Inefficient Asset Price Booms*

Daniel Neuhann†

December 2019

Abstract

Endogenous movements in the wealth distribution can generate asset price booms in which financial intermediaries increasingly engage in moral hazard and originate low-quality assets that are excessively exposed to aggregate risk. Central to the mechanism is a pecuniary externality whereby buyers of financial assets do not internalize that high asset prices harm origination incentives. Inefficient asset price booms occur only if risk-averse savers are sufficiently wealthy initially, and incentives decline only after a sufficiently long macroeconomic expansion. During the run-up, increasing asset prices instead allow for an efficient expansion of aggregate credit. I discuss implications for monetary policy and financial regulation.

JEL classification: G01, E32, E44

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

*This paper is based on Chapter 1 of my dissertation at the University of Pennsylvania. It was previously titled "Macroeconomic Effects of Secondary Market Trading." I am grateful for many helpful conversations with Harold Cole, Itay Goldstein, Dirk Krueger, and Guillermo Ordoñez. I thank Luigi Bocola, Selman Erol, John Geanakoplos, Urban Jermann, Michael J. Lee, Enrique Mendoza, Guillaume Plantin, 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 is a theoretical study of credit cycles. It is motivated by three empirical observations. First, credit growth is the best predictor of the likelihood of financial crises, and financial crises are typically preceded by asset price and credit booms (Mendoza and Terrones (2012) and Jorda, Schularick, and Taylor (2011)). Second, recent evidence suggests that capital is increasingly allocated to inefficient investment opportunities during credit booms.[1] Third, many crises are preceded by shifts in the relative importance of bank and non-bank financial institutions in lending.[2] One reading of these observations is that financial crises are partly rooted in a deterioration of private investment incentives over the credit cycle, and that they are linked to changes in the financial intermediation process.

I propose a dynamic framework in which increasing credit volumes and falling credit quality are both driven by endogenous movements in the wealth distribution across financial institutions. 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 boom and bust cycles arise only in specific macroeconomic conditions. I show that these include expansionary monetary policy and saving gluts, and that *temporary* shocks may be a sufficient to trigger eventual declines in asset quality. This allows for a wider accounting of the causes of financial instability than in previous work.

The model features financial intermediaries ("banks") who issue safe debt to risk-

[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 origination over the same period.

[2] Prior to the 2008 financial crisis in the U.S., there was increase in the size of shadow banks and other non-bank financial institutions (Brunnermeier (2009)). Adrian and Shin (2010) estimate that the combined balance sheet size of U.S. non-bank intermediaries such as hedge funds and broker-dealers was smaller than that of bank holding companies before 1990, but twice as large by 2007. Prior to the 1997 Asian financial crisis, there was increase in foreign investors shifting capital to Thailand (Kaminsky (2008)). White (2009) documents an increase in securitization prior to the Great Depression.

averse savers in order to originate risky assets (such as loans to households or firms). Origination is subject to moral hazard: savers are concerned that banks originate low-quality assets with excessive downside risk. This leads to an endogenous borrowing constraint that limits the amount of funds banks can borrow as a function of internal wealth and risk exposure. If wealth is scarce, banks can borrow more only by reducing risk exposure. This is accomplished by selling risky assets to financial market investors ("financiers") in asset markets. In the model, the key difference between banks and financiers is that financiers cannot originate risky assets. This segmentation of asset origination and warehousing has social value: transferring risk to financiers offers commitment against moral hazard in origination, and allows banks to issue more debt. As a result, a financial system with banks and financiers can extend more credit than one with only banks, holding fixed aggregate net worth.

Yet there is also a downside to asset sales. In practice, many financial markets are opaque, and bank balance sheets are hard to observe and contract upon. This leads me to assume that financiers neither observe the quality of individual assets traded, nor the total number of assets sold by each individual bank. That is, asset markets are non-exclusive and there is lack of commitment to asset retention. Kurlat (2013), Bigio (2015) and Eisfeldt (2004) explore similar market structures but assume that asset quality is exogenous. Lack of commitment allows for the possibility that some banks shirk and sell a large number of low-quality assets 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. Because the return to selling assets is determined by asset prices, this is the case if and only if asset prices are sufficiently high. This creates a link between macroeconomic conditions and origination incentives.

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 also feeds back to private incentives. In the model, this leads to a simple feedback rule: banks shirk if the asset price is above an endogenous threshold, and exert effort otherwise. This leads to an endogenous *maximal price* that is consistent with incentives. Since shirking has a private benefit, the maximal asset price is strictly lower than the expected return of good assets. This has the important implication that financiers are willing to buy at the maximal price as long as there are *enough* good assets. Hence asset markets remain liquid even if all participants know that some banks shirk. (Banks are willing to sell below par because of the shadow cost of risk exposure.)

So long as the asset price is below its upper bound, it adjusts freely to clear the asset market, and it is increasing in the ratio of financier wealth to bank wealth ("relative wealth"). But once the price reaches its upper bound, it cannot rise further. This is because all banks would then shirk, in which case financiers would no longer be willing to pay the requisite price. The wealth distribution can therefore be partitioned into two regions defined by a cutoff level of relative wealth. Below the cutoff, the asset price is low but increasing in relative wealth, all banks exert effort, and asset sales efficiently boost credit volumes. Above the threshold, the asset price is at its upper bound, credit volumes no longer grow, and shirking leads to an inefficient decline in credit quality (but the share of shirking banks is always such that financiers continue to earn non-negative rents.)

Why must there be shirking in equilibrium when financiers are wealthy? The reason is that price adjustments can no longer clear the asset market once relative wealth is above the cutoff. Excess demand at the maximal price must therefore be matched by an increase in supply at fixed prices. This can be achieved only by an increase in the fraction of shirking banks, who shirk precisely because they sell more assets than those who exert effort. Above the cutoff, the share of bad assets is thus strictly increasing in relative wealth. The mechanism can be understood through a classical trade-off between risk and incentives: some risk transfer allows banks to efficiently scale up credit and investment, while excessive risk transfer harms incentives. The key difference to classical agency theory is that the equilibrium allocation of risk is implemented by a market arrangement that is sensitive to the wealth distribution.

Limited commitment and asymmetric information are central to these results. Yet the model reveals an important state contingency: asset market frictions are entirely irrelevant if the asset price is low. Hence an upshot of the model is that the extent to which market incompleteness affects aggregate outcomes varies systematically with the wealth distribution (which in turn varies systematically over the credit cycle.) A practical implication is that asset markets which efficiently redistribute risk in typical conditions can at times become a nexus of financial instability. Importantly, asset prices do not reveal this increase in underlying risk: prices are high during periods of increasingly inefficient aggregate risk exposure.

The fact that banks shirk on the path of play is not a surprise to financiers; it is consistent with individual optimality. This is because financiers earn rents on all infra-marginal asset purchases even at the maximal price. Yet it is socially inefficient because banks shirk on their marginal investments. The key externality is that individual financiers do not internalize their effect on aggregate demand. To make this point, I show that a social planner can generate a strict Pareto improvement by restricting financiers' asset purchases. Contrary to canonical models of credit cycles, moreover, the equilibrium features a form of inefficient *underinvestment*: when asset prices are bounded, aggregate investment may not increase even when risk-bearing wealth does.

The model's dynamic predictions follow because the wealth distribution evolves endogenously. Specifically, relative wealth is typically pro-cyclical because financiers 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) may lead to an asset price and 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. Model dynamics are further shaped the fact that agents take the evolution of aggregate wealth as given. Hence they do not internalize that the trading of aggregate risk exposure may lead to excessive growth in relative wealth over time, and the economy transitions to regions of the state space where banks shirk even if all agents are fully rational and forward-looking. This effect is conceptually similar but substantively distinct from Lorenzoni (2008), since it operates through an over-accumulation of capital that excessively raises rather than lowers prices.

The nature and likelihood of credit booms depends on initial conditions. If financiers are very poor to begin with, they 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 growth is muted, and credit quality does not decline. At the same time, I show that relative financier wealth grows after good shocks for *any* initial wealth distribution as long as the risk-free rate is sufficiently low. This is because 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. Additionally, I find that *temporary* reductions in the risk-free rate can have persistent effects because wealth is a substitute for leverage: financiers do not need to rely on cheap borrowing once they have grown sufficiently wealthy. Even short-lived shocks to the risk-free rate may therefore set in motion inefficient asset price booms, pointing to a dynamic risk-taking channel of monetary policy. Another interpretation is that shocks to the wealth of *risk-averse* investors can lead to more risk-taking ex-post. Contrary to models with fire sale externalities, moreover, I find that capital requirements on banks may trigger or increase the extent of inefficient shirking. This is because capital requirements induce banks to sell fewer assets, raising asset prices.

The paper is organized as follows. Section 1.1 contains the literature review. Section 2 discusses the basic mechanisms in a static model with a fixed wealth distribution risk-free rate. Section 3 embeds the static model in an infinite-horizon framework to study the evolution of the wealth distribution and the risk-free rate. Section 4 discusses the empirical literature related to the model's predictions. Section 5 provides a detailed discussion of key model assumptions. The conclusion is in Section 6, and all proofs are in Appendix A. Extensions are in Appendices C and D.

1.1 Related Literature

The wealth distribution is a central object in much of macroeconomics. In models of financial intermediation (e.g. Bernanke and Gertler (1989), Gertler and Kiyotaki (2010), Gertler and Karadi (2011), Brunnermeier and Sannikov (2014), and He and Krishnamurthy (2011)) the net worth of financial intermediaries (the marginal buyers of risky assets) is the key state variable that determines lending volumes and prices, and net worth *shortages* can depress credit and/or raise risk premia. This means that market segmentation (the reason for financial intermediation) does not hamper investment when intermediaries are sufficiently well-capitalized. This paper considers two types of asset market participants, and shows that the marginal buyers of risky assets may be *too wealthy* because of a feedback from asset prices to incentives. This allows the model to speak to observed inefficiencies during good times.

Asset prices and leverage also depend on the wealth distribution in Fostel and Geanakoplos (2008). In their work, it is either because investors differ in beliefs, or because there is asymmetric information about issuer types. Here, the key dimension of heterogeneity is that agents differ in the degree to which they are susceptible to moral hazard. While Fostel and Geanakoplos (2008) draw implications for leverage and price contagion across asset classes, I study the role of leverage and asset prices in determining investment, asset quality, and the evolution of wealth itself. In their model, the share of bad assets grows and adverse selection is more severe when bad news lower asset prices, while here it grows in good times when asset prices are high.

