *and* $\phi^*(\chi^*) = 1$*.* k^* *is increasing in* w_B *and* w_F*, and* p^* *is increasing in* ω*.*

(iv) *If* $\omega \in (\bar{\omega}, \infty)$*, then* $\chi^* = \chi^{\max}$*,* $p^* = \bar{P}(\chi^{\max})$*, and* $\phi^*(\chi^*) = \frac{k(\chi^*)-a_F(\chi^*)}{(1-\chi^*)k(\chi^*)} \in [0, 1)$*. If* $q\hat{x}(\chi^{\max}) \neq \bar{P}(\chi^{\max})$*, then* $\phi^*(\chi^*)$ *is strictly decreasing in* ω *up to* $\phi^*(\chi^*) = \underline{\phi}(p^*, \chi^*)$*, and* $\mathbb{E}_z\mathbf{Y}(z)$ *is strictly decreasing in* w_F*. If* $q\hat{x}(\chi^{\max}) = \bar{P}(\chi^{\max})$*, then* $\phi^*(\chi^*)$ *and* $\mathbb{E}_z\mathbf{Y}(z)$ *are locally independent of* w_F*.*

(v) *Let* $\bar{\gamma}(p) = \mathbb{1}(q\hat{Y} > p)$ *and* $\underline{\gamma}(p) = 1 - \mathbb{1}(q\hat{Y} < p)$*. Then* $\underline{\omega}$ *and* $\bar{\omega}$ *are determined as follows:*

(a) *If* $\bar{\chi} < \underline{\chi}(q)$*, then* $\chi^{\max} = \bar{\chi}$ *and* $\bar{P}(\bar{\chi}) > \underline{P}(q)$*, Moreover,* $\bar{\omega} = \frac{\bar{\chi}\left(\bar{P}(\bar{\chi})-\bar{\gamma}\left(\bar{P}(\bar{\chi})\right)qY_l\right)}{1-q\rho\tilde{m}-q\bar{\chi}(\bar{P}(\bar{\chi})-\rho)}$*,* $\underline{\omega} = \frac{\bar{\chi}\left(\underline{P}(q)-\underline{\gamma}(\underline{P}(q))qY_l\right)}{1-q\rho\tilde{m}-q\bar{\chi}(\underline{P}(q)-\rho)}$*, and* $\bar{\omega} > \underline{\omega}$*.*

(b) *If* $\underline{\chi}(q) \leq \bar{\chi}$*, then* $\chi^{\max} = \underline{\chi}(q)$*, and* $\bar{P}(\chi^{\max}) = \underline{P}(q)$*. Moreover,* $\bar{\omega} = \frac{\underline{\chi}(q)(\underline{P}(q)-\bar{\gamma}(\underline{P}(q))qY_l)}{1-q\rho\tilde{m}-q\underline{\chi}(q)(\underline{P}(q)-\rho)}$ *and* $\underline{\omega} = \frac{\underline{\chi}(q)\left(\underline{P}(q)-\underline{\gamma}(\underline{P}(q))qY_l\right)}{1-q\rho\tilde{m}-q\underline{\chi}(q)(\underline{P}(q)-\rho)}$*, where* $\bar{\omega} = \underline{\omega}$ *if* $q\hat{Y} \neq \underline{P}(q)$ *and* $\bar{\omega} > \underline{\omega}$ *if* $q\hat{Y} = \underline{P}(q)$*.*

The intuition for the upper threshold is as follows. Assume that all banks exert effort. Then asset supply is increasing in w_B, and asset demand is increasing in w_F. Hence the equilibrium price is increasing relative financier wealth ω, and there exists a threshold $\bar{\omega}$ such that the equilibrium asset price is equal to the upper bound if $\omega = \bar{\omega}$. Since $\bar{P}(\chi) \geq \hat{y}$, the equilibrium price cannot increase further (otherwise all banks would shirk and no financier would buy). This presents a challenge for market clearing at $\omega > \underline{\omega}$. Since $\bar{P}(\chi) < \hat{Y}$, financiers are not willing to reduce asset demand if sufficiently many banks exert effort. So asset supply must increase to clear markets. Good banks are already selling the maximum number of assets consistent with effort. Hence the only way to further increase asset supply is for some banks to shirk. (Bad banks sell more assets than good banks). Shirking thus is a necessary equilibrium consequence of excess demand at $\bar{P}(\chi^{\max})$. Importantly, this method of clearing markets is consistent with individual rationality. Banks are indifferent between shirking and effort at $\bar{P}(\chi^{\max})$ by construction, and financiers prefer to buy risky assets so long as $\phi^* \geq \underline{\phi}(\bar{P}(\chi^{\max}), \chi^{\max})$.

The lower threshold is determined such that the equilibrium price equals its lower bound if $\omega = \underline{\omega}$. Hence financiers can afford to buy exactly $\chi^{\max} k^*(\chi^{\max})$ assets at price $\underline{P}(q)$ if $\omega = \underline{\omega}$. Since banks are not willing to sell below $\underline{P}(q)$, the equilibrium price is then pinned at its lower bound for all $\omega < \underline{\omega}$ and market volumes are fully determined by financier demand.

Finally, note that $\underline{\omega}$ and $\bar{\omega}$ may coincide if $\chi^{\max} = \underline{\chi}(q)$. The reason is that $\bar{P}(\chi)$ is strictly decreasing in χ. If $\underline{\chi}(q) < \bar{\chi}$, then an increase in χ may be sufficient to reach the upper bound even if $p = \underline{P}(q)$. Since χ is increasing in w_F when $\omega < \underline{\omega}$, growing asset demand alone may thus be sufficient to trigger bank shirking even if $p = \underline{P}(q)$.

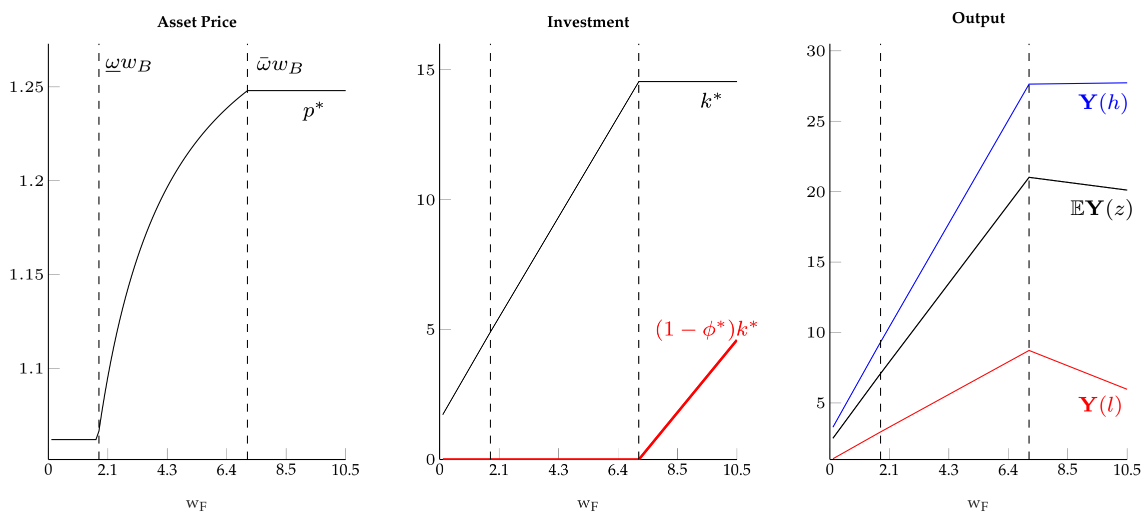

Figure 1: Equilibrium outcomes in the static model as a function of w_F. Parameter values: $\pi_h = 0.65$, $Y_h = 1..9$,$y_h = 1.92$, $Y_l = 0.6$, $y_l = 0$, $m = 0.05$, $q = 1$.

Figure 1 illustrates generic equilibrium outcomes as a function of w_F, holding w_B fixed. The only crucial assumption I make for illustrative purposes is that $\underline{\chi}(q) > \bar{\chi}$ so that $\underline{\omega} < \bar{\omega}$. The vertical dashed lines show the two thresholds for relative financier wealth. The left panel plots the asset price, the center panel plots total investment and investment in bad assets, and the right panel plots output. Below the first threshold, the asset price is constant but investment increases because banks leverage more if they sell a larger fraction of their assets. Since all banks exert effort, output increases in every state. In the intermediate region, investment increases because higher asset prices allow banks to borrow more. Output continues to increase state by state because no bank shirks. Once the upper threshold is reached, the asset price is fixed at $\bar{P}(\bar{\chi})$. Investment is now constant because further increases in financier wealth no longer translate into price increases. To clear markets, a growing number of banks shirk. This leads to a decline in expected output and an increase in downside risk. In contrast to the canonical literature on price-sensitive collateral constraints (e.g. Lorenzoni (2008)), moreover, the equilibrium features a form of *underinvestment*: credit volumes do not increase even though there is a increase in risk-bearing wealth.

It is useful to point out that the model does not need to rely on changes in the wealth distribution to generate declines in asset quality. If ω is sufficiently close to $\bar{\omega}$, for example, then negative shocks to Y_h may be enough to trigger shirking.[8]

Corollary 3 (Productivity Shocks)**.** *Let* $\bar{\chi} < \underline{\chi}(q)$. *Then* $\frac{\partial \bar{\omega}}{\partial Y_h} > 0$.

3.2 Efficiency

I now study the efficiency properties of competitive equilibrium. The key result is that the decline in asset quality at $\omega > \underline{\omega}$ is inefficient and stems from a pecuniary externality. Specifically, I show that allowing financiers to coordinate on a maximum per-capita investment in risky assets generates a strict Pareto improvement.

Proposition 7 (Pecuniary Externality)**.** *Assume that financiers coordinate on an investment policy that prohibits individual financiers from purchasing more than* $\bar{a}$ *risky assets. If* $\omega \in (\bar{\omega}, \infty)$, *then choosing* $\bar{a} = \chi^{\max} k^*(\chi^{\max}) < a_F^*(\chi^{\max})$ *strictly increases expected financier utility and does not lower expected bank utility.*

The basic mechanism is that financiers do not internalize that their demand for risky assets harms average asset quality. A binding cap on investment solves the coordination failure. It is worth pointing out that the pecuniary externality operates in the presence of price bounds. Hence it is reflected in excess demand at a fixed price.

At first pass, the inefficiency result may suggest that shutting down secondary mar-

[8] Similar results obtain with respect to Y_l and π_l. These are ommited for brevity.

kets and redistributing financier net worth to banks would boost investment. The next result shows that this is not necessarily true. Specifically, I provide conditions under which a marginal increase in financier wealth boosts total investment more than a marginal increase in bank net worth.

Proposition 8. *Let $q\hat{Y} > \underline{P}(q)$ and $\omega < \underline{\omega}$. Then $\frac{\partial k^*(\chi^*)}{\partial w_F} > \frac{\partial k^*(\chi^*)}{\partial w_B}$ if $Y_l - \rho > \frac{(1-q)\underline{P}(q)}{q}$. This condition is easier to satisfy if q is large, and it is satisfied if $q = 1$ and $y_h > Y_h$.*

A social planner who cares about investment will therefore opt for a system with both banks and financiers. The intuition is that asset sales provide a form of commitment against bank moral hazard. To see this, recall that ρ defines the pledgeable return on risky assets held by banks, while Y_l is the pledgeabe return on good risky assets held by financiers. If $y_h > Y_h$, then $\rho < Y_l$. Hence financiers can borrow more than banks per unit of investment. The reason is that only banks are subject to moral hazard. Financiers thus emerge as "specialists in leverage."[9] Secondary markets thus allow for sharper increases in credit volumes in the early stages of an expansion.

3.3 Dynamic Implications

To arrive at a theory of credit booms, it is necessary to study the evolution of the wealth distribution over time. Section 4 takes up this task in an infinite-horizon model. Because that model is not analytically tractable, it is useful to first gain intuition by studying the mechanisms driving wealth accumulation in the static model. I do so by characterizing end-of-period relative wealth $\omega'(z)$ conditional on the realized aggregate state z. Since banks and financiers are exposed to the same aggregate risk factor, credit volumes can increase only after good aggregate shocks. I will therefore focus specifically on end-of-period relative wealth given $z = h$. To study the role of leverage, I will assume that financiers strictly prefer to borrow if all banks exert effort ($\bar{P}(\chi^{\max}) < q\hat{Y}$).

Proposition 9. *Let $q\hat{Y} > \bar{P}(\chi^{\max})$. Then $\omega'(z)$ is as follows:*

[9] To see why ω must be small, note that the pass-through from financier wealth to risky asset purchases is strongest when the asset price is at its lower bound.

(i) If $\omega \leq \underline{\omega}$, then $\omega'(z) = \frac{\nu_F(z)\omega}{\nu_B(z)-(Y_z-\hat{Y})\left(\frac{\rho(1-\tilde{m})}{\underline{P}(q)-qY_l}\right)\omega}$, where $\nu_F(z) = \frac{Y_z-Y_l}{\underline{P}(q)-qY_l}$ and $\nu_B(z) = \frac{Y_z-\tilde{m}\rho}{1-q\tilde{m}\rho}$. There exists $\hat{q} < 1$ such that $\omega'(h) > \omega$ for all $\omega > 0$ if $q \geq \hat{q}$.

(ii) If $\omega \in (\underline{\omega}, \bar{\omega}]$, then $\omega'(z) = \frac{\bar{\chi}(Y_z-Y_l)}{Y_z-\tilde{m}\rho-\bar{\chi}(Y_z-\rho)}$ and $\omega'(h) > \bar{\omega}$ if $\frac{Y_h-Y_l}{Y_h-\tilde{m}\rho-\bar{\chi}(Y_h-\rho)} > \frac{\hat{y}-qY_l}{1-q\tilde{m}\rho-\bar{\chi}q(\hat{y}-\rho)}$.

(iii) If $\omega > \bar{\omega}$ and $\gamma^* = 1$, then $\omega'(z) = \frac{\phi^*\bar{\chi}(Y_z-Y_l)+(1-\phi^*)(y_z-y_l)}{\phi^*Y_z+(1-\phi^*)\bar{P}(\bar{\chi})-\tilde{m}\rho-\bar{\chi}(\phi^*Y_z+(1-\phi^*)\bar{P}(\bar{\chi})-\rho)}$, and $\omega'(h) > \frac{\bar{\chi}(Y_h-Y_l)}{Y_h-\tilde{m}\rho-\bar{\chi}(Y_h-\rho)}$ for y_h sufficiently close to Y_h.