The fact that this model's equilibrium is inefficient is partly due to a pecuniary externality in asset markets. Such externalities are also the focus of a large literature that studies credit booms in models where agent heterogeneity leads to price-sensitive borrowing constraints (e.g. Kiyotaki and Moore (1997), Lorenzoni (2008), Mendoza (2010), Bianchi (2011), 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 over-investment ex-ante. Inefficiencies arise because the natural holders of risky assets have too little wealth during bad times, leading to punitive cash-in-the-market pricing and low investment. While some similar effects are present in this model, the fundamental mechanism works in the opposite direction: investment efficiency declines when financiers have too much wealth, and cash-in-the-market pricing is a necessary condition for efficient investment rather than an ex-post manifestation of excess borrowing ex-ante. Moreover, I find the ratio of risky investment to risk-bearing wealth to be *low* when financier wealth is high.

Gorton and Ordoñez (2014) and Moreira and Savov (2017) study credit booms focusing on the information sensitivity of financial assets. Eisfeldt (2004), Kurlat (2013) and Bigio (2015) show that adverse selection hampers asset market liquidity. In their work, fundamental asset quality is exogenous, asset market volumes are pro-cyclical but inefficiently low at all times, and adverse selection serves primarily to amplify downturns. The opposite is true here: asset quality is endogenous, asset market volumes and prices are too *high* during booms, and excessive asset sales lead to inefficiencies during good times. My model also generates different policy implications: while redistributing wealth to asset buyers reduces market frictions in Kurlat (2013), the same policy may lead to inefficient originator moral hazard here.

Parlour and Plantin (2008) and Vanasco (2017) study the feedback from liquidity to asset quality in static models with fixed secondary market demand. In contrast, I use a dynamic macroeconomic model with an endogenous wealth distribution that determines secondary market demand, and show that a model with these features can generate credit cycles with falling asset quality. Contemporaneous work by Fukui (2018) and Caramp (2018) also considers the interaction of moral hazard and frictional secondary markets. Fukui (2018) shows that asset quality in the economy falls in response to a shock to aggregate asset productivity, but quality in the market rises. My model creates a decline in asset quality in both the economy and the market, in line with the data. In his model, shocks are to the current quality of assets. In my model, the mechanism operates only through the wealth distribution; there are no shocks to the expected productivity of assets, and realized productivity matters only insofar as it has distributional consequences. The presence of financial frictions in my model also generates predictions for the role of monetary policy in affecting the likelihood of crises. Caramp (2018) focuses on the liquidity value of financial assets, and studies interactions of asset quality with the supply of government bonds. My focus is on endogenous movements in the wealth distribution, and its sensitivity to interest rates independent of liquidity.

Previous work has studied the relationship between private moral hazard and booms and busts. Myerson (2012) considers a model in which credit cycles stem from the optimal life-cycle incentive contract for bankers, and actions are always efficient since the optimal contract deters moral hazard (i.e. only investment volume is distorted). I study a model with incomplete markets in which actions are inefficient during booms, and study the endogenous dynamics of wealth and asset prices.[3] 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 rising asset prices and risk-free rates, while low risk-free rates affect the likelihood of such a boom.

2 Static Model

2.1 Environment

Consider one-period economy populated by a unit mass each of risk-neutral banks and financiers, as we all as a unit mass of savers. Banks use own net worth w_B and debt issued to savers to originate pools of risky assets. The payoff generated by a pool depends on unobservable bank effort e and an aggregate state $z \in \{l, h\}$ that is the only source of risk. The probability of state z is $\pi_z \in (0, 1)$. The effort choice is binary: the bank either exerts effort or shirks, $e \in \{0, 1\}$.[4] The output generated by investing k units of capital is $Y_z k$ if banks exert effort and $y_z k$ if banks shirk. Hence output can also be written as

[3] Another literature focuses on how the severity of moral hazard varies with exogenous fundamentals. Examples include Povel, Singh, and Winton (2007) who study a static model where there is less monitoring when investors believe average asset quality to be high (as in a boom), and Axelson and Bond (2015) who study a two-period model where worker outside options and career paths depend on aggregate productivity. These papers do not focus on the (dynamics of) aggregate investment or asset prices.

[4] Appendix E 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.

$\mathbf{y}_z(e) = eY_z + (1-e)y_z$. I denote expected returns by $\hat{Y} = \mathbb{E}_z Y_z$ and $\hat{y} = \mathbb{E}_z y_z$, and returns are higher if $z = h$, $Y_h > Y_l$ and $y_h > y_l$. The private non-pecuniary benefit of shirking is mk. There is a risk-shifting problem: shirking is socially inefficient and increases volatility.

Assumption 1. *$\hat{Y} > \hat{y} + m$ and $y_h \geq Y_h$.*

Financiers are endowed with net worth w_F and issue debt to savers. They lack the expertise to originate risky assets themselves, and instead use their funds to purchase assets from banks. Their outside option is a constant-returns-to-scale *safe technology* that generates a risk-free return normalized to one. The *relative wealth* of financiers is

$$\omega \equiv \frac{w_F}{w_B}. \tag{1}$$

Savers cannot originate assets and are infinitely risk-averse as in Gennaioli, Shleifer, and Vishny (2013) and Caballero and Farhi (2014). Hence they lend to banks and financiers in exchange for risk-free debt. The resulting borrowing constraints for banks and financiers are equivalent to the zero-value-at-risk-collateral constraint derived by Fostel and Geanakoplos (2015) and assumed by e.g. Gromb and Vayanos (2002).[5]

To isolate the key mechanisms, I first assume that savers are deep-pocketed and elastically demand risk-free claims at the exogenous interest rate $1 + r_f$. I model these claims as zero-coupon bounds with face value one and 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 accommodate negative interest rates. The dynamic model in Section 3 endogenizes the risk-free rate using a wealth constraint for savers.

[5] The fact that savers have infinite (as opposed to finite) risk aversion is for tractability. See Section 5 for further discussion.

2.2 Asset Market

Risky assets trade in a competitive secondary market. Similar to Eisfeldt (2004), Bigio (2015), and Kurlat (2013), this market is non-exclusive and subject to asymmetric information: assets are traded individually, banks and financiers interact with multiple counterparties, and financiers neither observe the quality nor the total quantity of assets sold by a given bank. Hence banks have limited commitment with respect to the number of assets they sell. This prevents the perfect separation of types using retention as a signal.

Proceeds from asset sales can serve as collateral when borrowing from savers. I use $\underline{a} \in [0, k]$ to denote the number of assets sold to be used as collateral, and $a_B \in [\underline{a}, k]$ to denote the total number of assets sold. For simplicity, I assume that investment and asset sales take place simultaneously. Appendix D considers an extension with multiple rounds of investing and lending.

In equilibrium, additional collateral leads to higher debt levels and more investment. If debt and investment are observable, then $\underline{a}$ can be inferred from equilibrium outcomes. To account for this, I simply assume that $\underline{a}$ is observable. In practical terms, the assumption means that financiers may use a bank's reliance on asset sales (and the associated increase in leverage) as a signal of asset quality. But it is an imperfect signal because a_b remains unobservable. Hence there is *partial commitment*: pledging assets as collateral reveals a lower bound on asset sales, but does not prevent the bank from secretly selling more than $\underline{a}$ assets.

I define $\chi \equiv \underline{a}/k$ to be the ratio of collateralized asset sales to investment, and show below that χ is a sufficient statistic for banks' incentives to sell more than $\underline{a}$ assets. Hence it is without loss to assume that financiers will condition bids on χ only (rather than $\underline{a}$ and k individually). I use a market structure borrowed from the literature on directed search (e.g. Guerrieri, Shimer, and Wright (2010)) to jointly incorporate perfect competition, partial signaling with endogenous investment quantities, and borrowing constraints. Specifically, I assume that the asset market consists of a set of perfectly competitive submarkets indexed by $\chi \in [0, 1]$. Banks who pledge to sell at least a fraction χ of their assets 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. This construction allows financiers to condition bids on the asset issuer's χ (i.e. their leverage), but preserves perfect competition *conditional on observables*. The outcome is equivalent to allowing financiers to bid competitively for assets indexed by the issuer's χ, but simplifies notation and streamlines the analysis.

The market-clearing price in submarket χ is denoted by $P(\chi)$. A submarket is *active* if at least one bank offers a strictly positive number of assets for sale in this submarket, and 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 banks trading in submarket χ who exert effort is $\phi(\chi)$, and the fraction of good assets in the submarket is $f(\chi)$. The return generated in state z by purchasing assets in submarket χ is $x_z(\chi) = f(\chi)Y_z + (1 - f(\chi))y_z$, and the expected return is $\hat{x}(\chi) = E_z x_z(\chi)$.

2.3 Decision Problems

2.3.1 Banks

Banks maximize expected utility u_B by choosing investment $k \geq 0$, promised asset sales $\underline{a} \geq 0$, actual asset sales $a_B \in [\underline{a}, k]$, debt $b_B \geq 0$, and effort $e \in \{0, 1\}$. Recall that a_B may be larger than $\underline{a}$, and that choosing $a_B > \underline{a}$ does not impact the asset price at which the bank trades because a_B is unobservable. Let $m^*(e) = (1 - e)m$ denote the realized private benefit, which is equal to zero if banks exert effort ($e = 1$). Expected bank utility is

$$u_B = \sum_z \pi_z \max \left\{ \mathbf{y}_z(e)\,(k - a_B) - b_B + P(\chi)a_B, 0 \right\} + m^*(e)k \quad \text{where} \quad \chi = \frac{a}{k} \tag{2}$$

The max-operator imposes limited liability: banks cannot pay back more than their end-of-period net worth. Since savers only buy risk-free debt, banks must satisfy the following solvency constraint given equilibrium effort e^*,

$$\mathbf{y}_z(e^*)\,(k - \underline{a}) - b_B + P(\chi)\underline{a} \geq 0 \qquad \forall z \tag{3}$$

This can be interpreted a collateral constraint: all borrowing must be secured by either the worst-case return of retained assets given equilibrium effort, or by the proceeds from asset sales. Later I show that, if assets are traded, any deviation to $a_b > \underline{a}$ will not violate the solvency constraint. The bank budget constraint is

$$k \leq w_B + qb_B \tag{4}$$

Banks' decision problem is to maximize (2) subject to (3) and (4). I do not impose an incentive-compatibility constraint because banks may shirk on the path of play.

2.3.2 Financiers

Financiers maximize expected utility u_F by choosing asset purchases $a_F(\chi) \geq 0$ in every submarket, investment in the safe technology $s_F \geq 0$, and debt $b_F \geq 0$. Expected financier utility is

$$u_F = \sum_z \pi_z \max\left\{ \int x_z(\chi) a_F(\chi) + s_F - b_F, 0 \right\} \tag{5}$$

The max-operator imposes limited liability: financiers cannot pay back more than their end-of-period net worth. Since savers only buy risk-free debt, financiers must satisfy the following solvency constraint:

$$b_F \leq \int x_l(\chi) a_F(\chi) + s_F. \tag{6}$$

Hence all borrowing must be collateralized by the worst-case return of purchased assets or the safe technology. The financier budget constraint is

$$\int P(\chi) a_F(\chi) + s_F \leq w_F + q b_F \tag{7}$$

Financiers' decision problem is to maximize (5) subject to (6) and (7).

2.4 Equilibrium

A competitive equilibrium is defined as follows.

Definition 1. *A* competitive equilibrium *is a price schedule schedule $P : [0,1] \to \mathbb{R}_+$, a set of active submarkets $\mathcal{A}$, 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) the bank portfolio and effort decision jointly solve banks' decision problem given P and $\mathcal{A}$, (ii) the financier portfolio solves financiers' decision problem given P, $\mathcal{A}$, and bank effort decisions, and (iii) all active submarkets clear.*

Market-clearing is naturally imposed only in active submarkets. Because agents are price-takers, an unfortunate side effect is that one can render any arbitrary submarket χ

inactive by conjecturing $P(\chi) = 0$. Hence it is possible to construct competitive equilibria with essentially arbitrary $\mathcal{A}$, including no-trade equilibria. This conceptual issue is similar to standard signaling models where off-path beliefs can sustain a wide range of equlibria. I first characterize general model properties that hold for any set of active submarkets (Section 2.4.1), and then introduce a refinement to characterize the equilibrium set of active submarkets (Section 2.4.2).

2.4.1 Excessively High Asset Prices

The next proposition states the key property of the model: banks prefer to shirk if asset prices are too high. The result obtains because banks can deviate to selling more assets than pledged as collateral without affecting the asset price, allowing for double-deviations to selling and shirking. This deviation is particularly attractive when the bank is highly leveraged (has high χ).

Proposition 1. *There exists a threshold $\bar{\chi}$ such that banks shirk if $\chi > \bar{\chi}$, where*

$$\bar{\chi} = 1 - \left(\frac{m}{\hat{Y} - \hat{y}}\right). \tag{8}$$

If $\chi \leq \bar{\chi}$, there exists a minimum threshold price $\bar{P}(\chi)$ such that banks sell all assets ($a_B = k$) and shirk if $P(\chi) > \bar{P}(\chi)$. This threshold price is

$$\bar{P}(\chi) = \hat{Y} - \frac{m}{1-\chi}. \tag{9}$$

Banks are indifferent between shirking and effort if $P(\chi) = \bar{P}(\chi)$, and $\bar{P}(\chi)$ is decreasing in χ.

The first statement follows from a simple skin-in-the-game rule: if banks sell too large a fraction of their assets, they are longer exposed to the consequences of their effort decision, and so they prefer to shirk. The upper bound $\bar{\chi}$ can therefore be interpreted as a measure of the *liquidity* of bank assets: liquidity is high and all assets can be sold when the moral hazard problem is weak ($m \to 0$), while liquidity is low and no assets can be sold when the moral hazard problem is severe ($m \to \hat{Y} - \hat{y}$).

The intuition for the second result is as follows. Suppose a bank has pledged $\underline{a}$ assets as collateral, issued b_B in debt, and invested k. Since a_b is unobservable, 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 engage in a "double deviation" where it simultaneously sells more assets than it promised and shirks as a result. Because the objective function is linear, the optimal deviation is to sell *all* assets ($a_B = k$) and shirk. So the utility obtained upon a double deviation is $P(\chi)k - b_B + mk$. The alternative is to sell just as many assets as promised ($a_B = \underline{a}$) and exert effort, in which case utility is $\mathbb{E}_z Y_z(k - \underline{a}) - b_B + P(\chi)\underline{a}$. Since the bank sells more assets when it deviates, the deviation is profitable if the asset price is sufficiently high. It is important to note that savers are not hurt by such a deviation: since bank debt is collateralized by the returns from the first $\underline{a}$ units of assets sold, selling more assets does not violate the solvency constraint.

The next result is that the asset price *must* attain its upper bound if financiers are sufficiently wealthy. Because solvency constraints impose bounds on the supply and demand of assets that are functions of w_F and w_B, market-clearing prices reflect the relative wealth of financiers and banks. Since the upper bound on the asset price is below $\hat{Y}$, financiers earn profits if they buy good assets at the maximal price.[6] Hence asset prices are bid up to $\bar{P}(\chi)$. The following proposition formalizes this result, and also shows that any equilibrium *must feature shirking* if relative wealth grows further. The intuition is that asset prices cannot rise beyond $\bar{P}(\chi)$, since all banks would then shirk. But if all banks

[6] The next section shows that banks are willing to sell good assets at a price less than $\hat{Y}$ because there is a shadow cost of holding risk exposure.

shirk, financiers are no longer willing to pay $\bar{P}(\chi)$. The result is a mixed strategy equilibrium in which $P(\chi) = \bar{P}(\chi)$ in every active submarket, and excess demand is cleared by an increase in the fraction of banks who shirk, since shirking banks sell more assets than those who exert effort. That is, increasing the fraction of shirking banks leads to higher supply at fixed prices.

Proposition 2. *For any set of active submarkets $\mathcal{A}$, there exists a threshold $\bar{\omega}(\mathcal{A})$ such that a strictly positive fraction of banks shirk in equilibrium if $\omega > \bar{\omega}(\mathcal{A})$. If $\omega \geq \bar{\omega}(\mathcal{A})$, the fraction of shirking banks is increasing in ω*

Shirking induced by excessive financier demand is *socially inefficient*. This is because total investment cannot increase if asset prices are at the upper bound, and banks shirk on their marginal investments. Hence excess aggregate demand for risky assets leads to a deterioration in average quality without raising total quantities. This result is driven by a pecuniary externality: financiers do not internalize their impact on asset prices. Imposing a cap on asset purchases thus leads to a strict Pareto-improvement.

Proposition 3. *Let $\omega > \bar{\omega}(\mathcal{A})$. A social planner can strictly improve financier utility without harming bank utility by imposing a cap on financier's purchases of risky asset of the form $a_F \leq \bar{a}$.*

Asset sales also affect the distribution of aggregate risk across banks and financiers. Financiers exposure to aggregate risk (as proxied by the variance of individual end-of-period wealth $w'_F(z)$) increases in the quantity of risky assets it purchases, while bank exposure (as proxied by the variance of individual end-of-period wealth $w'_B(z)$) decreases. Cheap leverage (high q) further increases exposure. In the dynamic model, this effect leads to growth in relative financier wealth after good aggregate shocks.

Proposition 4. *If financiers optimally allocate funds across submarkets, the variance of $w'_F(z)$ is strictly increasing in $\int a_F(\chi)$. If financiers' solvency constraint* (6) *binds, then it is also increasing* *in q. Given investment k, the variance of $w'_B(z)$ is decreasing in a_B.*

2.4.2 Endogenous Credit Volumes

I now derive equilibrium limits on bank borrowing and show that asset sales allow banks to expand investment. I do so under the assumption that banks do not engage in the double deviation to shirking and selling discussed in Proposition 1. This leads to a characterization of credit volumes when asset prices are not "too high," and highlights the beneficial role of asset sales. Since moral hazard may reduce the pledgeable value of assets, I define the collateral value of a unit of risky investment

$$\Delta \equiv \frac{\pi_h Y_h + \pi_l Y_l - \pi_h y_h}{\pi_l}. \tag{10}$$

The numerator is the difference in expected returns between good and bad assets, assuming that the solvency constraint binds in the bad state conditional on shirking so that the bank receives none of the cash flows. This is the appropriate definition because the maximum borrowing capacity is attained if the solvency binds in the bad state conditional on shirking. It is scaled by π_l because there is less risk of default if the bad state is unlikely. In the absence of moral hazard, the pledgeable return on risky assets is the worst case return Y_l. Hence $Y_l - \Delta \geq 0$ is a measure of the degree to which moral hazard lowers pledgeability. It is increasing in $\pi_h(y_h - Y_h)$, the expected benefit of risk shifting.

I derive the borrowing constraint by solving for the maximum borrowing capacity consistent with solvency constraint (3). Since bad assets have lower worst-case returns than good assets, this turns out to require that banks exert effort. I also show that banks are willing to sell assets to obtain borrowing capacity only if asset prices are not too low. Consistent with the proposition, I later focus on equilibria with locally differentiable price functions.

Proposition 5. *Let $a_b = \underline{a}$ and $\chi \leq \bar{\chi}$. Bank borrowing is limited by the constraint*

$$b_B \leq \Big(\Delta - m/\pi_l\Big)k + \Big(P(\chi) - \Delta\Big)\underline{a}, \tag{11}$$

In addition, (11) *ensures that effort is incentive compatible conditional on $a_b = \underline{a}$ and $\chi \leq \bar{\chi}$. If constraint* (11) *binds and $P(\chi)$ is differentiable at the origin, there exists a threshold price $\underline{P}(q)$ such that banks are willing to increase borrowing capacity by selling assets if an only if $P(0) \geq \underline{P}(q)$. The threshold price is strictly decreasing in q, and is given by*

$$\underline{P}(q) \equiv \frac{\hat{Y} + \Delta\Pi(q)}{1 + \Pi(q)} \qquad \textit{where} \qquad \Pi(q) = \frac{q\hat{Y} - 1}{1 - q(\Delta - m/\pi_l)} \tag{12}$$

Even if banks do not engage in the double deviation to selling and shirking (as is the case if $a_b = \underline{a}$ and $\chi \leq \bar{\chi}$), their ability to issue debt is still limited by the need to stay solvent on the path of play. Since $y_l < Y_l$, the solvency constraint is particularly tight when banks do not exert effort, and banks would prefer to exert effort if they are solvent conditional on shirking. At the same time, banks would strictly prefer to shirk if the solvency constraint were to bind conditional on effort. Hence bank borrowing is not directly limited by the solvency constraint but rather by an incentive-compatibility constraint that can be restated as borrowing constraint (11). Borrowing capacity is increasing in $\underline{a}$ because banks are less willing to engage in risk-shifting if their balance sheet is less risky, and particularly so if the asset price is high. That is, asset sales allow banks to pledge a larger share of their investments as collateral, and so savers are willing to lend more. This captures the intuition that asset sales can be used to secure debt (Gorton and Metrick (2012b)). More generally, the fact that the borrowing constraint ensures effort *conditional* on $a_b = \underline{a}$ and $\chi \leq \bar{\chi}$ implies that shirking can occur *only* if high asset prices trigger the double deviation.