The intuition is as follows. If $\omega < \underline{\omega}$, then the asset price is at the lower bound and banks earn no rents on secondary markets. Thus, the utility earned by banks and financiers in state z is equal to $\nu_B(z)w_B$ and $\nu_F(z)w_F$. It follows that the evolution of relative wealth is thus determined by (i) the ratio of returns on equity $\nu_F(z)/\nu_B(z)$, and (ii) the volume of risky assets bought by financiers.(The second channel is reflected in the second term of the denominator in the expression for $\omega'(z)$, which shows that financiers buy more aggregate risk exposure if they are wealthy initially.) If $\omega \to 0$, then end-of-period relative wealth is entirely determined by the ratio of equity returns. Since financiers can lever more than banks at low interest rates, this ratio is strictly higher than 1 after good shocks if q is sufficiently close to 1. Financiers thus grow in relative terms no matter their initial wealth if interest rates are sufficiently low.

If $\omega \in (\underline{\omega}, \bar{\omega}]$, then the aggregate risk exposure of banks and financiers scales linearly with aggregate investment k^*. End-of-period relative net worth thus is a constant determined by relative leverage and the fraction of aggregate risk exposure held by financiers. Relative wealth after good shocks is thus increasing in $\bar{\chi}$, the fraction of assets sold to financiers. Since asset quality falls if and only if $\omega > \bar{\omega}$, the proposition also provides a condition such that end-of-period wealth exceeds $\bar{\omega}$ after a good shock. If $\omega > \bar{\omega}$, then financiers purchase an even larger fraction of aggregate risk exposure. This is because bad banks sell all their assets. This lead to an even sharper increase in relative wealth after good shocks than if $\omega \in (\underline{\omega}, \bar{\omega}]$.

Figure 2 provides an example in which $\omega'(h) > \omega$ for all ω, and $\omega'(h) > \bar{\omega}$ if $\omega \in (\underline{\omega}, \bar{\omega}]$. In a repeated version of the static model, a sequence of good shocks would thus eventually lead to shirking. Interestingly, this is the case precisely because secondary markets efficiently allocate aggregate risk exposure to financiers initially. The eventual decline in asset quality is thus the outcome of an efficient increase in credit volumes, given a suitable sequence of aggregate shocks.

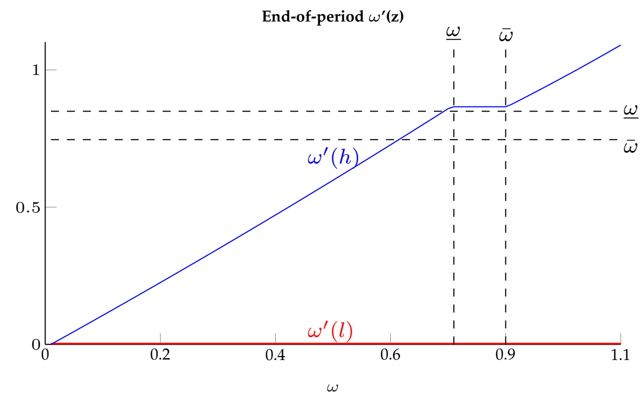

Figure 2: Dynamic implications within the static model. Parameter values: $\pi_h = 0.5$, $Y_h = 2.1$,$y_h = Y_h$, $Y_l = 0.4$, $y_l = 0.075$, $m = 0.08$, $q = 1$.

The fact that financiers hold a disproportionate share of aggregate risk also exposes them to downside risk. Financier wealth thus shrinks disproportionately after bad shocks, lowering future asset demand and credit volumes. This provides a link to the canonical literature on fire sales. Since shirking increases downside risk, an important difference is that *increases* in current wealth may lead to sharper declines in future wealth.

Corollary 4 (Implications for Fire Sales)**.** *Let $\omega > \bar{\omega}$ and $\phi^* > \underline{\phi}(\bar{P}(\chi^{\max}), \chi^{\max})$. Assume that financiers do not borrow ($\gamma^* = 0$). Then $w'_F(l)$ is strictly decreasing in w_F.*

4 Infinite Horizon Model

I now embed the static model in an infinite horizon model to study the endogenous evolution of the wealth distribution. I show below that this dynamic model can generate periods of growing credit volumes with increasing relative financier net worth and a growing fraction of shirking banks, akin to a credit boom with deteriorating credit quality. Time is discrete and indexed by $t = 0, 1, \ldots$. Banks and financiers are infinitely-lived and are endowed with initial net worth $w_{B,0}$ and $w_{F,0}$, respectively. They discount the future using discount factor $\beta \in (0,1)$. To focus on bank and financier incentives, I assume that savers live for one period, and that each saver generation is endowed with fixed net worth w_S.

The wealth constraint on savers implies that the risk-free interest rate is endogenously determined. This allows the model to generate predictions for the joint dynamics of the asset price and the risk-free rate. It also provides a counterweight to sustained financier growth, since higher interest rates make it harder for financiers to exploit leverage advantage. Since savers are short-lived, there are no dynamic effects stemming from wealth accumulation by savers. This allows for a clean analysis of the evolution of financier and bank wealth. It also keeps the state space small.

At the beginning of each period, banks originate assets with the same payoff structure as in the static model. Assets fully depreciate at the end of each period.[10] The aggregate shock is i.i.d. over time and is realized at the end of each period. This implies that the relevant state variables are the aggregate wealth distribution and own net worth. The private benefit of shirking is βm. As in Gertler and Karadi (2011), banks and financiers accumulate net worth over time, but may be forced to exit at the end of each period with probability $1 - \psi \in [0,1]$. In the event of an exit, they consume their net worth and are replaced by a new bank or financier. If they do not exit, they proceed to the period. For simplicity, entrants are endowed with the same net worth as those who exit. The value of a dollar of net worth will generically fluctuate with the endogenous aggregate wealth distribution $\mathbf{w} = \{\mathbf{w}_B, \mathbf{w}_F, \mathbf{w}_S\}$. This generates precautionary motives whereby financiers and banks want to carry net worth into states of the world where it is scarce in the aggregate. As in the static model, the only securities are one-period riskless bonds and risky assets traded in anonymous secondary markets. Hence there are no long-term contracts

[10] An alternative would be to assume that assets are long-lived. Given that assets can be good or bad and can be held by banks and financiers, this would introduce four additional endogenous state variables and greatly reduce the tractability of the dynamic model. It would also exacerbate the ineffiency. To see this, note that shirking does not necessarily lead to lower output if the aggregate state is good. With full depreciation, a single realization of the good shock may thus be enough to avoid the adverse consequences of shirking. This is not true if assets are long-lived, since shirking then leads to a persistent decline in quality of the asset stock, exposing the economy to additional downside risk at all future dates.

or reputational concerns.[11] The definition of equilibrium is the same as in the static model, but is augmented with the bond market-clearing condition $b_S(\mathbf{w}) = b_F(\mathbf{w}) + b_F(\mathbf{w})$, where b_S denotes savers' bond purchases.

4.1 Value Functions

I now briefly describe bank and financier decision problems in the infinite-horizon model. Most of the analysis carries over from the static model with small modifications. Let the optimal bank policy be $\{k^*, b^*, \underline{a}^*, a_B^*, e^*\}$, where $e^* \in \{0,1\}$ is the effort decision. Given asset price p, the bank value function can be written as

$$
\begin{aligned}
V_B(w_B, \mathbf{w}) = \quad & \mathbb{E}_z \beta \Big[(1-\psi) w_B'(z) + \psi V_B(w_B'(z), \mathbf{w}'(z)) \Big] + \beta m e^* k^* \\
\text{s.t.} \quad & w_B'(z) = \max \Big\{ (e^* Y_z + (1-e^*) y_z)\,(k^* - a_B^*) - b_B^* + p a_B^*, 0 \Big\} \\
& \mathbf{w}'(z) = \Gamma(\mathbf{w}, z).
\end{aligned}
$$

where $\Gamma(\cdot)$ is the law of motion for the aggregate net worth distribution. Guess and verify that the value function is linear in w_B, $V_B(w_B, \mathbf{w}) = v_B(\mathbf{w}) w_B$ for some unknown function $v_B(\mathbf{w})$, and define $\mu_B(\mathbf{w}) = (1-\psi) + \psi v_B(\mathbf{w})$. Finally, denote the (unique) active submarket given w by $\chi^*(\mathbf{w})$ and the fraction of bad banks by $\phi(\mathbf{w})$.

Definition 6. *Conditional on* $\mathbf{w}$*, each bank's expected value of a good and bad risky asset is, respectively,* $\tilde{Y}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z)) Y_z$ *and* $\tilde{y}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z)) y_z$*, and the expected value of one unit of a riskless asset is* $\tilde{\mu}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z))$*. The pledgeable return is* $\rho(\mathbf{w}) = \frac{\tilde{Y}_B(\mathbf{w}) - \pi_h \mu_B(\mathbf{w}'(h)) y_h}{\pi_l \mu_B(\mathbf{w}'(l)}$*. The moral hazard discount is* $\tilde{m}(\mathbf{w}) = 1 - m \cdot \left(\tilde{Y}_B(\mathbf{w}) - \pi_h \mu_B(\mathbf{w}'(h)) y_h \right)^{-1}$*.*

Proposition 10. *The borrowing and retention constraints are* $b_B \leq \rho(\mathbf{w}) \tilde{m}(\mathbf{w}) k + (P(\chi) - \rho(\mathbf{w})) a_B$ *and* $a_B \leq \bar{\chi}(\mathbf{w}) k$*, respectively, where* $\bar{\chi}(\mathbf{w}) = \left(\frac{\tilde{m}(\mathbf{w}) \rho(\mathbf{w}) - y_l}{\rho(\mathbf{w}) - y_l} \right)$*. The upper and lower*

[11] Reputations can be thought of as a mechanism to overcome anonymity in financial markets. Yet reputations are typically fragile (see e.g. Ordoñez (2013)) and, in dynamic settings, may even serve to *sustain* pooling equilibria in which both low-quality and high-quality assets are sold in secondary markets (Chari, Shourideh, and Zetlin-Jones (2014)). Empirically, Griffin, Lowery, and Saretto (2014) show that high-reputation issuers produced and sold lower-quality asset-backed securities during the run-up to the 2008 financial crisis than low-reputation issuers. This suggests that strong secondary market demand may provide the impetus to begin milking reputations. As a result, reputational concerns, at best, partially attenuate and, at worst, strengthen motives for harmful hidden trading in secondary markets.

bounds on the asset price are $\bar{P}(\chi) = \frac{1}{\tilde{\mu}_B(\mathbf{w})}\left(\tilde{Y}_B(\mathbf{w}) - \frac{m}{1-\chi}\right)$ *and* $\underline{P}(q,\mathbf{w}) = \frac{\tilde{Y}_B(\mathbf{w})+\rho(\mathbf{w})\Pi(q,\mathbf{w})}{\tilde{\mu}_B(\mathbf{w})+\Pi(q,\mathbf{w})}$*, respectively, where* $\Pi(q,\mathbf{w}) = \frac{q\tilde{Y}_B(\mathbf{w})-\tilde{\mu}_B(\mathbf{w})}{1-q\tilde{m}\rho(\mathbf{w})}$.

Now turn to financiers. Their value function is

$$
\begin{aligned}
V_F(w_F,\mathbf{w}) = \max_{a_F\geq 0, b_F \geq 0, s_F \geq 0} \quad & \mathbb{E}_z\beta\Big[(1-\psi)w'_F(z) + \psi V_F(w'_F(z),\mathbf{w}')\Big] \\
\text{s.t.} \quad & s_F + pa_F = w_F + qb_F && \text{(Budget constraint)} \\
& b_F \leq x_l(\mathbf{w})a_F + s_F && \text{(Borrowing constraint)} \\
& w'_F(z) = s_F + x_z(\mathbf{w})a_F - b_F && \text{(Individual law of motion)} \\
& \mathbf{w}'(z) = \Gamma(\mathbf{w},z) && \text{(Aggregate law of motion)}
\end{aligned}
$$

As with banks, guess and verify that the value function is linear w_F, $V_F(w_F,\mathbf{w}) = v_F(\mathbf{w})w_F$ for some unknown function $v_F(\mathbf{w})$, and define $\mu_F(\mathbf{w}) = (1-\psi) + \psi v_F(\mathbf{w})$.

Definition 7. *A financier's expected value conditional on* $\mathbf{w}$ *of a good risky asset, a bad risky asset, and an asset purchased in submarket* $\chi^*(\mathbf{w})$ *are, respectively,* $\tilde{Y}_F(\mathbf{w}) \equiv \mathbb{E}_z\mu_F(\mathbf{w}'(z))Y_z$, $\tilde{y}_F(\mathbf{w}) \equiv \mathbb{E}_z\mu_F(\mathbf{w}'(z))y_z$*, and* $\tilde{x}_F(\mathbf{w}) \equiv \mathbb{E}_z\mu_F(\mathbf{w}'(z))x_z$*. The expected value conditional on* $\mathbf{w}$ *of one unit of a riskless asset is* $\tilde{\mu}_F(\mathbf{w}) \equiv \mathbb{E}_z\mu_F(\mathbf{w}'(z))$.

Proposition 11. *Financiers are willing to purchase risky assets if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} \geq p$*. This condition holds if and only if* $\phi(\chi^*(\mathbf{w})) \geq \underline{\phi}\left(p,\chi^*(\mathbf{w})\right) \equiv \frac{\tilde{\mu}_F(\mathbf{w})p-\tilde{y}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})p-\tilde{y}_F(\mathbf{w})+\chi^*(\mathbf{w})\left(\tilde{Y}_F(\mathbf{w})-\tilde{\mu}_F(\mathbf{w})p\right)}$*. Financiers strictly prefer to leverage if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} > \frac{p}{q}$ *and are indifferent with respect to leverage if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} = \frac{p}{q}$.

Given the marginal value of dollar of net worth, the decision problems are thus identical to the static model, and optimal portfolios can be derived as before. I thus omit a detailed analysis here. The static model showed that the equilibrium can be solved in closed form conditional on q. The same is true here (given a guess for the value functions and the aggregate law of motion). This simplifies the computations.

Remark 2. *The double deviation to shirking and selling all assets lowers banks' risk exposure. Hence banks with precautionary motives are more likely to deviate than risk-neutral banks, and the upper bound on the asset price declines. Financiers with precautionary motives are less willing* *to buy low-quality assets at a given price than risk-neutral financiers. Nevertheless, there is scope for shirking because banks' precautionary motives lead to lower asset prices. Since precautionary motives decrease in own net worth, shirking is more likely when financiers are relatively wealthy.*

4.2 Saver portfolio

An important difference to the static model is that savers now have limited wealth, which they can use to buy risk-free debt or invest in the safe technology. The saver budget constraint is $qb_S + s_S = w_S$, where b_S is the number of bonds purchases and s_S is investment in the safe technology. (There is no need to differentiate between bank and financier debt because they are perfect substitutes.) The optimal saver portfolio then is as follows.

Proposition 12. *$b^*_S = w_S/q$ and $s_S = 0$ if $q < 1$, and $b^*_S \in [0, w_S/q]$ and $s^*_S = w_S - qb^*_S$ if $q = 1$.*

4.3 Computational Algorithm and Equilibrium Value Functions

I now describe the computional algorithm. Given that we can focus without loss of generality on symmetric portfolios and z is i.i.d., state variable consists of the wealth distribution own net worth. Given that we focus on symmetric equilibrium, we can compute the value function on discrete grids for $\mathbf{w}_B$ and $\mathbf{w}_F$. Solving the model presents two challenges. The first is that the wealth distribution evolves as a function of aggregate risk. The second is that the law of motion for $\mathbf{w}'_F$ need not be monotone in $\mathbf{w}_F$. To tackle these issues, I use the following algorithm loosely based on Krusell and Smith (1998).

1. Fix a discrete grid $[\epsilon, \bar{\mathbf{w}}_B] \times [\epsilon, \bar{\mathbf{w}}_F]$ for bank and financier wealth, where $\epsilon > 0$.
2. Guess value functions coefficents $v^0_B(\cdot), v^0_F(\cdot)$ and an aggregate law of motion $\Gamma^0(\cdot)$.
3. For each point on the grid, compute the static equilibrium given $\{v^0_B(\cdot), v^0_F(\cdot), \Gamma^0(\cdot)\}$.
4. Use optimal bank and financier portfolios to update the law of motion to $\Gamma^0_{updated}(\cdot)$.
5. Iterate to covergence on the law of motion, using the rule $\Gamma^1(\cdot) = \alpha_{lom}\Gamma^0_{updated}(\cdot) + (1 - \alpha_{lom})\Gamma^0(\cdot)$ to update the intial guess at the beginning of each iteration. Here

$\alpha_{lom} \in (0, 1)$ is the weight placed on the update.

6. Use the converged law of motion optimal portfolios to compute updated value function coefficients $\{v^0_{B,updated}(\cdot), v^0_{F,updated}(\cdot)\}$.

7. Iterate to convergence on $\{v^0_{B,updated}(\cdot), v^0_{F,updated}(\cdot)\}$ given $\Gamma^1(\cdot)$ using the updating rule $v^1_\theta(\cdot) = \alpha_v v^0_{\theta,updated}(\cdot) + (1 - \alpha_v) v^0_\theta(\cdot)$ for $\theta \in \{B, F\}$.

I repeat these steps until the value function and the law of motion have jointly converged. I update slowly ($\alpha_v, \alpha_{lom} < 1$) due to the non-monotonicity of the law of motion for $\mathbf{w}_F$.

Figure 3 plots representative cuts of the bank value function coefficients and the law of motion for $\mathbf{w}'_B$ as a function of $\mathbf{w}_B$, given a high and a low value of $\mathbf{w}_F$. The left

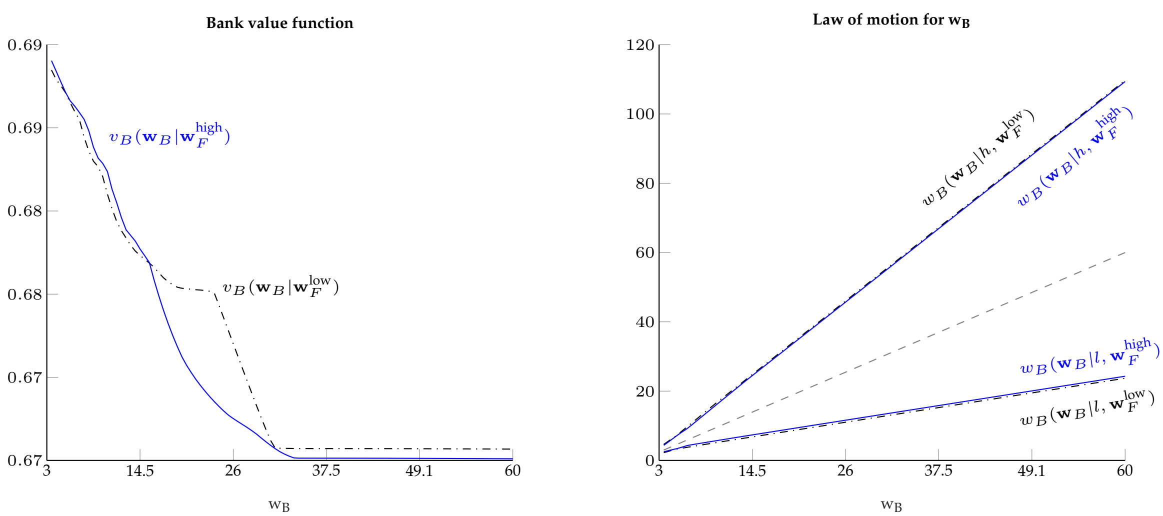

Figure 3: Bank value function coefficients and law of motion for bank wealth. $v_B(\cdot)$ and $w'_B(z)$ are plotted as a function of $\mathbf{w}_B$ conditional on two levels of financier wealth. Parameters as in Figure 5 below.

panel shows that the value function coefficient v_B is monotonically decreasing in $\mathbf{w}_B$. It is not monotone in $\mathbf{w}_F$, because asset sales allow banks to increase leverage the borrowing constraint is tight. If aggregate net worth is high, however, then financiers compete away bank intermediation rents by also issuing debt to savers. The right panel shows that banks retain less risk exposure when financiers are relatively wealthy. This reduces the variance of bank net worth across aggregate states.

Figure 4 plots representative cuts of the financier value function coefficients and the law of motion for $\mathbf{w}'_F$ as a function of $\mathbf{w}_F$, given a high and a low value of $\mathbf{w}_B$. The left

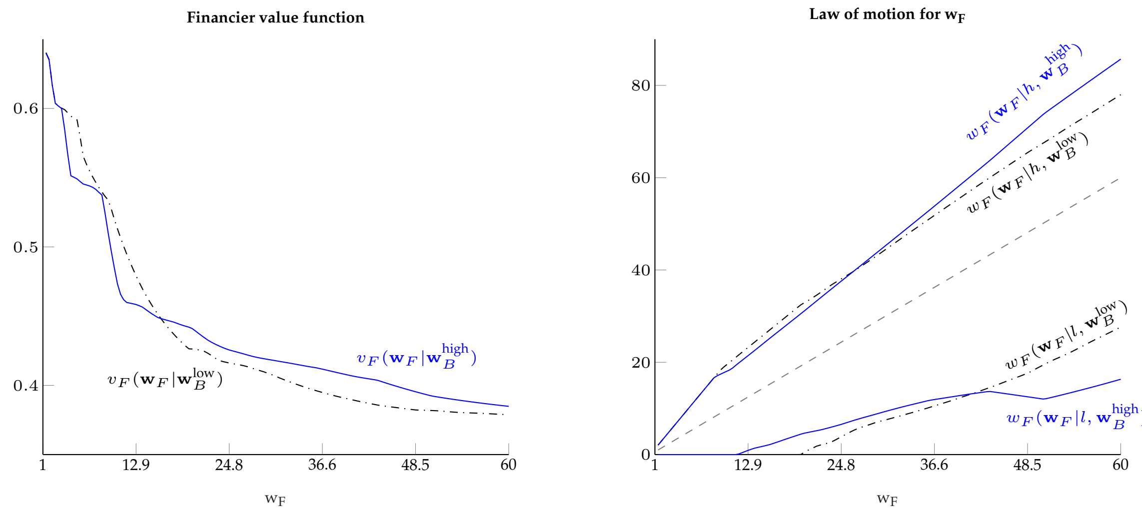

Figure 4: Financier value function coefficients and law of motion for financier wealth. $v_F(\cdot)$ and $w_B'(z)$ are plotted as a function of $\mathbf{w}_F$ conditional on two levels of bank wealth. Parameters as in Figure 5 below.

panel plots value function coefficients v_F as a function of aggregate bank wealth $\mathbf{w}_F$ for high and low values of aggregate bank wealth. Due to competition among financiers, the value function coefficient is monotonically decreasing in $\mathbf{w}_F$, but it need not be monotone in $\mathbf{w}_B$. The solid line on right panel demonstrates the non-monotonicity of the law of motion $\mathbf{w}_F$. The reason is that banks shirk when financiers are sufficiently wealthy. Since bad assets carry more downside risk, this effect leads to an increase in financier risk exposure and a decline in financier wealth after a bad shock (see Corollary 4). Eventually, the fraction of shirking banks reaches its maximum level and $\mathbf{w}_F'(l)$ again increases in $\mathbf{w}_F$.

4.4 Simulations

I now show that the dynamic model can generate periods of growing credit volumes with increasing relative financier net worth and a growing fraction of shirking banks, akin to a credit boom with deteriorating credit quality. To do so, I study equilibrium outcomes given a sequence of aggregate shocks and an initial wealth distribution. Since credit volumes increase only if net worth increases, I focus on a sequence of good aggregate shocks. Since banks and financiers are exposed to the same risks, $\mathbf{w}_F$ and $\mathbf{w}_B$ are positively correlated. Asset prices and secondary market volumes increase only if the *relative* wealth of financiers increases, however.

Figure 5 shows equilibrium outcomes given a sequence of eight good aggregate shocks followed by two negative shocks. (Note that shocks are realized at the *end* of each period, while wealth distribution is depicted at the *beginning* of the period.) Initial relative wealth is $\omega_0 = 0.5$. In the example at hand $\omega_0 < \underline{\omega}$, and prices are initally given by $q = 1$ and $p = \underline{P}(q)$. The top left panel shows that ω increases after good shocks, but crashes after the initial bad shock. The mechanism is that financiers exploit their leverage advantage to grow are highly exposed to downside risk as a result. The top right panel shows that the asset price and the risk-free rate rate rise during the boom. The bottom left panel plots total investment and investment by banks who shirk. The fraction of shirking banks grows steadily due to the increase in ω. Finally, the bottom right panel plots output, with the dashed line showing counterfactual output given the same investment but no shirking. Shirking thus leads to excess downside risk.

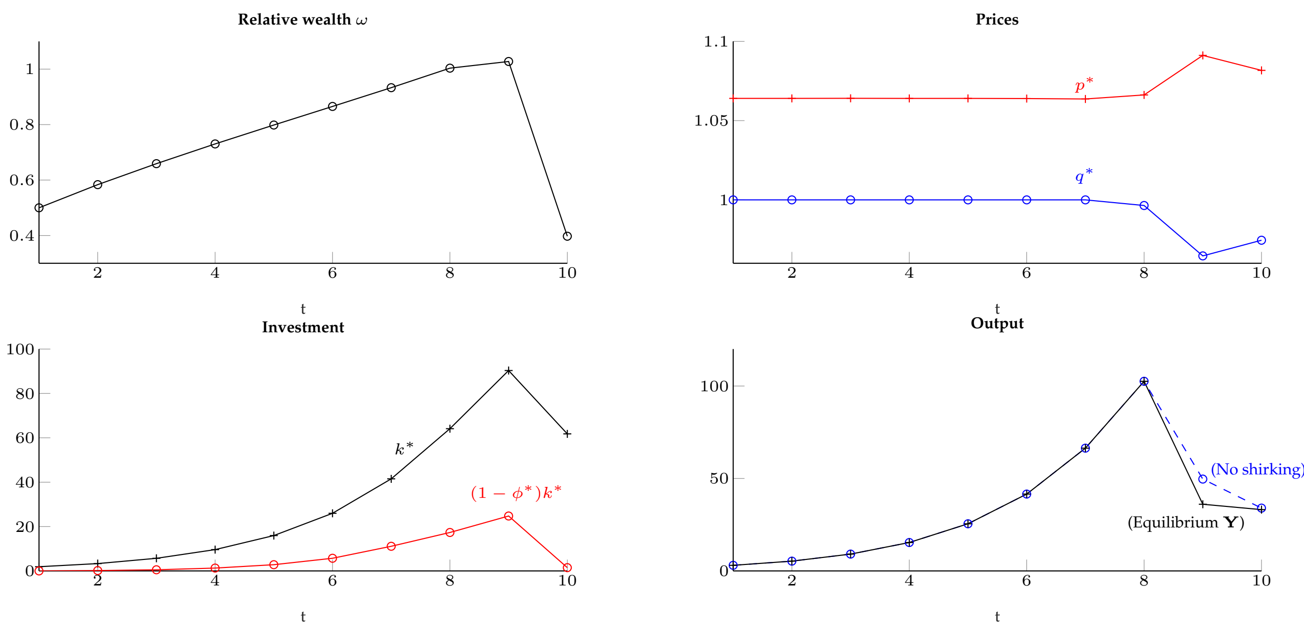

Figure 5: Equilibrium outcomes in the dynamic model given a sequence of eight good aggregate shocks and one negative shock. Parameters: $\pi_h = 0.65$, $Y_h = 1.6$, $y_h = 1.6$, $Y_l = 0.55$, $y_l = 0$, $m = 0.05$, $\beta = 0.95$, $\psi = 0.8$. Saver wealth: $w_S = 40$. Initial wealth distribution: $w_{B,0} = 0.8$, $w_{F,0} = 0.4$.