The second part of the proposition characterizes conditions such that banks *want* to sell good assets to increase borrowing capacity. The cost of selling is that the asset price must lie strictly below the expected return $\hat{Y}$ by Proposition 1. Banks are willing suffer this loss only if the price is not too low relative to the shadow value of additional borrowing capacity. This shadow value can be measured using levered bank profits $\Pi(q)$. Since borrowing costs are decreasing in q, so is $\underline{P}(q)$. Thus a useful interpretation of the asset market arrangement is that financiers sell balance sheet capacity to banks and are compensated by relatively low asset prices. Banks in turn sell assets below expected value to defray the shadow of cost of risk. In the dynamic model, rents earned by financiers contribute to the growth of relative wealth during good times.

The next step is to characterize bank's optimal choice of asset sales. The equilibrium concept defined above is not sufficiently restrictive to do so. This is because market-clearing is imposed only in active submarkets, and so off-equilibrium prices can be chosen to impose arbitrary restrictions on the set of active markets. For example, a conjecture of $P(\chi) = 0$ for all $\chi \neq \chi^*$ ensures that trade occurs in at most submarket χ^*. The same conceptual issue arises when determining off-equilibrium beliefs in general signaling games. Dubey and Geanakoplos (2002) develop a refinement suitable for the general equilibrium framework with multiple submarkets I use here. Their idea is to introduce a fictitious buyer who purchases an infinitesimal quantity in inactive markets at a price that is linked to *market-clearing* prices in active submarkets. So if trade occurs in submarket χ^* at price $P^*(\chi^*)$, then off-equilibrium prices in inactive submarkets that are sufficiently "similar" to χ^* must be close to $P^*(\chi^*)$. There is an intuitive notion of market similarity here: since incentives to engage in the double deviation to shirking and selling are increasing in χ, prices should not decrease when a seller moves from an active submarket to a slightly lower χ. (That is, prices should not decrease if the bank chooses lower leverage). To transparently characterize the main mechanism, I use the simplest possible version of this refinement: the asset price is constant across all submarkets that are *consistent with bank effort* (as defined below), and equal to zero otherwise. This refinement permits maximal trade in good assets, but provides market discipline in that banks who are known to shirk with probability 1 receive a price of zero.

Proposition 1 shows that banks shirk for sure if $\chi > \bar{\chi}$. Proposition 5 implies that there does not exist an equilibrium with trade in good assets in submarket χ if $\bar{P}(\chi) < \underline{P}(q)$. Hence if we define $\hat{\chi}(q)$ to be the threshold submarket where $\bar{P}(\hat{\chi}(q)) = \underline{P}(q)$, we can define the set of submarkets *consistent with bank effort* $\mathcal{E}$ as the union of all submarkets in which banks *may* choose to exert effort in equilibrium,

$$\mathcal{E} = [0, \chi^e] \qquad \text{where} \qquad \chi^e \equiv \min\{\hat{\chi}(q), \bar{\chi}\}. \tag{13}$$

That is, if trade takes place in submarket $\chi' \neq \mathcal{E}$, traded assets are bad with probability one, while if assets are traded in submarket $\chi'' \in \mathcal{E}$, they may be good. This leads to the following definition of refined competitive equilibrium.

Definition 2. *A refined equilibrium is a competitive equilibrium such that* $P(\chi) = p > 0$ *for all* $\chi \in \mathcal{E}$ *and* $P(\chi) = 0$ *for all* $\chi \neq \mathcal{E}$.

Since χ is related to leverage, this has the intuitive interpretation that financiers do not purchase from banks who are too highly levered. It also provides the closest analogue to the *anonymity* of financial markets assumed in Bigio (2015), Kurlat (2013) and Eisfeldt (2004), which automatically generates a flat pricing schedule for traded assets. Nevertheless, the basic mechanisms do not hinge on the particulars of the refinement. As long as banks do not suffer large price penalties upon a switch to a submarket when doing so does not harm incentives, they will use asset sales to increase borrowing capacity. Moreover, Proposition 1 shows that shirking obtains when financiers are sufficiently wealthy for any refinement. In addition, Appendix C considers an alternative refinement in which asset prices are strictly decreasing in $\chi \in \mathcal{E}$, and shows that this leads to more shirking.

I focus on the interesting case where $\mathcal{E}$ is non-empty. This requires that $\hat{\chi}(q) > 0$. Since $\underline{P}(q)$ is decreasing in q, this in turn requires that q is sufficiently close to one.

Assumption 2. *q is such that $\hat{\chi}(q) > 0$.*

The next proposition provides a full characterization of refined competitive equilibrium. Given a fraction ϕ of banks who exert effort, I define aggregate output in state z to be $\mathbf{Y}_z = (\phi Y_z + (1-\phi) y_z)\, k$, and expected aggregate output to be $\mathbb{E}_z \mathbf{Y}_z$.

Proposition 6. *There exists a unique symmetric refined equilibrium characterized by a single active submarket χ^*, asset price p^*, and two thresholds for relative wealth, $\underline{\omega}$ and $\bar{\omega} \geq \underline{\omega}$. This equilibrium has the following structure.*

(i) If $\omega \leq \underline{\omega}$, then $p^* = \underline{P}(q)$, and all banks exert effort. Total investment and expected output are increasing in w_F. If $y_h > Y_h$ and q is sufficiently close to one, a marginal wealth reallocation from banks to financiers increases aggregate investment.

(ii) If $\omega \in (\underline{\omega}, \bar{\omega}]$, then banks sell the highest fraction of assets consistent with effort, $\chi^* = \chi^e$, and all banks exert effort. Total investment, expected output, and p^* are increasing in w_F.

(iii) If $\omega > \bar{\omega}$, then $\chi^* = \chi^e$, the asset price is at its upper bound, $p^* = \bar{P}(\chi^*)$, and a strictly positive fraction of banks shirk. Investment is constant in w_F and the fraction of shirking banks is increasing in ω. Expected output is decreasing in w_F, and the variance of aggregate output is increasing in w_F.

(iv) $\bar{\omega} > \underline{\omega}$ if $\bar{\chi} > \hat{\chi}(q)$, and $\bar{\omega} = \underline{\omega}$ otherwise.

The intuition for the upper threshold was established by Proposition 1. The lower threshold is such that the equilibrium price equals its lower bound if $\omega = \underline{\omega}$, in which case financiers extract all rents from asset sales. The last part of the proposition states that it may be the case that the asset price's upper and lower bound on the price coincide, $\underline{P}(q) = \bar{P}(\chi^*)$, in which case shirking may occur even though prices haven't appreciated above the lower bound. This is the case if q is low and the value of additional borrowing capacity is small. In such scenarios, asset market trading may thus hamper asset quality without first boosting aggregate investment.

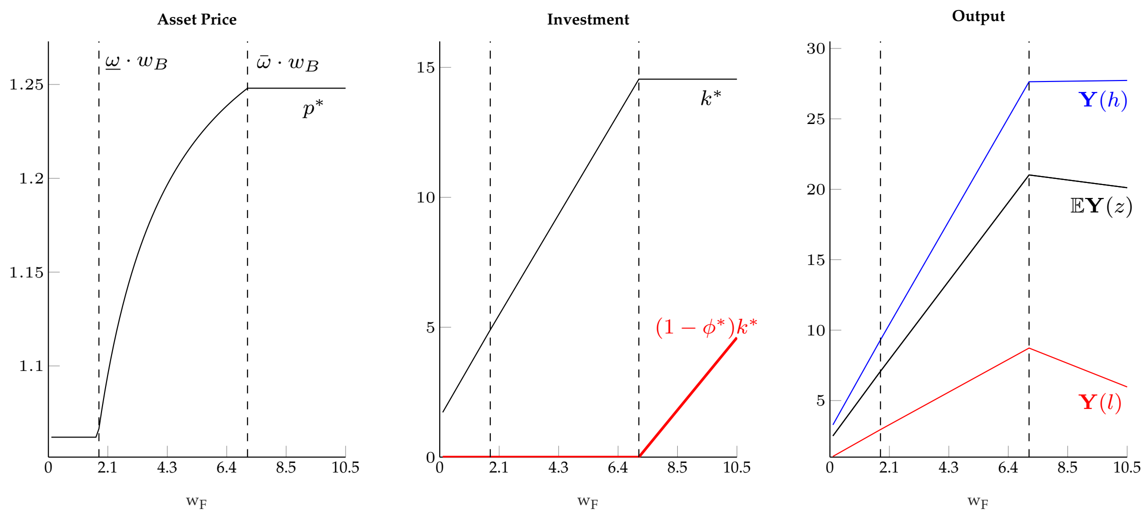

Figure 1: Equilibrium outcomes in the static model as a function of w_F. Parameter values: $\pi_h = 0.65$, $Y_h = 1.9$,$y_h = 1.92$, $Y_l = 0.6$, $y_l = 0$, $m = 0.05$, $q = 1$.

Figure 1 illustrates the proposition by plotting equilibrium outcomes as a function of w_F. Parameters are such that $\bar{\chi} > \hat{\chi}(q)$. The vertical dashed lines show the two thresholds $\underline{\omega}$ and $\bar{\omega}$. The left panel plots the asset price, the center panel plots total investment and investment in bad assets (in red), and the right panel plots output (in expectation and state by state). Below the first threshold, the asset price is constant but investment increases because banks can borrow more if they sell a larger fraction of their assets. Since all banks exert effort, output increases state by state and in expectation. Moreover, aggregate investment increases more sharply in w_F than it would have after an increase in w_B. This is because asset sales provide a form of commitment against bank moral hazard: while the collateral value of a risky asset is Δ if held by banks, it is Y_l if held by financiers. Under the stated conditions, we have $\Delta < Y_l$, and so financiers can pledge more. That is, financiers are "specialists in leverage" that help to efficiently expand credit volumes.