Figure 6 shows the importance of the initial wealth distribution. All parameters are the same as in Figure 5. The only difference is that $w_{B,0} = 3.2$ so that $\omega_0 = 0.125$. Relative financier wealth grows while the risk-free rate remains low. However, growing $\mathbf{w}_B$ leads to an appreciating risk-free rate, reducing financiers' leverage advantage and leading to a decline in ω. As a result, there is no shirking in equilibrium. By period eight, banks are no longer borrowing constrained and secondary markets are inactive. Asset prices rise only because unconstrained banks are not willing to sell below expected value. Since financiers do not purchase any assets, banks retain all aggregate risk exposure and ω increases after a *bad* shock. Indeed, $\omega_{10} > \omega_0$. Bad shocks may thus produce a wealth distribution that is conducive to future credit booms with falling asset quality.

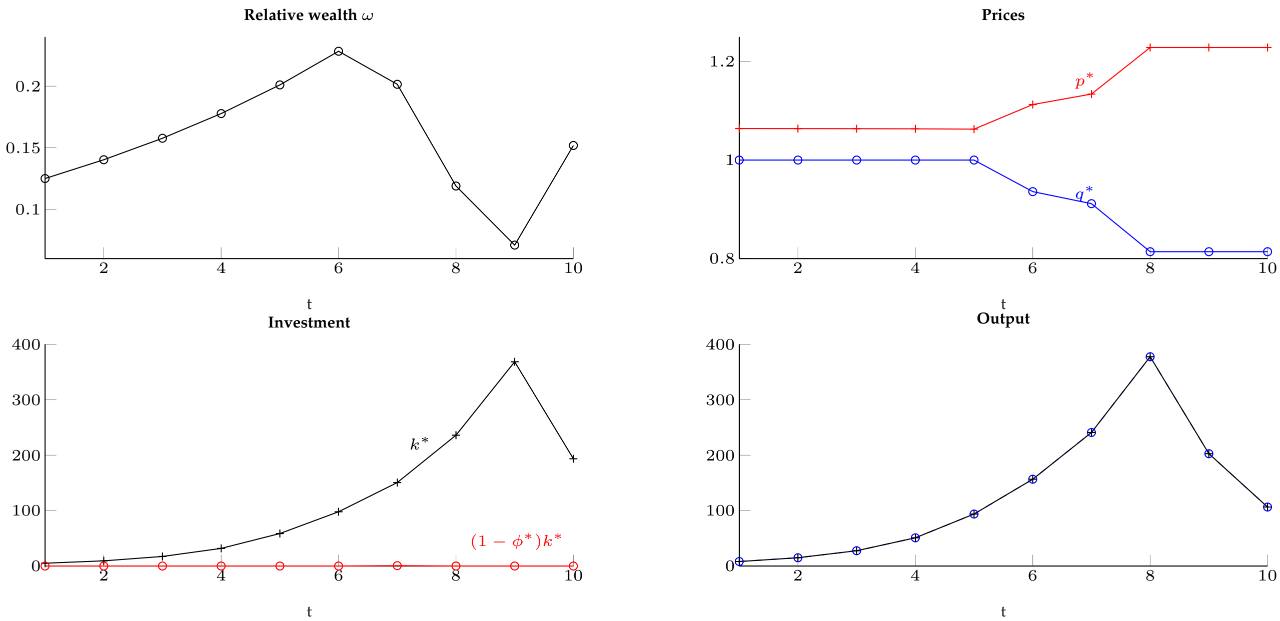

Figure 6: Equilibrium outcomes in the dynamic model given a sequence of eight good aggregate shocks and two negative shocks. Parameters and saver wealth as in Figure 5. Initial wealth distribution: $w_{B,0} = 3.2$, $w_{F,0} = 0.4$.

Remark 3. *The likelihood of credit booms with falling asset quality is largely determined by initial conditions. If ω is large to begin with, then a single positive shock may be enough to trigger shirking. If ω is small to begin with, then banks shirk only after a sufficiently long (and thus unlikely) sequence of good shocks. However, if banks initially hold a large share of risky assets, then a negative shock may increase the likelihood of a subsequent boom with falling asset quality (see Figure 6). If ω grows after good shocks, then ϕ is decreasing in the duration of the boom. Large crises thus occur with lower probability than small crises.*

Figure 7 shows the importance of saver demand (saving gluts) in triggering credit booms with falling asset quality. Parameters and the initial wealth distribution are the

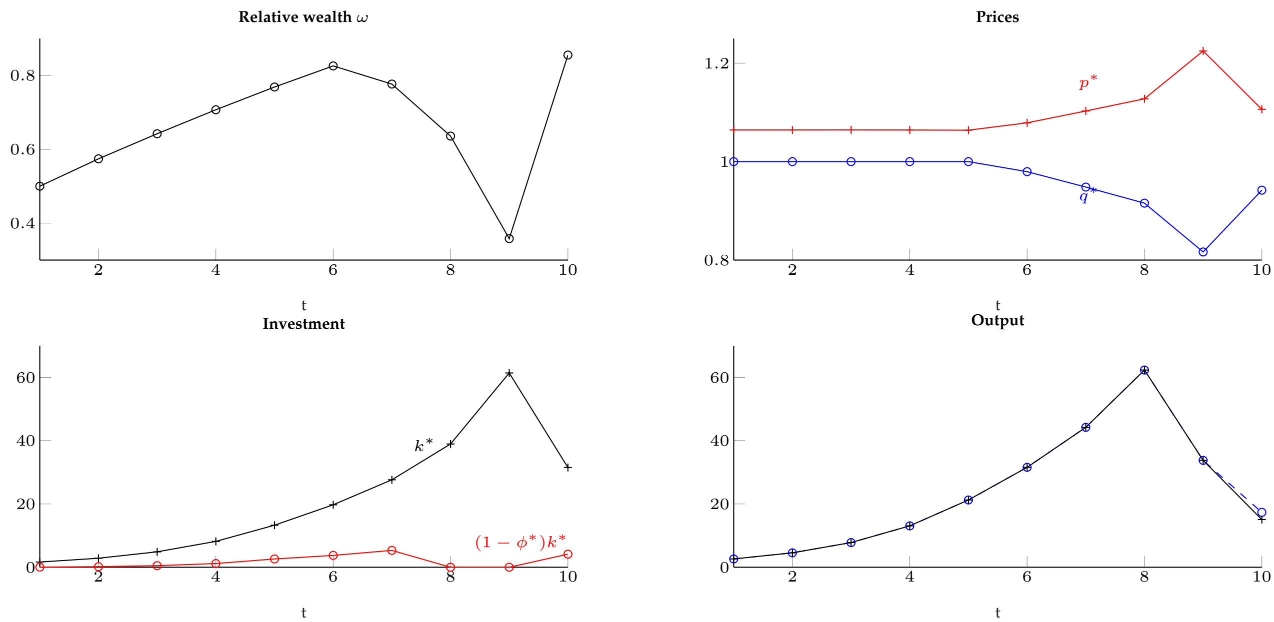

Figure 7: Equilibrium outcomes in the dynamic model given a sequence of eight good aggregate shocks and two negative shocks. Parameters and initial wealth distribution as in Figure 5. Saver wealth: $w_S = 10$.

same as in Figure 5. The difference is that saver wealth is 10 instead of 40. As a result, q falls more rapidly and ω grows more slowly. Fewer banks shirk and banks retain a larger fraction of total risk exposure. The asset price increases more than in Figure 5 only because banks demand a higher minimum price when q is low. However, secondary market volumes remain low.

4.5 Comparison

I now briefly contrast the present model's predictions with the extant literature on price-sensitive collateral constraints (e.g. Mendoza (2010), Gertler and Karadi (2011), Lorenzoni (2008), and Davila and Korinek (2018)). In these papers, agents take excessive risks ex-ante because they do not internalize that their risk-taking worsens fire-sales after bad shocks. While the production technology is exogenous in these models, one could extend them to include a choice of technology along a risk-return frontier. Since borrowing constraints are less likely to bind when agents are wealthy, agents may then switch to a risky but more productive technology after a sequence of good shocks, leading to more aggregate risk-taking. There are important normative and positive differences between such a model and the one presented here. The key normative difference is that shirking simultaneously reduces expected productivity and increases risk exposure. As such, banks only shirk if they can sell bad quality assets to others. This has differential implications for policy. In the canonical theory, risk-taking can be managed via capital requirements or debt taxes (Bianchi and Mendoza (2013)). Here, Proposition 14 shows that bank capital requirements may exacerbate the inefficiency, and Proposition 13 suggests that it may be efficient to regulate unlevered investors (financiers) rather than banks. The key positive difference is as follows. In the canonical theory, risk-taking is not privately inefficient. Here, it is shirking at the expense of other investors. My theory thus predicts lower (rather than higher) expected productivity and increased fraud and misrepresentations during credit booms. This is true empirically (see Empirical Prediction 3 below.)

5 Empirical Predictions and Evidence

This section describes the model's empirical content. The first step is to identify empirical counterparts to banks, financiers, and savers. In the model, banks are distinguished from financiers by the fact that they originate claims on the real economy. Financiers instead purchase assets produced by banks, but do not themselves originate financial claims on the real sector. I therefore define banks' empirical counterparts to be equity holders of financial institutions who originate loans to the real economy, such as commercial banks, mortgage companies, and credit unions. I define financiers to be the equity holders of financial institutions who primarily buy and hold existing claims on the real sector, such as hedge funds, broker dealers, asset managers, pension funds, mutual funds, and insurance companies. (This definition implies a close link between financiers and what is commonly referred to as the "shadow banking system".) I define savers to be creditors to financial institutions, such as depositors, corporations with excess cash, and sovereign wealth funds.[12] In line with the literature (e.g. Mendoza and Terrones (2012) and Schu-

[12] Egan, Hortacsu, and Matvos (2017) show that half of all deposits in large US commercial banks are not covered by deposit insurance, and provide evidence that depositors are sensitive to bank default risk. Deposit insurance alone is thus not sufficient to eliminate borrowing constraints.

larick and Taylor (2012)), I define credit booms to be periods of increasing gross credit volumes. In the model there is a monotone mapping from gross credit volumes ($b_B + b_F$) to investment k. I thus use k as the model's basic measure of gross credit. Due to the wealth of evidence available for credit boom and bust centered on the 2008 Financial Crisis, I focus primarily on this episode. I also point out which empirical regularities hold more broadly.

Prediction 1. *The relative size of financier to bank balance sheets increases during credit booms.*

Evidence. Precise estimates for the balance sheet size of all non-bank intermediaries are hard to obtain. However, 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. The 2016 Global Shadow Banking Monitoring Report of the Financial Stability Board[13] compiles aggregate balance sheet data for 21 advanced economies and the Euro Area. It finds that the share of global financial assets held by shadow banks steadily increased from 25.8% in 2002 to 29.2% in 2007, while the share held by commercial banks remained constant (43% in 2002 and 43.5% in 2007).

Prediction 2. *Credit booms coincide with secondary market asset price and trading volume booms.*

Evidence. Gorton and Metrick (2012b), 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. Chernenko, Hanson, and Sunderam (forthcoming) provide evidence of declining yield spreads for mortgage-backed securities and collateralized debt obligations from 2003 to 2007. 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 loan syndication booms for the Great Depression and Asian Financial Crisis, respectively.

[13] Retrieved from FSB website on December 10, 2017. Link.

Prediction 3. *The average quality of newly originated assets declines during credit booms, and banks increasingly engage in moral hazard during booms. The average quality of assets retained by originators is higher than that of assets sold on secondary markets.*

Evidence. Keys, Mukherjee, Seru, and Vig (2010), Piskorski, Seru, and Witkin (2015), and Griffin and Maturana (2016) study mortgage lending and provide evidence of falling credit standards and increasing fraud and misrepresentations during the 2000-2007 U.S. credit boom. Griffin and Maturana (2016) show that fraud is concentrated among securitized loans. Purnanandam (2011) shows that originate-to-distribute banks originated and securitized excessively poor-quality mortgages prior to 2007, and argues that banks did not screen borrowers. Consistent with the notion that banks shirked under the guise of efficient risk reallocation, he finds stronger effects among capital-constrained banks. Downing, Jaffee, and Wallace (2009) show that mortgage-backed securities sold to special-purpose vehicles were subject to private information and of lower quality.

Prediction 4. *At the height of the boom, financiers are disproportionately exposed to downside risk. Hence financier balance sheets contract more sharply than bank balance sheets during crises.*

Evidence. Greenlaw, Hatzius, Kashyap, and Shin (2008) show that non-bank intermediaries who purchased mortgage-backed securities were more exposed to downside risk than loan-originating banks. Brunnermeier (2009) documents run in 2007-2008 on secondary market traders exposed to mortgage-backed securities, such as Lehman Brothers, Bear Stearns and Washington Mutual.

Prediction 5. *Credit booms with falling asset quality are more likely to occur after an inflow of savings or during periods of expansionary monetary policy.*

Evidence. Mendoza and Quadrini (2010) show that US foreign credit market borrowing rose from approximately 30% of GDP in 2000 to approximately 60% of GDP in 2005. Caballero and Krishnamurthy (2009) argue that monetary policy was expansionary during the early stages of the pre-2008 U.S. credit boom.

Prediction 6. *Credit booms driven by secondary market trading are a predictor of crises, and longer booms predict sharper crises. Bank equity returns fall during credit booms.*[14]

Evidence. Jorda, Schularick, and Taylor (2011) argue that credit booms are the best predictor of financial crises. Schularick and Taylor (2012), Mendoza and Terrones (2012), and Reinhart and Rogoff (2009) show that longer credit booms predict sharper crises. Baron and Xiong (2017) show that bank credit expansions forecast bank equity crashes and declining bank equity returns.

6 Policy Implications

This section discusses the theory's positive implications for policy. The goal is to shed light on the novel motives for regulation that arise in the model, and to study the impact of various real-world policies. For simplicity, I do so using the static model. I study policies designed to alleviate the pecuniary externality, and those which might contribute to credit booms in the first place. An analysis of optimal policy is left for future work.

6.1 Alleviating the Inefficiency

The basic source of inefficiency is that some banks sell too many assets when there is strong demand. There are two ways to address this. The first is to prevent banks from selling too many assets. The second is to reduce asset demand. In the context of the simple model presented here, both approaches lead to equivalent outcomes.