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. In the upper region, instead, the asset price is fixed at upper bound $\bar{P}(\chi^e)$. Investment is now constant because further increases in w_F 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)), the equilibrium features a form of *underinvestment*: investment does not increase even though risk-bearing wealth does. This is because prices are bounded from above, which implies that the additional collateral obtained via asset sales is also bounded. Note also that productivity shocks alone may trigger shirking, fixing ω.

Corollary 1 (Productivity Shocks)**.** *Let* $\chi^e = \bar{\chi}$. *Then* $\frac{\partial \bar{\omega}}{\partial Y_h} > 0$.

2.4.3 Implications for Regulation

I now briefly comment on the model's implications for regulation. The frequent use of bank capital requirements is often motivated by the idea that they are an effective means of curbing excessive risk-taking in models where fire sale externalities lead to over-borrowing ex-ante (see e.g. Bianchi and Mendoza (2012)). The next proposition shows that their effects are quite different in this model: binding bank capital requirements may *cause* inefficient moral hazard by raising asset prices. The intuition is that banks sell risky assets to increase leverage, and so they sell less if leverage is capped. This leads to a reduction in asset supply that raises the asset price and hampers incentives. Put differently, regulation may interfere with market discipline by artificially inflating asset prices.

Proposition 7. *Assume that banks are restricted to originating no more than a multiple λ of their net worth, $k \leq \lambda w_B$. There exist parameters such that a binding capital requirement triggers shirking even though no bank would shirk without the requirement.*

Which policies might work better in this model? The root source of inefficiency is that markets are incomplete and trading is non-exclusive and subject to limited commitment. A direct solution therefore is to decrease information frictions in asset markets. If this is infeasible, the problem can be addressed by regulating asset supply using a skin-in-the-game rule $\chi \leq \chi^e$, or by regulating asset demand. These approaches are equivalent in the model, but the information required to implement in practice may. Skin-in-the-game rules require measures of bank risk exposure for each asset class, and regulators must be able to observe bank trades in real time. Restrictions on asset demand can be implemented by taxing trades, regulating the asset side of financier balance sheets, or directly redistributing financier equity. This would require regulating a wider swath of financial investors.

3 Dynamic Model

I now embed the static model in an infinite-horizon model to study the endogenous evolution of the wealth distribution. I do so in two steps. First, I consider an overlapping generations economy where each generation of banks and financiers lives for one period and the bond price q is fixed. This structure permits me to analytically characterize the key determinants of the evolution of the wealth distribution. Second, I consider a model with long-lived banks and financiers where the bond price is endogenous because savers are wealth-constrained. Due to the presence of aggregate risk, I must rely on numerical illustrations for this version of the model, but I can draw out implications for the joint behavior of asset prices and the risk-free rate, and show robustness of the basic mechanisms from the overlapping generations model.

In both dynamic models, time is discrete and indexed by $t = 0, 1, \ldots$, and there are successive generations of savers who each live for one period. All technologies are as in the static model, and the aggregate shock is i.i.d. Invested capital k depreciates at the end of every period, but wealth can be carried forward across periods. These assumptions allow me to clearly describe the role of asset market trading in shaping the evolution of wealth. Given realized aggregate shock z, I denote relative wealth at the beginning of period t by $\omega_t(z)$, and initial relative wealth by ω_0. When there is no risk of confusion, I denote current (beginning of period) relative wealth simply by ω, and updated relative wealth at the end of the period by $\omega'(z)$.

3.1 Overlapping Generations

I begin by considering an overlapping generations framework where banks and financiers live for one period and q is fixed for all periods. Each generation of banks and financiers has the same objective as in the static model, namely to maximize expected utility over the current period, and each agent passes on their end-of-life wealth on to the next generation.[7] Model dynamics are then equivalent to an infinitely-repeated version of the static model, with the wealth distribution the only link across periods.

Since wealth dynamics depend on leverage, it will be useful to restate the financier

[7] To make this precise, assume that each new generation is born with an endowment of consumption good that it trades for the old generation's capital at a unit price. The old consume the consumption good and exit, the young use the obtained capital to invest in the current period.

solvency constraint as $b_F = \gamma\left[x_l(\chi)a_F + s_F\right]$, where γ encodes the optimal borrowing decision. Let $p^*(\omega)$ denote the asset price given ω, and $\hat{x}^*(\omega)$ the expected return on purchased assets. Then it is easy to verify that optimal borrowing satisfies $\gamma^*(\omega) = 1$ if $\hat{x}^*(\omega)/p^*(\omega) > 1/q$, $\gamma^*(\omega) = 0$ if $\hat{x}^*(\omega)/p^*(\omega) < 1/q$, and $\gamma^*(\omega) \in [0, 1]$ otherwise. Assuming that $\chi^e = \bar{\chi}$, we can then summarize equilibrium leverage ratios at $\omega = \bar{\omega}$ by

$$\lambda_F = \frac{a_F(\bar{\omega})}{w_F} = \frac{1}{p^*(\bar{\omega}) - q\gamma^*(\bar{\omega})Y_l} \qquad \text{and} \qquad \lambda_B = \frac{k(\bar{\omega})}{w_B} = \frac{1}{1 - q(1-\bar{\chi})y_l - q\bar{\chi}p^*(\bar{\omega})}, \tag{14}$$

and the associated rates of return by $r_F(z) = Y_z - \gamma^*(\bar{\omega})Y_l$ and $r_B(z) = Y_z - y_l$. Finally, let $\lambda_B^0 = \lim_{\bar{\chi}\to 0} \lambda_B$ denote bank leverage in the limit where risky assets become highly illiquid, and $\epsilon_B^0 = \lim_{\bar{\chi}\to 0} \frac{\partial \lambda_B}{\partial \bar{\chi}}/\lambda_B^0$ the associated semi-elasticity. With these definitions in hand, the next result characterizes equilibrium wealth dynamics. A corollary shows that asset quality may decline after a sequence of good aggregate shocks.

Proposition 8. *Given current relative wealth ω, the dynamics of relative wealth in the overlapping generations economy are as follows.*

(i) If $\omega < \underline{\omega}$, there exists a cutoff bond price $q^C < 1$ such that $\omega'(h) > \omega$ if and only if $q \geq q^C$. The cutoff price q^C is strictly decreasing in ω.

(ii) If $\omega \in (\underline{\omega}, \bar{\omega}]$, then $\omega'(z) = \left(\frac{r_F(z)}{r_B(z)}\right)\left(\frac{\bar{\chi}}{1-\bar{\chi}}\right)$, $\bar{\omega} = \bar{\chi}\left(\frac{\lambda_B}{\lambda_F}\right)$, and $\omega'(h) > \bar{\omega}$ if and only if

$$\bar{\chi} < \chi^0 \equiv \frac{r_F(h)\lambda_F - r_B(h)\lambda_B^0}{r_F(h)\lambda_F\epsilon_B^0 - r_B(h)\lambda_B^0}. \tag{15}$$

Moreover, $\lim_{q\to 1} \chi^0 > 0$ if and only if $r_F(h)\lambda_F > r_B(h)\lambda_B^0$.

(iii) If $\omega'(h) > \omega$ for $\omega \in (\underline{\omega}, \bar{\omega}]$, then $\omega'(h) > \omega$ for any $\omega > \bar{\omega}$.

Corollary 2. *Assume that* (15) *holds given $q = 1$ and consider a sequence of N periods of good shocks. If q is sufficiently close to 1, there exists a smallest integer T such that $\omega_N > \bar{\omega}$ if $N \geq T$.*

Moreover, the asset price and investment are monotonically increasing in t for all $t \leq T$, while the fraction of shirking banks is increasing in t for all $t \in [T, N]$.

That is, as long as the risk-free rate is sufficiently low, relative wealth grows after good shocks even if ω_0 is arbitrarily small. If the interest rate is relatively high, instead, then financiers must be wealthy to begin with in order to buy a sufficiently large share of the stock of aggregate risk. The intuition is that low interest rates permit financiers to take on relatively more leverage than banks since they are not subject to moral hazard. Taken together, this leads to an increase in relative wealth after good shocks. Given the appropriate parametric restrictions, moreover, relative wealth continues to grow even when $\omega > \underline{\omega}$. Since banks shirk when relative wealth is sufficiently high, a sequence of *good* aggregate shocks may thus lead to an inefficient decline in asset quality. It follows that the model can generate credit and asset price booms with gradually decreasing asset quality. Interestingly, this is the case precisely *because* asset markets efficiently allocate aggregate risk exposure to financiers when relative wealth is low (indeed, this motive for trade is particularly strong when q is high). Figure 2 illustrates this mechanism by plotting $\omega'(z)$ given current relative wealth ω. Parameters are such that $\omega > \bar{\omega}$ given a sufficiently long sequence of aggregate shocks.

The second part of Proposition 8 clarifies the role of asset liquidity in shaping the dynamics of relative wealth and asset quality. When $\bar{\chi}$ is high, banks can sell a larger fraction of their assets. This has two effects: the first is that financiers purchase a larger share of the stock of aggregate risk exposure, which contributes to an increase in relative wealth after good shocks. The second is that it leads to increase in the *supply* of assets, which lowers asset prices and makes it more difficult for financiers to reach level of wealth required to trigger shirking (that is, $\bar{\omega}$ increases). The proposition shows that the second effect dominates, and that financiers grow beyond the upper bound $\bar{\omega}$ only if $\bar{\chi}$ is small

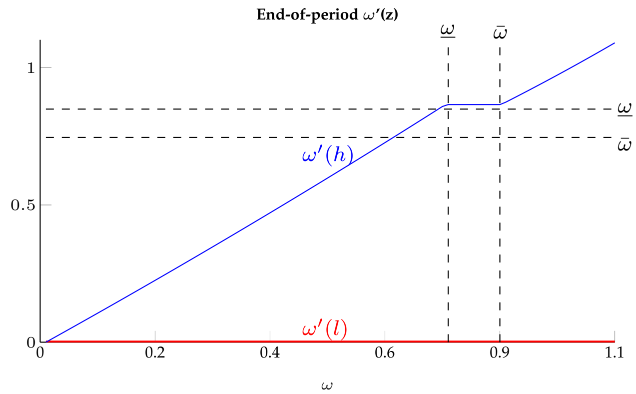

Figure 2: The dynamics of relative wealth in the overlapping generations model. Parameter values: $\pi_h = 0.5$, $Y_h = 2.1$,$y_h = Y_h$, $Y_l = 0.4$, $y_l = 0.075$, $m = 0.08$, $q = 1$.

and they earn sufficiently high levered returns relative to banks when asset liquidity is low (as defined by Condition (15)). A weaker moral hazard problem (a decrease in m that increases $\bar{\chi}$) can therefore simultaneously increase asset market volumes and relative wealth but reduce the scope for excess asset sales. In this sense, hidden trading on asset markets can reduce welfare only if the moral hazard problem is sufficiently severe.