Proposition 13. *Shirking can be eliminated by (i) a skin-in-the-game rule $\chi \leq \chi^{\max}$, or (ii) a restriction of financier asset demand of the form $a_F \leq \bar{a}_F = \chi^{\max} k(\chi^{\max})$. Both lead to the same equilibrium allocation.*

While there are no differences between these policies in the model, there are differences in their real-world application. Skin-in-the-game rules require appropriate measures of

[14] Section 4.4 shows that q declines during booms. Hence intermedition rents (and thus bank excess returns) decline during credit expansions.

bank-level risk exposure for each asset class, and regulators must be able to observe bank trades in real time and sufficient detail. Hence they are informationally intensive. Restrictions on asset demand can be implemented by taxing trades, regulating the asset side of financier balance sheets, or directly redistributing financier equity. While these tools may be less informationally intensive than skin-in-the-game rules, they would ideally be employed pro-cyclically, and would require the regulation of almost all financial institutions. This leads to additional complexity. Notably, the motive for regulating financiers is entirely independent of leverage. Hence it applies equally to unlevered asset managers such as BlackRock, Inc. that recently escaped tighter regulatory attention.

6.2 Expansionary Monetary Policy and Saving Gluts

I now argue that monetary policy is a plausible trigger for credit booms with falling asset quality. I interpret a monetary expansion simply as a reduction in the risk-free rate (an increase in the bond price). The argument follows directly from Proposition 9, which states that there exists a cutoff bond price $\hat{q}$ such that, given any $\omega \leq \underline{\omega}$, end-of-period relative wealth grows after good shocks if $q \geq \hat{q}$. An immediate corollary is that there exists such a cutoff bond price for each initial relative wealth $\omega \leq \underline{\omega}$.

Corollary 5. *If $\omega \leq \underline{\omega}$, then there exists a cutoff bond price $q^C(\omega) < \hat{q}$ such that $\omega'(h) > \omega$ if $q \geq q^C(\omega)$, and $q^C(\omega)$ is strictly decreasing.*

Observation 2. *If $q < q^C(\omega)$ initially, then a sufficiently large reduction in the risk-free rate leads to an increase in relative financier wealth after good shocks. Monetary expansions can thus trigger credit booms with growing relative financier wealth that would have not occured otherwise.*

Lower borrowing costs thus lead to the relative growth of financiers, which in turn boosts asset prices and eventually triggers bank shirking. This can be interpreted as a dynamic risk-taking channel of monetary policy. It should be noted that this mechanism is active after any shock that triggers a reduction in the risk-free rate. A global saving gluts that leads to an inflow of savings (an increase in w_S) is thus an equally plausible trigger of credit booms with falling investment efficiency. Importantly, even *temporary* shocks to the interest rate may be sufficient. To make this point, let the bond price prior to the monetary expansion be q^*, and let q^{MP} denote the bond price during the expansion. Finally, let $\tilde{w}(z|q^{MP})$ denote relative wealth at the end of the monetary expansion.

Observation 3. *Suppose $q^* < q^C(\omega)$. If $q^{MP} > q^C(\omega)$ and $q^* > q^C(\tilde{w}(h|q^{MP}))$, then relative financier wealth continues to grow after good shocks after the monetary policy expansion ends.*

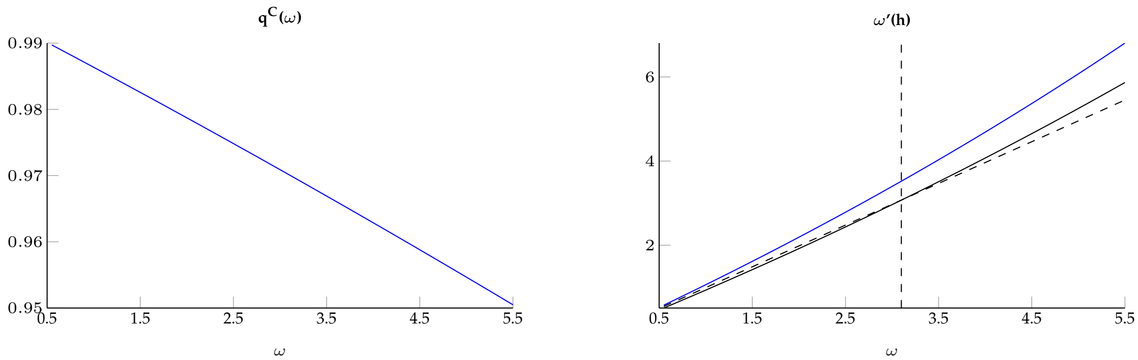

Figure 8: Persistent effects of monetary policy or saving gluts. Parameters: $\pi = 0.65, Yh = 2.5, y_h = 2.5, Y_l = 0.815, y_l = 0, m = 0.03$. Bond prices: $q^* = 0.97$, and $q^{MP} = 0.99$.

Figure 8 illustrates this mechanism graphically. The left panel depicts the downward-sloping cutoff price. The right panel depicts the evolution of relative net worth after good shocks. The lower solid line plots $\omega'(h)$ in the absence of monetary policy, the upward sloping dashed line is the 45-degree line, and the vertical dashed line plots the cutoff level for initial relative wealth ω_0 such that $\omega'(h) \geq \omega_0$ in the absence of monetary policy. Relative financier wealth thus contracts to the left of the dashed line but expands to its right. The upper solid line plots $\omega'(h)$ during the monetary expansion. Relative wealth now grows for any ω_0. If ω crosses the dashed line during the expansion, then relative financier wealth grows even if the stimulus is removed. This suggests an asymmetry in the conduct of monetary policy: the policy adjustment required to reign in a credit boom once it is underway may be larger than the initial expansion.

6.3 Bank Capital Requirements

Lastly, I turn to the effects of bank capital requirements, which have been identified as an effective tool for managing overborrowing in canonical models with fire sales (e.g. Lorenzoni (2008)). I model these as restricting total bank investment to a multiple of net worth, $k \leq \lambda^{CR} w_B$. Since many non-bank financial institutions were not subject to regulation prior to 2008, I assume that financiers are exempt. I denote equilibrium outcomes given capital requirements by superscript CR. I assume for simplicity that $\chi^{\max} = \bar{\chi}$ and that financiers' borrowing constraint binds if no bank shirks ($q\hat{Y} > \bar{P}(\bar{\chi})$).

Proposition 14. *Let* $\chi^{\max} = \bar{\chi}$, $\omega \in [\underline{\omega}, \bar{\omega})$ *and* $q\hat{Y} > \bar{P}(\bar{\chi})$. *Hence* $\phi^* = 0$. *If* $k^* > \lambda^{CR} w_b > \frac{w_B}{1-q\tilde{m}\rho}$, *then* $p^{CR} > p^*$, *and there exists* $\underline{\lambda}^{CR} > (\frac{1}{1-q\tilde{m}\rho})$ *such that* $\phi^{CR} > 0$ *if* $\lambda^{CR} < \underline{\lambda}^{CR}$.

Banks sell risky assets to increase leverage. If leverage is capped by the capital requirement, banks sell fewer assets. Capital requirements thus act like a negative supply shock that raises prices, and sufficiently tight capital requirements lead to shirking.

7 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 borrowing constraints, but investment efficiency declines when lenders sell too many assets. 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 models 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

ADRIAN, T., AND H. S. SHIN (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), 418–437.

ADRIAN, T., AND H. S. SHIN (2014): "Procyclical Leverage and Value-at-Risk," *Review of Financial Studies*, 27(2), 373–403.

ATTAR, A., T. MARIOTTI, AND F. SALANIÉ (2011): "Nonexclusive Competition in the Market for Lemons," *Econometrica*, 79(6), 1869–1918.

BARON, M., AND W. XIONG (2017): "Credit Expansion and Neglected Crash Risk," *Quarterly Journal of Economics*, 132, 713–764.

BERNANKE, B., AND M. GERTLER (1989): "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79(1), 14–31.

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

——— (2016): "Efficient Bailouts?," *American Economic Review*, 106(12), 3607–3659.

BIANCHI, J., AND E. G. MENDOZA (2013): "Optimal Time-Consistent Macroprudential Policy," *NBER Working Paper 19704*.

BIGIO, S. (2015): "Endogenous Liquidity and the Business Cycle," *American Economic Review*, 105(6), 1883–1927.

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.

BRUNNERMEIER, M. K., AND Y. SANNIKOV (2014): "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104(2), 379–421.

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.

CHERNENKO, S., S. G. HANSON, AND A. SUNDERAM (forthcoming): "Who Neglects Risk? Investor Experience and the Credit Boom," *Journal of Financial Economics*.

CORSETTI, GIANCARLO, P. P., AND N. ROUBINI (1999): "Paper tigers?: A model of the Asian crisis," *European Economic Review*, 43(7), 1211–1236.

DAVILA, E., AND A. KORINEK (2018): "Pecuniary Externalities in Economies with Financial Frictions," *Review of Economic Studies*, 85(1), 352–395.

DOWNING, C., D. JAFFEE, AND N. WALLACE (2009): "Is the Market for Mortgage-Backed Securities a Market for Lemons?," *The Review of Financial Studies*, 22(7), 2457–2494.

DUBEY, P., AND J. GEANAKOPLOS (2002): "Competitive Pooling: Rothschild-Stiglitz Reconsidered," *The Quarterly Journal of Economics*, 117(4), 1529–1570.

EGAN, M., A. HORTACSU, AND G. MATVOS (2017): "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *American Economic Review*, 107(1), 169–216.

EISFELDT, A. L. (2004): "Endogenous Liquidity in Asset Markets," *Journal of Finance*, 59(1), 1–30.

GERTLER, M., AND P. KARADI (2011): "A Model of Unconventional Monetary Policy," *Journal of Monetary Economics*, 58, 17–34.

GORTON, G., AND A. METRICK (2012a): "Securitization," *The Handbook of the Economics of Finance*.

GORTON, G., AND A. METRICK (2012b): "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 Undewriter Reputation: Do Reputable Underwriters," *Review of Financial Studies*, 27(10), 2872–2925.

GRIFFIN, J., AND G. MATURANA (2016): "Who Facilitated Misreporting in Securitized Loans?," *Review of Financial Studies*, 29(7), 1671–1708.

GROMB, D., AND D. VAYANOS (2002): "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics*, 66, 361–407.

HE, Z., AND A. KRISHNAMURTHY (2011): "A Model of Capital and Crises," *Review of Economic Studies*, 79(2), 735–777.

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.

JORDA, O., M. SCHULARICK, AND A. M. TAYLOR (2011): "Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons," *IMF Economic Review*, 59(2), 340–378.

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.

KRUSELL, P., AND A. A. SMITH (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journa of Political Economy*, 106(5), 867–896.

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.

MARTINEZ-MIERA, D., AND R. REPULLO (2017): "Search for Yield," *Econometrica*, 85(2), 351–378.

MENDOZA, E. G. (2010): "Sudden Stops, Financial Crises, and Leverage," *American Economic Review*, 100, 1941–1966.

MENDOZA, E. G., AND V. QUADRINI (2010): "Financial globalization, financial crises and contagion," *Journal of Monetary Economics*, 57, 24–39.

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.

MOREIRA, A., AND A. SAVOV (forthcoming): "The macroeconomics of shadow banking," *The Journal of Finance*.

MYERSON, R. B. (2012): "A Model of Moral-Hazard Credit Cycles," *Journal of Political Economy*, 120(5), 847–878.

ORDOÑEZ, G. (2013): "Fragility of Reputation and Clustering of Risk-Taking," *Theoretical Economics*, 8(3), 653–700.

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.

PURNANANDAM, A. (2011): "Originate-to-distribute Model and the Subprime Mortgage Crisis," *The Review of Financial Studies*, 24(6), 1881–1915.

REINHART, C., AND K. ROGOFF (2009): *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton.

SCHNEIDER, M., AND A. TORNELL (2004): "Balance Sheet Effects, Bailout Guarantees and Financial Crises," *The Review of Economic Studies*, 71(3), 883–913.

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. (2017): "The Downside of Asset Screening for Market Liquidity," *Journal of Finance*, 72(5), 1937–1982.

WHITE, E. N. (2009): "Lessons from the Great American Real Estate Boom and Bust of the 1920s," *NBER Working Paper 15573*.

A Proofs

A.1 Proposition 1

Proof. Follows immediately from the text. □

A.2 Proposition 2

Proof. The budget constraint binds at the optimum. Recall that $P(\chi) \leq \bar{P}(\chi) = \hat{Y} - \frac{m}{1-\chi} < \hat{Y}$. Hence banks sell only if the borrowing constraint binds. Since effort is assumed to be optimal, let $a_B = \underline{a}$. Then $k(\chi) = \frac{w_B}{1-q\tilde{m}\rho-\chi q(p-\rho)}$ and $b_B(\chi) = \Big(\tilde{m}\rho + \chi(p-\rho)\Big)k(\chi)$ by the binding borrowing constraint, and the indirect utility function given $\chi = \frac{a}{k}$ is $u_B(\chi) = \hat{Y}(k-\underline{a}) - b_B + P(\chi)\underline{a} = \left[\frac{\hat{Y}-\tilde{m}\rho-(\hat{Y}-\rho)\chi}{1-q\tilde{m}\rho-q\chi(P(\chi)-\rho)}\right] w_B$. Since $P(\chi)$ is differentiable, so is $u_B(\chi)$. So $\frac{\partial u_B(\chi)}{\partial \chi} \geq 0$ if and only if $\tilde{u} \equiv -(\hat{Y}-\rho)\Big(1 - q\tilde{m}\rho - q\left(P(\chi)-\rho\right)\chi\Big) + \Big(\hat{Y} - \tilde{m}\rho - (\hat{Y}-\rho)\chi\Big)\Big(qP(\chi) + q\chi P'(\chi) - q\rho\Big) \geq 0$. Since $P(\chi)$ is weakly decreasing, $\tilde{u} \leq \hat{u} \equiv -(\hat{Y}-\rho)\left(1-q\tilde{m}\rho\right) + \Big(\hat{Y} - \tilde{m}\rho\Big) q(P(\chi)-\rho)$. Hence $\frac{\partial u_B(\chi)}{\partial \chi} \geq 0$ only if $\hat{u} \geq 0$, and $\hat{u} \geq 0$ if and only if $P(\chi) \geq \frac{\hat{Y}-\rho+\hat{Y}q(1-\tilde{m})\rho}{q(\hat{Y}-\tilde{m}\rho)}$. Rearranging yields $P(\chi) \geq \underline{P}(q)$. $\Pi(q)$ is strictly increasing in q because $q \leq 1$. Hence $\underline{P}(\frac{1}{\hat{Y}}) = \hat{Y}$ because $\Pi(\frac{1}{\hat{Y}}) = 0$. □