As illustrated by Figure 2, 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. An important difference to the canonical literature on fire sales is that *increases* in current wealth may lead to sharper declines in future wealth.

Corollary 3. *Let* $\omega > \bar{\omega}$ *and* $s_F = b_F = 0$. *Then* $w'_F(l)$ *is strictly decreasing in* w_F.

3.1.1 Implications for Monetary Policy

The first statement of Proposition 8 reveals a subtle dynamic interaction between wealth and interest rates that is relevant for monetary policy: because high wealth is a substitute for low risk-free rates when purchasing assets, *temporary* shocks to the risk-free rate may be sufficient to induce sustained growth in relative wealth.

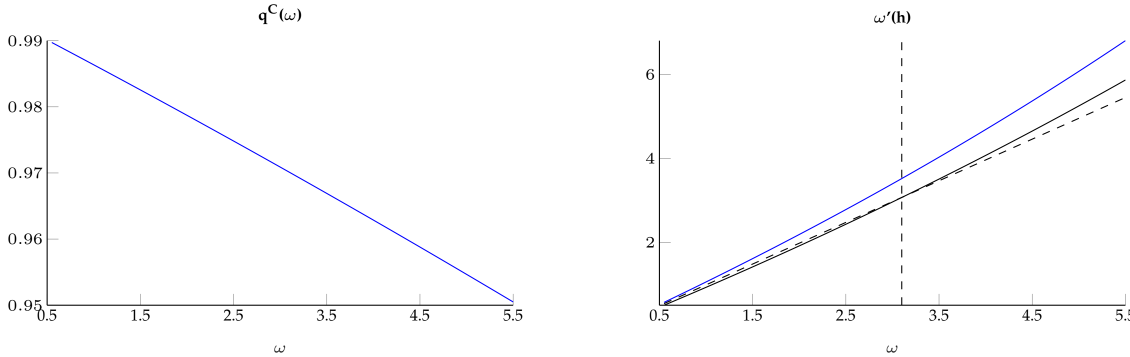

Figure 3: Persistent effects of temporary interest rate shocks. Parameters: $\pi = 0.65, Yh = 2.5, y_h = 2.5, Y_l = 0.815, y_l = 0, m = 0.03$. *Bond prices:* $q^* = 0.97$, *and* $q^s = 0.99$.

Figure 3 illustrates this mechanism graphically. The left panel depicts the downward-sloping cutoff price q^C from Proposition 8 as a function of ω. The right panel depicts the evolution of relative net worth after good shocks, also as a function of ω. I consider a temporary increase in the bond price from baseline level q^* to $q^s > q^*$. The lower black solid line plots $\omega'(h)$ in the baseline. The upper blue solid line plots $\omega'(h)$ given the interest rate shock (where $q = q^s$). The upward sloping dashed line is the 45-degree line. Hence relative wealth grows if the plot lies above the 45-degree line, and the vertical dashed line depicts the cutoff level for relative wealth such that $\omega'(h) \geq \omega$ in the baseline with $q = q^*$.

For any initial condition to the left of this cutoff, relative wealth contracts in the baseline but grows given the interest rate shock. Hence there exist initial conditions such that relative wealth begins to grow if and only if interest rates are low. If the shock lasts until ω crosses the vertical dashed line, relative wealth continues to grow even when the stimulus ends. This result can be interpreted as a dynamic risk-taking channel of monetary policy, or as a model of the long-run response to a temporary saving glut. It also suggests an asymmetry in the conduct of policy: the contraction required to reign in an asset price boom is larger than the initial stimulus.

3.2 Long-lived Agents

I now consider the dynamic model with infinitely-lived banks and financiers who are endowed with initial net worth $w_{B,0}$ and $w_{F,0}$, respectively, and have discount factor $\beta \in (0,1)$. To study the joint dynamics of asset prices and the risk-free rate, I endow savers with a wealth constraint and the option to invest in the safe technology. Note that an endogenous risk-free rate provides a counterweight to sustained financier growth because higher interest rates reduce financier's relative leverage advantage. For simplicity, I assume that savers live for one period and that each saver generation is endowed with fixed net worth w_S, which they they can either invest in risk-free debt b_S or in the safe technology s_S. The fact that savers do not accumulate wealth leads to a faster increase of the risk-free rate, and works against the basic mechanism that financiers grow wealthy during good times because they are highly levered.

Observation 1. *The optimal saver portfolio satisfies $b_S^* = w_S/q$ and $s_S = 0$ if $q < 1$, and $b_S^* \in [0, w_S/q]$ and $s_S^* = w_S - qb_S^*$ if $q = 1$.*

Since capital k fully depreciates at the end of each period[8], and the aggregate shock is i.i.d. the relevant state variables are the aggregate wealth distribution $\mathbf{w} = \{\mathbf{w}_B, \mathbf{w}_F, \mathbf{w}_S\}$ and own wealth. As in Gertler and Karadi (2011), banks and financiers accumulate wealth 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 next period. For simplicity, entrants are endowed with the same net worth as those who exit, and I normalize the private benefit of shirking to βm. As in the static model, the only securities are one-period riskless bonds

[8] If capital fully depreciates, a single realization of the good shock is enough to circumvent the negative consequences of having originated bad assets. If assets are durable, shirking instead leads to a persistent decline in the quality of the capital stock. Hence durable assets exacerbate the key inefficiency.

and risky assets are traded under limited commitment[9] The equilibrium concept is as before, but augmented by bond market-clearing condition $b_S = b_F + b_F$.

I describe the value functions and computational algorithm in Appendix B. The key observation is that value functions are linear in own net worth, $V_\theta(w_\theta, \mathbf{w}) = v_\theta(\mathbf{w})w_\theta$ for $\theta \in \{B, F\}$. Conditional on $v_\theta(\mathbf{w})$, objective functions are thus linear, and equilibrium within a given period can therefore be solved as in the static model given a suitable modification of the expected value of risky and safe investments. The key difference is that the marginal value of net worth fluctuates with the wealth distribution. 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.

Since interest rates are low when bank and financier net worth is low, the marginal value of net worth is particularly high after bad shocks that deplete capital. (For financiers, an additional consideration is that the marginal value of net worth is low when *other* financiers are wealthy and asset prices are high.) Financiers with precautionary motives are therefore less willing to buy risky assets on the margin, and particularly so if banks shirk. While this leads to stronger market discipline against shirking, there is a countervailing force: because selling assets lowers banks' risk exposure, the cutoff price at which they engage in the double-deviation to shirking and selling is *lower* than in the static model. As a result, the dynamic model still features a well-defined interior threshold price at which banks engage in inefficient shirking.

There is also a capital accumulation externality that is conceptually similar to Loren-

[9] That is, there are no long-term contracts or reputational concerns. While reputations are a means of overcoming commitment problems in financial markets, they are typically fragile (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. Thus, reputational concerns at best partially attenuate and at worst strengthen motives for nefarious hidden trading.

zoni (2008) but manifests itself in high rather than low asset prices: because agents take the evolution of the aggregate wealth distribution as given, individual financiers do not internalize that their purchases of aggregate risk exposure may lead to excessive growth in relative wealth over time. The economy thus transitions to regions of the state space where banks shirk even if agents are fully forward-looking.

3.3 Simulations

I now use simulations to show that the dynamic model with long-lived agents can generate asset price and credit booms with deteriorating credit quality: periods of growing credit volumes with increasing relative financier net worth and a growing fraction of shirking banks. Hence the full dynamic model replicates the qualitative mechanisms established in the overlapping generations framework. Since credit volumes increase only if net worth increases, I study a sequence of good shocks.

Figure 4 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 the wealth distribution is depicted at the beginning of the period.) Initial relative wealth is $\omega_0 = 0.5$. Parameters are such that $\omega_0 < \underline{\omega}$, and initial prices are $q = 1$ and $p = \underline{P}(1)$. Recall that there may be shirking even if the asset price is at its lower bound as long as the lower bound coincides with the upper bound given $\chi = \chi^*$. By Proposition 6, this occurs if $\chi^e = \hat{\chi}(q)$. In the dynamic model, this is an endogenous condition that depends on the law of motion for the wealth distribution. Hence it cannot be verified ex-ante. Yet even in this case, the fraction of shirking banks increases in ω because aggregate demand increases.

The top left panel shows that ω increases after good shocks, but crashes after the first bad shock. This is because financiers exploit their leverage advantage to grow after good shocks, and 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 risk-free rate rises because both financiers and banks accumulate wealth after good shocks, increasing the supply of bonds, and the asset price rises as financiers become relatively wealthy. The bottom left panel plots total investment and investment by banks who shirk. The fraction of shirking banks grows steadily as ω increases. The bottom right panel plots output, with the dashed line showing counterfactual output given the same investment but no shirking. Hence shirking leads to excess downside risk. Since the share of shirking banks is increasing in the duration of the boom, and long sequences of good shocks are less likely than short ones, large recessions occur with lower probability than small recessions.