A.3 Corollary 1

Proof. Follows from $\frac{\bar{P}(\chi)}{\partial\chi} < 0$ and $\lim_{\chi\to 1} = -\infty$. $\underline{\chi}(q)$ is such that $\bar{P}(\underline{\chi}(q)) = \underline{P}(q)$. $\square$

A.4 Corollary 2

Proof. Follows from the definition of $\underline{\chi}(q)$. $\square$

A.5 Proposition 3

Proof. Follows directly from the proof of Proposition 2 given $P(\chi) = p$. $\square$

A.6 Proposition 4

Proof. Since $\bar{P}(\chi)$ is strictly decreasing, $p < \bar{P}(\chi)$ for all $\chi < \chi^{\max}$. Hence it is ex-post optimal to exert effort for any $\chi < \chi^{\max}$, and b_B is bounded by the good bank's borrowing constraint. Since expected utility is increasing in k and b_B, and $\chi^{\max}$ maximally relaxes the borrowing constraint, bad banks can do no better than $\chi^* = \chi^{\max}$. $\square$

A.7 Lemma 1

Proof. Given some fixed $b_F \geq 0$ satisfying the borrowing constraint, financiers solve the program $\max_{0\leq s_F\leq w_F+qb_F} \quad u_F(s_F|b_F) = \frac{\hat{x}(\chi)}{p}(w_F + qb_F - s_F) - b_F + s_F$ Hence $\frac{\partial u_F}{\partial s_F} < 0$ if $p < \hat{x}(\chi)$, $\frac{\partial u_F}{\partial s_F} > 0$ if $p > \hat{x}(\chi)$ and $\frac{\partial u_F}{\partial s_F} = 0$ if $p = \hat{x}(\chi)$. $\square$

A.8 Lemma 2

Proof. If $\tilde{r}(\hat{x}(\chi), p) > \frac{1}{q}$, then the expected return on investment is higher than the cost of capital. Given risk-neutrality, it is optimal to borrow as much as possible. If $\tilde{r}(\hat{x}(\chi), p) < \frac{1}{q}$, then the cost of capital is higher than the return on investment, and it is optimal to not issue debt. If $\tilde{r}(\hat{x}(\chi), p) = \frac{1}{q}$, then any borrowing decision is optimal. $\square$

A.9 Proposition 5

Proof. The budget constraint and the borrowing constraint imply $a_F \leq \bar{a}_F \equiv \frac{w_F+(q-1)s_F}{p-qx_L(\chi)}$. Since $q \leq 1$, $\frac{\partial \bar{a}_F}{\partial s_F} \leq 0$ and $s_F^* = 0$ if $\hat{x}(\chi) > p$. The remaining statements follow from Lemma 1 and Lemma 2. $\square$

A.10 Proposition 6

Proof. Uniqueness of the symmetric equilibrium is immediate from noting that (i) all banks and financiers are identical, (ii) $P(\chi) = p^*$ for all $\chi \in [0, \chi^{\max}]$ by definition, and (iii) all constraints are convex. Existence follows from $q > \underline{q}$, which implies that $\chi^{\max} > 0$.

Statement (i) characterizes the symmetric equilibrium borrowing decision given individually inderminate optimal leverage ($q\hat{x}(\chi^*) = p^*$). Under symmetry, $a_F^*(\chi^*) = \frac{w_F}{p^* - \gamma^* q x_l(\chi^*)}$. Solving the market-clearing condition $a_F^*(\chi^*) = \mathbf{a}_B(\chi^*)$ for γ^* yields the result. The $\min$ and $\max$ operators ensure feasibility, $\gamma^* \in [0,1]$.

The first step in proving the remainder of the proposition is to construct the cutoffs $\underline{\omega}$ and $\bar{\omega}$. Guess and verify that $p^* \in (\underline{P}(q), \bar{P}(\chi^*))$, $\chi^* = \chi^{\max}$ and $\phi^*(\chi^*) = 1$. Given p, we then have $k^* = \frac{w_B}{1 - q\tilde{m}\rho - \chi^{\max} q(p-\rho)}$ and $a_F(\chi^*) = \frac{w_F}{p - \gamma^* q Y_l}$, where γ^* is the optimal borrowing decision from Lemma 2. Solving the market-clearing condition $a_F(\chi^{\max}) = \chi^{\max} k(\chi^{\max})$ for p yields

$$p = \hat{p}(\omega|\chi^{\max}) \equiv \frac{\omega\left(1 - q\rho(\tilde{m} - \chi^{\max})\right) + \chi^{\max}\gamma^* q Y_l}{\chi^{\max}(1 + q\omega)}$$

Market-clearing can be equivalently stated as $\tilde{a}_F(\chi^{\max}) = \chi^{\max}\tilde{k}(\chi^{\max})$, where $\tilde{a}_F(\chi^{\max}) = \omega\left(p - \gamma^* q Y_l\right)^{-1}$ and $\tilde{k}(\chi^{\max}) = \left(1 - q\tilde{m}\rho - \chi^{\max} q(p-\rho)\right)^{-1}$. For fixed γ^*, it follows that $\frac{\partial \tilde{a}_F(\chi^{\max})}{\partial \omega} > 0$ and $\frac{\partial \tilde{a}_F(\chi^{\max})}{\partial p} < 0$, while $\frac{\partial \tilde{k}(\chi^{\max})}{\partial \omega} = 0$ and $\frac{\partial \tilde{k}(\chi^{\max})}{\partial p} > 0$.

The comparative statics of $\hat{p}(\omega|\chi^{\max})$ with respect to ω are as follows. If $q\hat{Y} > \hat{p}(\omega|\chi^{\max})$ or $q\hat{Y} < \hat{p}(\omega|\chi^{\max})$, then γ^* is fixed at 1 and 0, respectively. Hence $\frac{\partial \hat{p}(\omega|\chi^{\max})}{\partial \omega} > 0$ because $\frac{\partial \tilde{a}_F(\chi^{\max})}{\partial \omega} > 0$. If $q\hat{Y} = \hat{p}(\omega|\chi^{\max})$, then $\hat{p}(\omega|\chi^{\max})$ is a constant because $q\hat{Y}$ is a constant. To verify market-clearing, note that $\gamma^* = \frac{\hat{p}(\omega|\chi^{\max})\chi^{\max}k(\chi^{\max}) - w_F}{qY_l\chi^{\max}k(\chi^{\max})}$ by Statement (i). Hence $a_F^* = \chi^{\max}k(\chi^{\max})$, and γ^* adjusts to account for ω given fixed prices within the interval $[0,1]$. It follows that $\frac{\partial \hat{p}(\omega|\chi^{\max})}{\partial \omega} \geq 0$ for all $\omega > 0$.

To construct the cutoffs $\underline{\omega}$ and $\bar{\omega}$, define the *lowest* and *highest* leverage consistent with individual optimality for any p and q to be $\bar{\gamma}(p) = \mathbb{1}(q\hat{Y} > p)$ and $\underline{\gamma}(p) = 1 - \mathbb{1}(q\hat{Y} < p)$, respectively. Note that $\bar{\gamma}(p) \neq \underline{\gamma}(p)$ if and only if $q\hat{Y} = p$. Then define $\hat{p}^-(\omega|\chi^{\max}) \equiv \frac{\omega(1 - q\rho(\tilde{m} - \chi^{\max})) + \chi^{\max}\underline{\gamma}(\underline{P}(q))qY_l}{\chi^{\max}(1+q\omega)}$ and $\hat{p}^+(\omega|\chi^{\max}) \equiv \frac{\omega(1 - q\rho(\tilde{m} - \chi^{\max})) + \chi^{\max}\bar{\gamma}(\bar{P}(\chi^{\max}))qY_l}{\chi^{\max}(1+q\omega)}$ and note that $\hat{p}^-(\omega|\chi^{\max})$ and $\hat{p}^+(\omega|\chi^{\max})$ are both strictly increasing in ω.

To construct the lower cutoff, recall that banks sell only if $p^* \geq \underline{P}(q)$. Since $\hat{p}(\omega|\chi^{\max})$ is weakly increasing in ω, define $\underline{\omega}$ to be the *smallest* ω such that $\underline{P}(q) = \hat{p}(\omega|\chi^{\max})$. Then $\hat{p}^-(\underline{\omega}|\chi^{\max}) = \underline{P}(q)$ by construction, and so $\underline{\omega}$ is unique because $\hat{p}^-(\omega|\chi^{\max})$ is strictly increasing. It follows that $\hat{p}(\omega|\chi^{\max}) < \underline{P}(q)$ if $\omega < \underline{\omega}$. Since $p^* \geq \underline{P}(q)$ by bank optimality and banks are indifferent at $\underline{P}(q)$ by construction, we have that $p^* = \underline{P}(q)$ and $\chi < \chi^{\max}$ if $\omega < \underline{\omega}$. By defintion of $\chi^{\max}$, moreover, $\phi^* = 1$ if $\omega < \underline{\omega}$. Since $\underline{P}(q) > 0$ we have $\underline{\omega} > 0$. Solving $\hat{p}^-(\underline{\omega}|\chi^{\max}) = \underline{P}(q)$ gives the stated cutoff.

To construct the upper cutoff, recall that banks exert effort only if $p \leq \bar{P}(\chi^{\max})$. Since $\hat{p}(\omega|\chi^{\max})$ is weakly increasing in ω, define $\bar{\omega}$ to be the *largest* ω such that $\bar{P}(\chi^{\max}) = \hat{p}(\bar{\omega}|\chi^{\max})$. Then $\hat{p}^+(\bar{\omega}|\chi^{\max}) = \underline{P}(q)$ by construction, and $\bar{\omega}$ is unique because $\hat{p}^+(\omega|\chi^{\max})$ is strictly increasing. By construction of $\bar{P}(\chi)$, it follows that $\phi^* = 1$ if $\omega \leq \bar{\omega}$. Solving $\hat{p}^+(\bar{\omega}|\chi^{\max}) = \bar{P}(\chi^{\max})$ gives the stated cutoff.

Next determine the relative size of the two cutoffs. If $\underline{\chi}(q) \leq \bar{\chi}$, then $\chi^{\max} = \underline{\chi}(q)$. Hence $\hat{p}(\underline{\omega}|\chi^{\max}) = \underline{P}(q) = \bar{P}(\chi^{\max})$. If $q\hat{Y} \neq \underline{P}(q)$, then $\bar{\gamma}(\underline{P}(q)) = \underline{\gamma}(\underline{P}(q))$ and $\underline{\omega} = \bar{\omega}$. If $q\hat{Y} = \underline{P}(q)$, then $\bar{\gamma}(\underline{P}(q)) < \underline{\gamma}(\underline{P}(q))$ and $\bar{\omega} > \underline{\omega}$. If instead $\underline{\chi}(q) > \bar{\chi}$, then $\bar{P}(\chi^{\max}) > \underline{P}(q)$ and $\bar{\gamma}(\bar{P}(\chi^{\max})) \leq \underline{\gamma}(\underline{P}(q))$. Hence $\bar{\omega} > \underline{\omega}$. This proves statement (v). The proof of statements (ii)-(iv) is as follows.

(ii) If $\omega \in (0, \underline{\omega}]$, we have already shown that $p^* = \underline{P}(q)$ and $\phi^* = 1$. Hence $\chi^* = \frac{a_F^*}{k(\chi)}$, where $k^* = \frac{w_B + q(\underline{P}(q) - \rho)a_F^*}{1 - q\tilde{m}\rho}$ and $a_F^* = \frac{w_F}{\underline{P}(q) - \gamma^* q Y_l}$. So $\frac{\partial k^*}{\partial w_B} > 0$. If $q\hat{Y} \neq \underline{P}(q)$, then $\gamma^* = 1$ or $\gamma^* = 1$. Hence $\frac{\partial k^*}{\partial w_F} > 0$. If $q\hat{Y} = \underline{P}(q)$, then $\gamma \in [0, 1]$ and a_F^* and k^* are locally independent of w_F.

(iii) Now let $\omega \in (\underline{\omega}, \bar{\omega}]$. This interval is measure zero if $\bar{\chi} \leq \underline{\chi}(q)$. Hence let $\bar{\chi} > \underline{\chi}(q)$. By the arguments above, we then have $p^* = \hat{p}(\omega|\bar{\chi})$, and $\frac{\partial \hat{p}(\omega|\bar{\chi})}{\partial \omega} \geq 0$. Since $\chi^* = \bar{\chi}$, it follows that k^* is strictly increasing in w_B and weakly increasing in w_F.

(iv) The asset price is constant at $\bar{P}(\chi^{\max})$, and $\chi = \chi^{\max}$ throughout. Hence k^* is increasing in w_B but constant in w_F. The market-clearing condition is $\phi\chi^{\max}k^* + (1-\phi)k^* = \frac{w_F}{\bar{P}(\chi^{\max}) - q\gamma^*(y_l + f(Y_l - y_l))} = \frac{w_F}{\bar{P}(\chi^{\max}) - q\gamma^* y_l - q\gamma^* f(Y_l - y_l)}$ where $f = \frac{\phi\chi^{\max}}{\phi\chi^{\max} + (1-\phi)}$. Then

$$\phi^* = \frac{(\bar{P}(\chi^{\max}) - q\gamma^* y_l)k^*(\chi) - w_F}{q\gamma^*(Y_l - y_l)\chi^{\max}k^*(\chi^{\max}) + (\bar{P}(\chi^{\max}) - q\gamma^* y_l)(1 - \chi^{\max})k^*(\chi)},$$

where k^* is linear in w_B. There are two cases to consider. If $q\left(f^*\hat{Y} + (1-f^*)\hat{y}\right) \neq \bar{P}(\chi^{\max})$, then $\gamma^* = 1$ or $\gamma^* = 0$. Hence ϕ^* is strictly decreasing in ω and w_F. Since k^* is constant in w_F, $\mathbb{E}\mathbf{Y}(z)$ is strictly decreasing in w_F. If $q\left(f^*\hat{Y} + (1-f^*)\hat{y}\right) = \bar{P}(\chi^{\max})$, then f^* is a constant. Hence ϕ^*, k^* and $\mathbb{E}\mathbf{Y}(z)$ are constant in w_F. The latter case obtains if (i) it is strictly optimal to borrow if no one else does, and (ii) it is strictly optimal to not borrow if everybody else does. These conditions may be satisfied for some q because excess demand (and thus the fraction of bad banks) is increasing in financier leverage. Hence there may exist an interval for w_F on which asset quality is constant and γ adjusts to clear the market. Finally, ϕ^* is bounded below by $\underline{\phi}(\bar{P}(\chi^{\max}), \chi^{\max})$ since financiers cease to buy otherwise.

□

A.11 Corollary 3

Proof. If $\bar{\chi} < \underline{\chi}(q)$, then $\chi^{\max} = \bar{\chi}$. Hence $\bar{P}(\chi^{\max}) = \hat{y}$ and $\bar{\omega} = \frac{\bar{\chi}(\hat{y} - \bar{\gamma}(\hat{y})qY_l)}{1 - q\rho\tilde{m} - q\bar{\chi}(\hat{y} - \rho)}$. We want to show that $\frac{\partial \bar{\omega}}{\partial Y_h} > 0$. A sufficient condition is that $\frac{\partial \bar{\omega}}{\partial Y_h} > 0$ given $\bar{\gamma}(\hat{y}) = 1$. Assume that this is the case. Then $\bar{\omega} = \frac{B}{1 - qA}$ where $B = \bar{\chi}(\hat{y} - qY_l)$ and $A = \rho(\tilde{m} - \bar{\chi}) + \bar{\chi}\hat{y}$. Recall that $\bar{\chi} = \frac{\tilde{m}\rho - y_l}{\rho - y_l}$, which can be rewritten as $\bar{\chi} = \frac{\hat{Y} - \hat{y} - m}{\hat{Y} - \hat{y}}$. Hence $\rho(\tilde{m} - \bar{\chi}) = \frac{m y_l}{\hat{Y} - \hat{y}}$, $A = \frac{\hat{y}(\hat{Y} - \hat{y}) - m(\hat{y} - y_l)}{\hat{Y} - \hat{y}}$, and $B = \frac{(\hat{Y} - \hat{y} - m)(\hat{y} - qY_l)}{\hat{Y} - \hat{y}}$, where $\frac{\partial A}{\partial Y_h} > 0$ and $\frac{\partial B}{\partial Y_h} > 0$. Hence $\frac{\partial \bar{\omega}}{\partial Y_h} > 0$. □

A.12 Proposition 7

Proof. Let financiers choose an individually optimal and feasible portfolio $\{a_F, b_F, s_F\}$. The budget constraint implies $s_F = w_F + qb_F - pa_F$. Proposition 6 implies that $\phi^* < 1$ if $\omega > \bar{\omega}$. By market clearing, $a_F = \phi^*\chi^*k + (1-\phi^*)k^*$ where χ^* and k^* are constant for all $[\bar{\omega}, \infty)$. Hence $\phi^*(a_F) = \frac{k^* - a_F}{(1-\chi^*)k}$ and $f^*(a_F) = \frac{\chi^*(k^* - a_F}{a_F(1-\chi^*)}$. In the symmetric equilibrium where all financiers choose the same portfolio[15], the expected utility given a_F is $u_F(a_F) = \left(f^*(a_F)\hat{Y} + (1 - f^*(a_F))\hat{y}\right) a_F + (q-1)b_F + w_F - \bar{P}(\chi^*)a_F$. Now consider a coordinated investment policy $a_F = \bar{a}$. By definition of $f^*(a_F)$, $u_F(\bar{a}) = \frac{\hat{Y}\chi^*(k^* - \bar{a})}{1-\chi^*} + \frac{\hat{y}(\bar{a} - \chi^* k^*)}{1-\chi^*} + (q - 1)b_F + w_F - \bar{P}(\chi^*)\bar{a}$. Observe that $\bar{P}(\bar{\chi}) = \hat{y}$. Hence $\bar{P}(\chi) \geq \hat{y}$ for all $\chi \in [0, \chi^{\max}]$. Given fixed b_F, $\frac{\partial u_F(\bar{a})}{\partial \bar{a}} = -\frac{\hat{Y}\chi^*}{1-\chi^*} + \frac{\hat{y}}{1-\chi^*} - \bar{P}(\chi^*) \leq -(\hat{Y} - \hat{y})\left(\frac{\chi^*}{1-\chi^*}\right) < 0$. Since $y_l < Y_l < 1$ and $\frac{\partial f^*(\underline{a})}{\partial \underline{a}} > 0$, moreover, the borrowing constraint is weakly relaxed. A marginal reduction in $\underline{a}$ thus strictly increases financier welfare given $a_F^* > \chi^* k^*$. Finally, note that bank welfare is a constant since χ^*, k^*, and $p^* = \bar{P}(\chi^*)$ are constant. $\square$

A.13 Proposition 8

Proof. $\gamma^* = 1$ since $q\hat{Y} > \underline{P}$ and $\omega < \underline{\omega}$. So $k^* = \frac{w_B + q(\underline{P}(q) - \rho)\left(\frac{w_F}{\underline{P}(q) - qY_l}\right)}{1 - q\tilde{m}\rho}$ by Proposition 6. Hence $\frac{\partial k^*(\chi^*)}{\partial w_F} > \frac{\partial k^*(\chi^*)}{\partial w_B} \Leftrightarrow q(\underline{P}(q) - \rho) > \underline{P}(q) - qY_l \Leftrightarrow Y_l - \rho > \frac{(1-q)\underline{P}(q)}{q}$. The right-hand is strictly decreasing in q since $\underline{P}(q)$ is strictly decreasing. By definition, $\rho = \frac{\mathbb{E}_z Y_z - \pi_h y_h}{\pi_l}$. Hence $\rho < Y_l$ if $y_h > Y_h$. $\square$

A.14 Proposition 9

Proof. $\gamma^* = 1$ since $q\hat{Y} > \bar{P}(\chi^{\max})$. Assume first that $\omega > \underline{\omega}$. Then $k^* = \frac{w_B + q(\underline{P}(q) - \rho)a_F^*}{1 - q\tilde{m}\rho}$ by Proposition 6. $\underline{P}(q)$ is equivalently stated as $\underline{P}(q) = \frac{\hat{Y} - \rho + q\hat{Y}\rho(1-\tilde{m})}{q(\hat{Y} - \tilde{m}\rho)}$, and so $k^* = \frac{w_B}{1 - q\tilde{m}\rho} + \left(\frac{\hat{Y} - \rho}{\hat{Y} - \tilde{m}\rho}\right) a_F^*$. Recall that $w_B'(z) = Y_z(k - \underline{a}) + p\underline{a} - b_B$. Then the binding borrowing constraint gives $w_B'(z) = \left(\frac{Y_z - \tilde{m}\rho}{1 - q\tilde{m}\rho}\right) w_B - (Y_z - \hat{Y})\rho(1 - \tilde{m})a_F^*$. By definition, $w_F'(z) = Y_z a_F - b_F$. By Proposition 5, $w_F'(z) = (Y_z - Y_l)a_F^*$ and $a_F^* = \frac{w_F}{\underline{P}(q) - qY_l}$. To derive the cutoff bond price, observe that $Y_l \geq \rho > \tilde{m}\rho$ since $y_h \geq Y_h$, and note that $E_z v_F(z) \geq E_z v_B(z)$ if and only if $\underline{P}(q) \leq A \equiv \frac{\hat{Y} - Y_l + q\hat{Y}(Y_l - \tilde{m}\rho)}{\hat{Y} - \tilde{m}\rho}$. This expression can be rearranged to show that $A = \frac{\hat{Y} - \rho + q\hat{Y}(1-\tilde{m})\rho + (q\hat{Y} - 1)(Y_l - \rho)}{\hat{Y} - \tilde{m}\rho} \geq \frac{\hat{Y} - \rho + q\hat{Y}(1-\tilde{m})\rho}{\hat{Y} - \tilde{m}\rho} = \underline{P}(1)$. Hence $E_z v_F(z) \geq E_z v_B(z)$ if $q = 1$. Since $Y_l > \tilde{m}\rho$, it follows that $\frac{1}{\underline{P}(q) - qY_l} > \frac{1}{1 - q\tilde{m}\rho}$ if $q = 1$. Hence $v_F(h) > v_B(h)$ if $q = 1$. The existence of $\hat{q} < 1$ follows from continuity of $v_F(h) - v_B(h)$. This proves the first statement.

Now consider the second statement. Let $\omega \in (\underline{\omega}, \bar{\omega}]$, and note that $\bar{P}(\bar{\chi}) = \hat{y}$. Then $\underline{a}^* = \bar{\chi}k^*$, and so $w_B'(z) = (Y_z - \tilde{m}\rho - \bar{\chi}(Y_z - \rho))\, k^*$. Market-clearing requires $\bar{\chi}k^* = a_F^*$. Hence $w_F'(z) = (Y_z - Y_l)a_F^* = \bar{\chi}(Y_z - Y_l)k^*$, and $\omega'(z) = \frac{\bar{\chi}(Y_z - Y_l)}{Y_z - \tilde{m}\rho - \bar{\chi}(Y_z - \rho)}$. The inequality follows immediately from the definition of $\bar{\omega}$. This proves the second statement.

Now consider the third statement. Let $\omega > \bar{\omega}$ and assume that financiers continue to borrow ($\gamma^* = 1$). Then $\mathbf{a}_B = \phi^*\bar{\chi}k^* + (1 - \phi^*)k^*$. For good banks, $w_B'(z) = (Y_z - \tilde{m}\rho - \bar{\chi}(Y_z - \rho))\, k^*$. For bad banks, $w_B'(z) = \bar{P}(\bar{\chi})k^* - b_B = \left(\bar{P}(\bar{\chi}) - \tilde{m}\rho - \bar{\chi}(\bar{P}(\bar{\chi}) - \rho)\right) k^*$. Hence $\mathbf{w}_B'(z) = \left(\phi^* Y_z + (1 - \phi^*)\bar{P}(\bar{\chi}) - \tilde{m}\rho - \bar{\chi}(\phi^* Y_z + (1 - \phi^*)\bar{P}(\bar{\chi}) - \rho)\right) k^*$. Now turn to financiers and

[15] The equilibrium allocation is necessarily symmetric unless financiers portfolios are indeterminate due to indifference. In any asymmetric allocation due to indifference, the market-clearing condition is unchanged. Hence the fraction of shirking banks remains the same, as does k^* and $\chi^{\max}$. In this case, $u_F(a_F)$ represents average utility and the subsequent welfare analysis remains valid.

let $f^* = \frac{\phi^*\bar{\chi}}{\phi^*\bar{\chi}+(1-\phi^*)}$. Then $w'_F(z) = (f^*(Y_z - Y_l) + (1-f^*)(y_z - y_l))a^*_F$. By market-clearing, $a^*_F = \phi^*\bar{\chi}k^* + (1-\phi^*)k^*$. Hence $w'_F(z) = (\phi^*\bar{\chi}(Y_z - Y_l) + (1-\phi^*)(y_z - y_l))k^*$, and $\omega'(z) = \frac{\phi^*\bar{\chi}(Y_z - Y_l)+(1-\phi^*)(y_z-y_l)}{\phi^* Y_z + (1-\phi^*)\bar{P}(\bar{\chi}) - \tilde{m}\rho - \bar{\chi}(\phi^* Y_z + (1-\phi^*)\bar{P}(\bar{\chi}) - \rho)}$. □

A.15 Corollary 4

Proof. $\phi^* > \underline{\phi}(\bar{P}(\chi^{\max}), \chi^{\max})$ implies that financiers do not invest in the safe technology. Adapating Proposition 9 to the case $\gamma = 0$ and $\omega > \bar{\omega}$ yields $w'_F(z) = (\phi^*\bar{\chi}Y_z + (1-\phi^*)y_z)k^*$, where k^* is a constant. The result follows from $y_l < Y_l$ and $\frac{\partial\phi^*}{\partial w_F} < 0$. □

A.16 Proposition 10

Proof. Banks are risk-neutral conditional on the state-contingent marginal values defined in Definition 6. Hence the derivations from the static model apply.

□

A.17 Proposition 11

Proof. Financiers are risk-neutral conditional on the state-contingent marginal values defined in Definition 7. Hence the derivations from the static model apply. □

A.18 Proposition 12

Proof. Follows immediately from comparing the rate of return on bonds $\frac{1}{q}$ and the sure return of 1 delivered by the safe technology. □

A.19 Proposition 13

Proof. Follows immediately from observing that banks strictly prefer to shirk if only if $a_B > \chi^{\max}k$. The skin-in-the-game rule directly prevents banks from selling more than $\chi^{\max}k$ assets. Restricting financier demand indirectly ensures that banks cannot sell more than $\chi^{\max}k$ assets by affecting the market-clearing condition. □

A.20 Corollary 5

Proof. Follows immediately from Proposition 9 and noting that $\frac{\partial\omega'(h)}{\partial\omega} > 0$ if $\omega \leq \underline{\omega}$. □

A.21 Proposition 14

Proof. Note that $k = k^0 = \frac{w_B}{1-q\tilde{m}\rho}$ if $\underline{a} = 0$. Hence $\underline{a} > 0$ if $k = \lambda^{CR}w_B$. Specifically, $k = \lambda^{CR}w_B$ iff $\chi^{CR}(\lambda^{CR}) = \left(\lambda^{CR}(1 - q\tilde{m}\rho) - 1\right) / \left(\lambda^{CR}(p^* - \rho)q\right)$ and $\lim_{\lambda^{CR}\to(1-q\tilde{m}\rho)^{-1}} \chi^{CR}(\lambda^{CR}) = 0$. Assume $\phi^* = 0$. Market-clearing is $\frac{w_F}{p-qY_l} = \chi^{CR}\lambda^{CR}w_B$. So $p^{CR} = \frac{w_F}{\chi^{CR}\lambda^{CR}w_B} + qY_l$ and $\lim_{\lambda^{CR}\to(1-q\tilde{m}\rho)^{-1}} p^{CR} = \infty > \bar{P}(0)$. Hence $\phi^{CR} > 0$ for λ^{CR} close to $(1-q\tilde{m}\rho)^{-1}$. □

Online Appendix. Not for Publication.