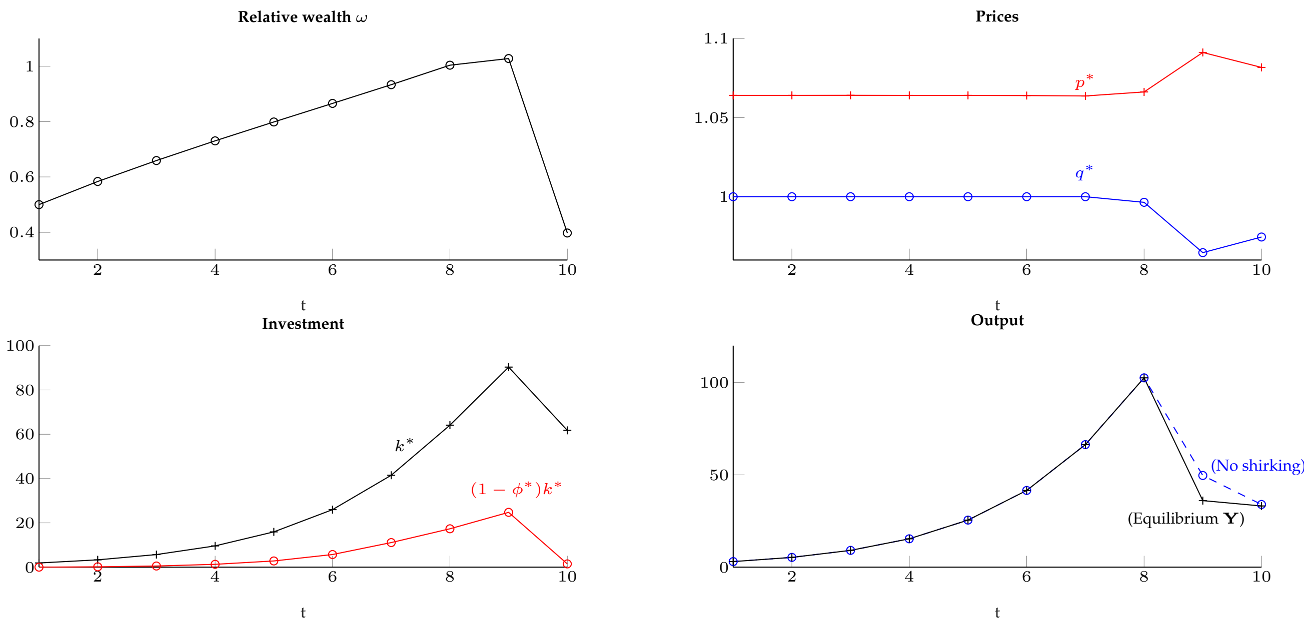

Figure 4: Equilibrium outcomes in the dynamic model given a sequence of eight good aggregate shocks and one negative shock. Parameters: $\pi_h = 0.65$, $Y_h = 1.6$, $y_h = 1.6$, $Y_l = 0.55$, $y_l = 0$, $m = 0.05$, $\beta = 0.95$, $\psi = 0.8$. *Saver wealth:* $w_S = 40$. *Initial wealth distribution:* $w_{B,0} = 0.8$, $w_{F,0} = 0.4$.

Figure 5 shows the importance of saver wealth in determining aggregate outcomes, and the model's rich state-contingent dynamics more generally. Parameters are the same

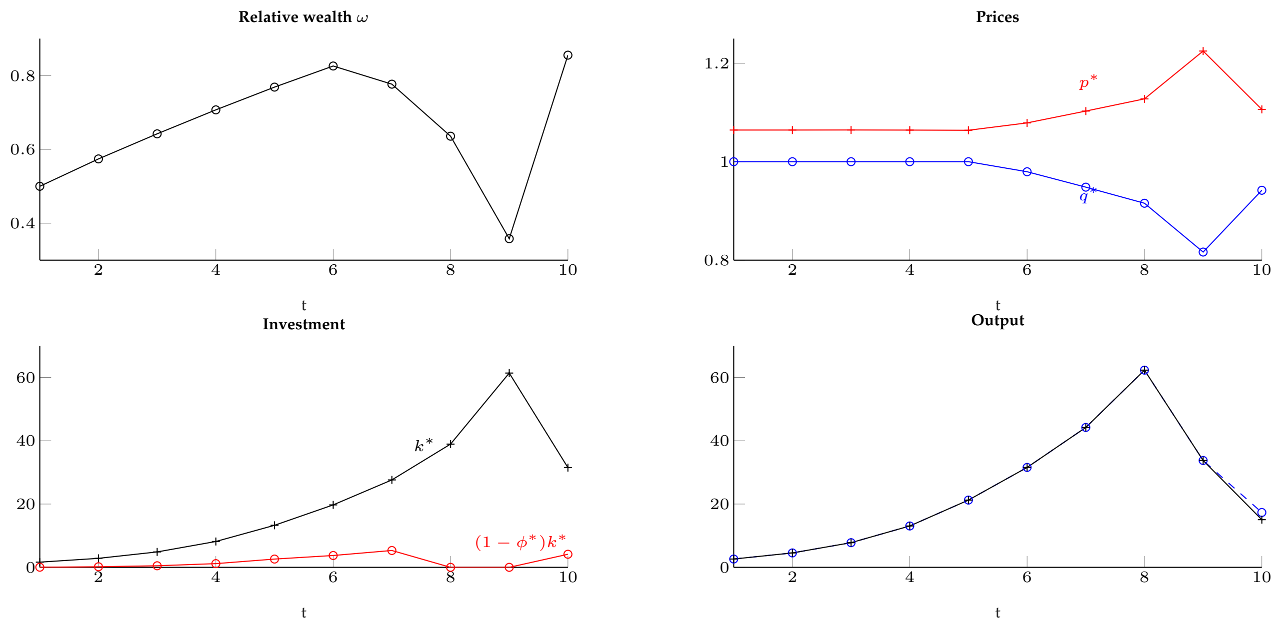

Figure 5: Equilibrium outcomes in the dynamic model given a sequence of eight good aggregate shocks and two negative shocks. Parameters and initial wealth distribution as in Figure 4. Saver wealth: $w_S = 10$.

as in Figure 4, except that saver wealth is 10 instead of 40. As a result, q falls more rapidly and ω grows more slowly. The asset price rises only because banks are not willing to sell below $\underline{P}(q)$, which is decreasing in q. (That is, the shadow cost of retaining assets falls as q falls, and so banks are less willing to sell assets at low prices.) However, asset market quantities are low because financiers are relatively poor, and this is reflected in the low share of shirking banks. Hence the credit cycle is less severe – there is less investment and fewer bad assets – when saver wealth is low. Interestingly, the fact that financiers do not always grow relative to banks means that banks maintain a higher share of aggregate risk during periods 8 and 9. This implies that ω now spikes after the bad shock, potentially creating the conditions that would trigger a more severe credit cycle going forward. Indeed, the destruction of bank capital also serves to lower the risk-free rate, which would allow financiers to again exploit their leverage advantage.

4 Relationship to the Empirical Literature

This section argues that the model delivers relationships between observables that are consistent with the empirical record on credit booms and financial crises. While these are not direct tests of the model, the fact that the model is consistent with a range of documented regularities may lend some credibility to the proposed mechanisms.

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 counterpart to be equity holders of financial institutions who originate loans to the real economy, such as commercial banks or mortgage companies. 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. I define savers to be creditors to financial institutions, such as depositors, corporations with excess cash, and sovereign wealth funds.[10] In line with the literature (e.g. Mendoza and Terrones (2012) and Schularick and Taylor (2012)), I define credit booms to be periods of increasing gross credit volumes. Since there is monotone mapping from gross credit volumes $b_B + b_F$ to investment, I use k as the model's measure of gross credit. The model generates the following predictions for observable outcomes.

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

For the pre-2008 U.S. credit boom., Adrian and Shin (2010) estimate that the combined balance sheet size of non-bank intermediaries such as hedge funds and broker-dealers

[10] Egan, Hortacsu, and Matvos (2017) show that half of all deposits in large U.S. 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 solvency constraints on banks.

was smaller than that of bank holding companies before 1990, but twice as large by 2007.

Prediction 2. *At the height of the boom, financiers are disproportionately exposed to downside risk. Financier balance sheets contract more sharply than bank balance sheets during crises.*

Greenlaw, Hatzius, Kashyap, and Shin (2008) study the 2008 financial crisis and show that non-bank intermediaries who purchased mortgage-backed securities were more exposed to downside risk than loan-originating banks, and suffered greater losses.

Prediction 3. *Credit booms coincide with increasing secondary market prices and volumes.*

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 1990, they issued more than $3.5 trillion in 2006. Chernenko, Hanson, and Sunderam (2016) 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.

Prediction 4. *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.*

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 (which can be interpreted as moral hazard on the path of play) 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 5. *Credit booms with falling asset quality are more likely to occur after an inflow of savings or during periods of expansionary monetary policy.*

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.*[11]

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.

[11] The simulations in Section 3.3 show that q declines during booms. Hence intermediation rents (and thus bank excess returns) decline during credit expansions.

5 Discussion of Assumptions

Two assumptions are critical to the model's results. 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, issuers of asset-backed securities are free to trade *non-exclusively* with multiple counterparties both at a given date and over time. Hence it is reasonable to assume that issuers cannot commit to not selling assets in the future, and that traders cannot contract on the trading partners of their counterparties or the total quantity of assets sold by a given issuer. Attar, Mariotti, and Salanié (2011) theoretically analyze non-exclusive markets with adverse selection, and show that it may indeed be impossible to screen by quantity if sellers trade with multiple buyers. Third, the balance sheets of financial institutions are hard to observe in real time. Hence financiers cannot properly assess a given banks exposure to a particular asset pool. Fourth, banks have access to a wide variety of financial securities with which to reduce exposure various risks. Ashcraft, Gooriah, and Kermani (forthcoming) document empirically that "complex financial innovations like CDOs enabled informed parties in the commercial mortgage backed securitization pipeline to reduce their skin-in-the-game in a way not observable to other market participants."

The second is that the contract space is such that financiers cannot claw back losses from banks in the event that they were sold low-quality assets. This can be motivated by the fact that it has proven difficult to hold financial institutions legally liable for documented instances of fraud in mortgage origination, or by the following example in which loan pools consist of individual assets whose payoffs have identical support and effort is thus not ex-post observable. Since the key motivation for asset sales is precisely to transfer risk exposure, financiers will therefore necessarily be exposed to the risk of moral hazard.

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 in the baseline model.*

Individual assets can be interpreted as claims on the real sector that either pay a fixed return or default, such as 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 assumption is convenient because savers lend only against risk-free debt. The downside is that shirking is observable ex-post at the level of bank balance sheets. The form of market incompleteness assumed here is thus equivalent to the non-contractability of bank portfolio returns, which is reasonable because it is difficult to observe or contract upon bank balance sheets in practice. A straightforward way to further relax 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. Effort is then not observable from portfolio returns ex-post, and yet there is a motive for selling assets to relax borrowing constraints.