B Extension to Signaling Equilibrium

The refined competitive equilibrium concept employed in the main text implies that the price schedule is constant across all submarkets consistent with effort. Individual banks thus cannot affect their terms of trade by moving to another submarket. This is consistent with the general equilibrium setting emphasized here, and particularly so if no bank shirks. If the upper bound on the asset price binds, however, then it may be sensible to allow banks to signal their intent to produce high quality assets by pledging fewer assets as collateral. To see why this is the case, recall that $\bar{P}(\chi) = \hat{Y} - \frac{m}{1-\chi}$ is strictly decreasing in χ. If there is excess demand at $\bar{P}(\chi^*)$ (as is the case when some banks shirk), it may be reasonable to assume that an individual bank who promises to retain $1-\chi' > 1-\chi^*$ assets can transact at price $\bar{P}(\chi') > \bar{P}(\chi^*)$ because financiers believe they are marginally less likely to shirk at χ^*. (Whether or not financiers are willing to buy naturally depends on the (expected) fraction of shirking banks at χ^* and χ'. To make the argument simple, I simply assume that banks can always transact at $\bar{P}(\chi)$ for all $\chi \in [0, \chi^{\max}]$ whenever $p^* = \bar{P}(\chi^*)$). I refer to this equilibrium concept as the *signaling equilibrium*. Since it allows banks to sell at the maximum feasible price after any deviation, it provides the maximum incentives to deviate to another submarket. In this sense, it is the polar opposite of the refined equilibrium, which provided no incentives to deviate. However, I now show that the basic mechanisms highlighted so far are robust to this alternative method of modeling off-equilibrium prices.

B.1 Optimal Bank Portfolio in the Signaling Region

The first step is to characterize the optimal bank portfolio under the assumption that $P(\chi) = \bar{P}(\chi)$. As before, we can take as given that the borrowing and budget constraints bind. Investment and bond issuances conditional on χ are thus given by

$$k(\chi) = \frac{w_B}{1 - q\tilde{m}\rho - \chi q(\bar{P}(\chi) - \rho)} \qquad \text{and} \qquad b_B(\chi) = \Big(\tilde{m}\rho - \chi(\bar{P}(\chi) - \rho)\Big)k(\chi)$$

By construction, banks who shirk obtain the same utility as banks who exert effort. Conditional on χ, we can therefore write bank utility as

$$u_B(\chi) = \left(\bar{P}(\chi) + m\right)k(\chi) - b_B(\chi) = \left(\frac{\hat{Y} - \tilde{m}\rho - (\hat{Y} - \rho)\chi}{1 - q\tilde{m}\rho - q\chi(\hat{Y} - \rho) + \frac{qm\chi}{1-\chi}}\right) w_B$$

Individual optimality requires $\chi^* = \arg\max_{\chi \in [0, \chi^{\max}]} u_B(\chi)$. (It is easy to verify that there is a unique optimum). The key difference to the baseline model is that financiers now internalize the effect of χ on prices. The direct effect of a reduction in χ is to reduce leverage. On the other hand, lower χ boosts prices on all inframarginal assets posted as collateral. This is due to a signaling effect: by pledging fewer assets as collateral, the bank credibly signals that it will continue to exert effort at higher prices. The optimal χ trades off these two effects. An important difference to standard models of signaling is that the upper bound on the asset price is continuous in χ. This means that small deviations from χ^* cannot lead to discrete jumps in the terms of trade. I restrict attention to the natural case where signaling is costly in the sense that reductions in χ force the bank to delever.

Assumption 5 (Costly Signaling). $\frac{\partial k(\chi)}{\partial \chi} \geq 0 \; \forall \, \chi \in [0, \bar{\chi}]$.

Observation 4. *Assumption 5 holds if and only if* $\bar{\chi} \leq \frac{\hat{Y}-\rho-m}{\hat{Y}-\hat{y}+\hat{Y}-\rho}$.

Proof. $\frac{\partial k(\chi)}{\partial \chi} = k(\chi)\left(\frac{q}{1-q\tilde{m}\rho-\chi q(\bar{P}(\chi)-\rho)}\right)\left(\bar{P}(\chi)-\rho+\chi\bar{P}'(\chi)\right)$ where $\bar{P}'(\chi) = -\frac{m}{(1-\chi)^2}$. $\bar{P}(\chi)$ and $\bar{P}'(\chi)$ are strictly decreasing in χ. Hence Assumption 5 holds iff $\frac{\partial k(\chi)}{\partial \chi}|\chi = \bar{\chi} \geq 0$. $\square$

Assumption 5 does *not* imply that it is not optimal for the bank to signal. This is because it may be privately beneficial borrow slightly less if doing so means the bank can sell all inframarginal assets at a higher price. In the aggregate, however, the fraction of shirking banks is determined by excess demand at the threshold price. The next result shows that signaling unambiguously lowers the total market value of good risky assets that are for sale if Assumption 5 holds.

Proposition 15. *Define* $MV(\chi) = \bar{P}(\chi) \cdot \chi \cdot k(\chi)$ *to be the market value of risky assets if no farmer shirks. If Assumption* (5) *holds, then* $\frac{\partial MV(\chi)}{\partial \chi} > 0$ *for all* $\chi \in [0, \bar{\chi}]$.

Proof. $\frac{\partial MV(\chi)}{\partial \chi} = \bar{P}(\chi)k(\chi) + \chi\bar{P}'(\chi)k(\chi) + \bar{P}(\chi)\chi k'(\chi)$. By Observation 4, $\bar{P}(\chi) - \rho + \chi\bar{P}'(\chi) \geq 0$. So $\left(\bar{P}(\chi)+\chi\bar{P}'(\chi)\right)k(\chi) > 0$. Since $k'(\chi) \geq 0$ it follows that $\frac{\partial MV(\chi)}{\partial \chi} > 0$. $\square$

Hence signaling contributes to growing excess demand. The next figure provides an example in which signaling is privately optimal but leads to more shirking overall due to a reduction in asset supply. It follows that the possibility of signaling can strengthen the key mechanisms discussed in the baseline model.

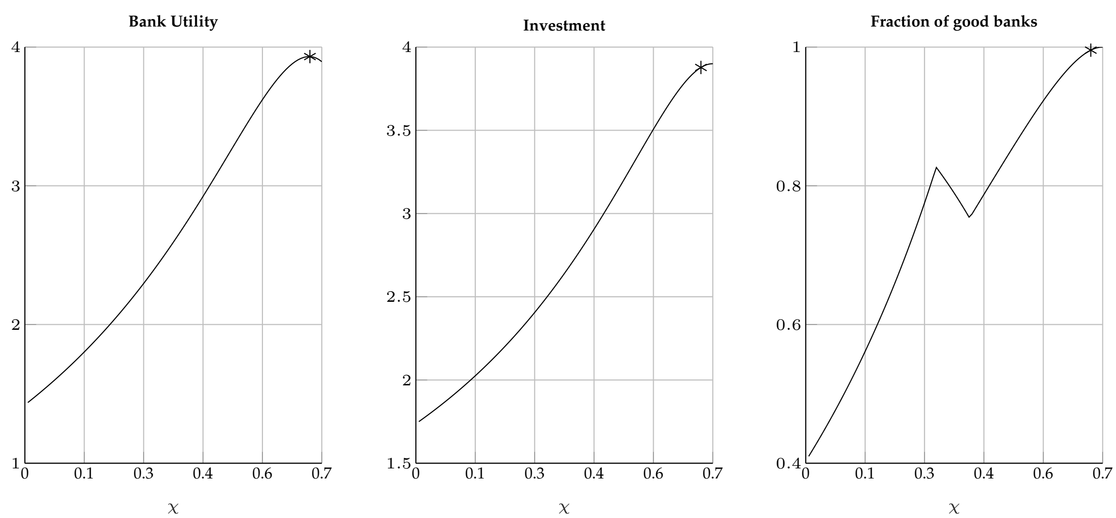

Figure 9: Optimal portfolio and equilibrium outcomes in the signaling equilibrium as a function of χ. *The asterisk denotes the individually optimal choice* χ^*. *All parameters as in Figure 5.The non-monotonicity between 0.3 and 0.4 occurs because financiers are indifferent toward leverage in this region. Since* $P(\chi)$ *is decreasing in* χ, *reductions in* χ *lead to price increases. To keep financiers indifferent,* ϕ *must increase. To see why financiers must be indifferent, note that* ϕ *would be such that individual financiers would borrow if no one else did, and such that no financier would borrow if everybody else did.*

The signaling equilibrium also retains the key property of baseline model: the fraction of shirking banks is increasing in the wealth of financiers.

Proposition 16. *ϕ^* is increasing in w_F in any signaling equilibrium.*

Proof. Given $P(\chi) = \bar{P}(\chi)$ for any $\chi \in [0, \bar{\chi}]$, the invididually optimal χ^* is independent of w_F, and $P(\chi^*) = \bar{P}(\chi^*)$ is a constant. Since q is a constant, so is investment $k(\chi^*)$ and good banks' total asset supply $\chi^* k(\chi^*)$. Since financier demand is increasing in w_F, so is excess demand for good bank assets. Market-clearing thus requires ϕ^* to be increasing. □

Even in the signaling equilibrium, the upper bound on the asset price only binds if financiers are sufficiently wealthy. The equilibrium wealth dynamics in the absence of shirking thus are the same as in the baseline model. Since signaling may lead to an increase in the fraction of shirking banks, moreover, financiers may hold an even larger fraction of aggregate risk exposure in the shirking region. The signaling equilibrium thus replicates the key economic mechanisms from the baseline model. I chose the baseline equilibrium concept because it allows for a more transparent closed-form characterization of equilibrium outcomes.

C Extension to staggered borrowing

This section provides a micro-foundation for the assumption that banks can commit to pledging $\underline{a}$ assets as collateral when issuing debt to savers. The idea is to introduce multiple rounds of borrowing and investment within a given period. Proceeds from asset sales can then be used to pay off debt issued in a previous round, freeing up borrowing capacity for a new round of debt issuance and investment. As in the static model, asset sales thus serve to boost borrowing capacity, but there is no need for commitment.

Let there be N financing rounds in a single period. Within each round, banks can issue debt at price q, invest, and sell assets at a price p_n that is a function of the bank's current observable balance sheet (or equivalently, the submarket it trades in in every period). Let the bank's bond purchases, asset sales and capital investment in round $n = 0 \dots N$ be denoted by $\tilde{b}_n$, $\tilde{a}_n$ and $\tilde{k}_n$ respectively. Banks cannot commit to future sales, can only sell only assets they have already originated, and can use proceeds from current sales to pay off outstanding debt. Asset sales must therefore satisfy the constraints $\tilde{a}_n \leq \tilde{k}_{n-1}$ for $n \geq 1$ and $\tilde{a}_0 = 0$. The round-0 IC constraint then is

$$\sum_z \pi_z \left[Y_z \tilde{k}_0 - \tilde{b}_0 \right] \geq \sum_z \pi_z \left[\max \left\{ y_z \tilde{k}_0 - \tilde{b}_0, 0 \right\} \right] + m \tilde{k}_0.$$

Suppose that, in round 1, the bank sells $\tilde{a}_1$ assets in exchange for $p\tilde{a}_1$ in revenue, and uses this revenue to pay off existing debt. After issuing $\tilde{b}_1$ in new debt, total outstanding debt is $\tilde{b}_0 + \tilde{b}_1 - p\tilde{a}_1$. Hence the round-1 IC constraint is

$$\begin{aligned} &\sum_z \pi_z \left[Y_z(\tilde{k}_0 + \tilde{k}_1 - a_1) - \tilde{b}_0 - \tilde{b}_1 + p_1 \tilde{a}_1 \right] \\ \geq &\sum_z \pi_z \left[\max \left\{ y_z(\tilde{k}_0 + \tilde{k}_1 - a_1) - \tilde{b}_0 - \tilde{b}_1 + p_1 \tilde{a}_1, 0 \right\} \right] + m(\tilde{k}_0 + \tilde{k}_1). \end{aligned}$$

Iterating forward shows that the round-N IC constraint is

$$\sum_z \pi_z \left[Y_z \left(\sum_{n=0}^{N} \tilde{k}_n - \sum_{n=1}^{N} \tilde{a}_n \right) - \sum_{n=0}^{N} \tilde{b}_n + \sum_{n=1}^{N} p_n \tilde{a}_n \right] \geq$$
$$\sum_z \pi_z \left[\max \left\{ y_z \left(\sum_{n=0}^{N} \tilde{k}_n - \sum_{n=1}^{N} \tilde{a}_n \right) - \sum_{n=0}^{N} \tilde{b}_n + \sum_{n=1}^{N} p_n \tilde{a}_n, 0 \right\} \right] + m \sum_{n=0}^{N} \tilde{k}_n.$$

We can then define $k = \sum_{n=0}^{N} \tilde{k}_n$, $b = \sum_{n=0}^{N} \tilde{b}_n$, $\underline{a} = \sum_{n=1}^{N} \tilde{a}_n$, and $\tilde{P} = \frac{\sum_{n=1}^{N} p_n \tilde{a}_n}{\underline{a}}$ to arrive at the incentive constraint form the baseline model. Note that it is easy for $\tilde{P}$ to incorporate at least as much information as in the static model. To allow for hidden violations of the retention constraint, simply assume that the bank can sell off additional assets in each round, and let these trades be unobservable unless banks use their proceeds to pay off existing debt. This illustrates that the baseline model requires the partial commitment assumption only to allow for simultaneous asset sales and borrowing.

D Robustness to Alternative Moral Hazard Specification

Consider an alternative specification of the effort decision in which banks can choose to exert effort at the level of individual assets rather than at the level of the pool. Fixing k and $\underline{a}$, let $L \leq k$ denote the number of low-quality assets, and let mL denote the associated private benefit. As is true in practice, 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 the pool consists of a continuum of assets, financiers 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

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

This problem is linear in L (up to a binding limited liablity constraint). Hence banks will either choose to shirk on all assets or not at all. Moreover, whether or not the limited-liability constraint binds is determined by the same constraints as in the baseline model. Conditional on letting financiers use a random selection rule, there is thus no loss of generality in assuming that the bank either shirks on all assets or on none.