6 Conclusion

This paper offers a theory of credit cycles in which the distribution of wealth and aggregate risk across financial intermediaries determines asset prices, credit volumes and investment efficiency. Some risk transfer from banks to financial market investors boosts credit volumes by relaxing borrowing constraints, but investment efficiency declines when banks sell too many assets. The latter channel dominates when asset prices are sufficiently high, providing a novel link from run-ups in asset prices to credit quality. Because buyers of risky assets grow wealthy after good shocks, inefficient asset price booms coincide with macroeconomic expansions. The models predictions are in line with empirical evidence on credit booms and the role of securitization in prominent financial crises. Low risk-free rates are a trigger of inefficient asset price booms, providing a novel dynamic link from expansionary monetary policy and saving gluts to investment inefficiency. This leads to novel implications for regulation and monetary policy. There are two main avenues for future research. The first is to study the optimal design of policy when there is a feedback from asset prices to credit quality. The second is a quantitative evaluation of the model.

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.

ASHCRAFT, A. B., K. GOORIAH, AND A. KERMANI (forthcoming): "Does Skin-in-the-Game Affect Security Performance?," *Journal of Financial Economics*.

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

AXELSON, U., AND P. BOND (2015): "Wall Street Occupations," *Journal of Finance*, 70(5), 1949–1996.

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.

BIANCHI, J., AND E. G. MENDOZA (2012): "Overborrowing, Financial Crises and 'Macro-prudential' Policy," *University of Pennsylvania Working Paper*.

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 E. FARHI (2014): "The Safety Trap," *NBER Working Paper 19927*.

CABALLERO, R. J., AND A. KRISHNAMURTHY (2009): "Global Imbalances and Financial Fragility," *American Economic Review, Papers and Proceedings*, 99(2), 584–88.

CARAMP, N. (2018): "Sowing the Seeds of Financial Crises: Endogenous Asset Creation and Adverse Selection," *University of California Davis Working Paper*.

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 (2016): "Who Neglects Risk? Investor Experience and the Credit Boom," *Journal of Financial Economics*, 122(2), 248–269.

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.

FOSTEL, A., AND J. GEANAKOPLOS (2008): "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98(4), 1211–1244.

——— (2015): "Leverage and Default in Binomial Economies: A Complete Characterization," *Econometrica*, 83(6), 2191–2229.

FUKUI, M. (2018): "Asset Quality Cycles," *Journal of Monetary Cycles*, 95, 97–108.

GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2013): "A Model of Shadow Banking," *The Journal of Finance*, 68(4), 1331–1363.

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

GERTLER, M., AND N. KIYOTAKI (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis," in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3, pp. 547–599. Elsevier.

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.

GUERRIERI, V., R. SHIMER, AND R. WRIGHT (2010): "Adverse Selection in Competitive Search Equilibrium," *Econometrica*, 78(6), 1823–1862.

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.

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," *Journal 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 (2017): "The Macroeconomics of Shadow Banking," *The Journal of Finance*, 72(6), 2381–2432.

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.

POVEL, P., R. SINGH, AND A. WINTON (2007): "Booms, Busts, and Fraud," *Review of Financial Studies*, 20(4), 1219–1254.

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.

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. The bank's incentive-compatibility constraint is

$$\sum_z \pi_z \left[\max\{Y_z\left(k-a_B\right)-b_B+P(\chi)a_B,0\}\right] \geq \sum_z \pi_z \left[\max\left\{y_z\left(k-a_B\right)-b_B+P(\chi)a_B,0\right\}\right]+mk. \tag{16}$$

Assume first that the solvency constraint does not bind in any state. Then (16) can be written as

$$\hat{Y}\left(k-a_B\right) \geq \hat{y}\left(k-a_B\right)+mk. \tag{17}$$

This can be restated as

$$\frac{a_B}{k} \leq \frac{\hat{Y}-\hat{y}-m}{\hat{Y}-\hat{y}}$$

Since $\chi = \frac{a}{k}$ and $\underline{a} \leq a_B$ and the solvency constraint must be respected on the equilibrium path, the first result follows. Next consider the second statement. Since χ is fixed, the bank has promised to sell at least $\underline{a}$ assets and has invested k. Suppose first that the bank does indeed sell $\underline{a}$ in assets, $a_b = \underline{a}$, and exerts effort. Since the solvency constraint must hold on the path of play, bank utility is $\hat{Y}(k - \underline{a}) - b_B + P(\chi)\underline{a}$. Now recall that the bank has the option secretly more than $\underline{a}$ assets. Since $P(\chi) \leq \hat{Y}$ without loss of generality, a deviation to some $a_B > \underline{a}$ cannot be profitable conditional on effort. Hence let the bank shirk. Since selling more assets does not violate the solvency constraint, the objective function is linear and, if a profitable deviation to some $a_B > \underline{a}$ exists, then the optimal deviation is $a_B = k$. Bank utility given the deviation is $P(\chi)k - b_B + mk$. The deviation is profitable if $P(\chi) \geq \bar{P}(\chi)$. □

A.2 Proposition 2

Suppose all banks in submarket χ exert effort. Then the expected gross return of buying assets in χ is $\frac{\hat{Y}}{P(\chi)}$. Since $P(\chi) \leq \bar{P}(\chi)$ by Proposition 1, $\frac{\hat{Y}}{P(\chi)} > 1$ so that the safe technology is dominated. Since all risk aggregate, it is without loss to assume that individual financiers invest in at most submarket. (This does not preclude that multiple submarkets are active.) By solvency constraint (6), $a_F \leq \frac{w_F}{P(\chi) - qY_l}$, and with equality if $\frac{\hat{Y}}{P(\chi^*)} > 1$. Hence for each χ, there exists a $\bar{\omega}$ such that $P(\chi) = \bar{P}$ if $\omega \geq \bar{\omega}$, and the market cannot clear at $\bar{P}(\chi)$ if all banks exert effort. Conversely, there does not exist an asset market equilibrium with $P > \bar{P}$, since all banks would shirk but $\bar{P}(\bar{\chi}) = \hat{y}$. Hence the unique equilibrium must be such that $P(\chi^*) = \bar{P}(\chi^*)$ with some banks shirking. In this case, aggregate supply is $\phi(\chi^*)\chi k + (1 - \phi(\chi^*))k$ and is strictly decreasing in $\phi(\chi^*)$. The fraction of good assets trading in the market is $f(\chi^*) = \frac{\phi(\chi^*)\chi^* k}{\phi(\chi^*)\chi k + (1 - \phi(\chi^*))k}$. Financiers invest only in risky assets if $\hat{x}(\chi^*) = f(\chi^*)\hat{Y} + (1 - f(\chi^*))\hat{y} > \bar{P}(\chi^*)$, and may invest in the safe technology if the condition holds with equality.

A.3 Proposition 3

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 2 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[12], 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 $\chi \in [0, \bar{\chi}]$. Given

[12] 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 χ^e. In this case, $u_F(a_F)$ represents average utility and the subsequent welfare analysis remains valid.

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. □

A.4 Proposition 4

Proof. Since each financier has zero mass, asset quality in each submarket is invariant to individual portfolios. By financiers' solvency constraint, $w_F'(z) = \int x_z(\chi)a_F(\chi) + s_F - b_F$. Since there is only aggregate risk, it is without loss to assume that each financier optimally invests in at most one submarket, say χ^*. Hence $Var(w_F'(z)) = Var\left(x_z(\chi^*)\right) a_F^2(\chi^*)$, which is strictly increasing in $a_F(\chi^*)$. If the solvency constraint binds, $a_F(\chi^*) = \frac{w_F}{P(\chi^*) - qx_l(\chi)}$ is strictly increasing in q. By banks' solvency constraint, $w_B'(z) = \hat{\mathbf{y}}_z(e)\left(k - a_B\right) - b_B + P(\chi)a_B$. Hence $Var(w_B'(z)) = Var(\hat{\mathbf{y}}_z(e))\left(k - a_B\right)^2$. □

A.5 Proposition 5

Proof. Begin by deriving the borrowing constraint. Since $a_b = \underline{a}$ and $\chi \leq \bar{\chi}$, the double deviation from Proposition 1 is ruled out. Suppose first that banks are expected to shirk. Solvency constraint (3) implies that $y_z\left(k - \underline{a}\right) - b_B + P(\chi)\underline{a} \geq 0$ for all z. Hence banks prefer to shirk if

$$\hat{y}\left(k - \underline{a}\right) + mk > \hat{Y}\left(k - \underline{a}\right) \quad \Leftrightarrow \quad \chi > \bar{\chi}$$

a contradiction with $\chi \leq \bar{\chi}$. Next suppose that banks exert effort. Since $y_l < Y_l$, banks can borrow more if they violate the solvency constraint *conditional on shirking* (i.e. off the equilibrium path) than if they respect it for all states and all actions. The incentive-compatibility constraint is

$$\hat{Y}\left(k - a_B\right) - b_B + pa_B \geq \pi_h\left[y_z\left(k - a_B\right) - b_B + P(\chi)a_B\right] + mk. \tag{18}$$

which is equivalent to $b_B \leq (\Delta - m/\pi_l)k + (P(\chi) - \Delta)\underline{a}$. What remains to be shown is that this is a tighter constraint than solvency constraint (3). Given $e = 1$, the solvency constraint is $b_B \leq Y_l k + (P(\chi) - Y_l)\underline{a}$. Since $\Delta \leq Y_l$ because $y_h \geq Y_h$, the result follows.

Next consider the lower bound on the asset price. Let $\tilde{\Delta} = \Delta - m/\pi_l$. If the borrowing constraint binds, investment and debt satisfy $k(\chi) = w_B/\left(1 - q\tilde{\Delta} - q\chi(p - \Delta)\right)$ and $b_B(\chi) = (\tilde{\Delta} + \chi(p - \Delta))k(\chi)$. Hence expected bank utility is

$$u_B(\chi) = \frac{\left(\hat{Y} - \tilde{\Delta} - (\hat{Y} - \Delta)\chi\right) w_B}{1 - q\tilde{\Delta} - q\chi(P(\chi) - \Delta)}$$

Differentiating with respect to χ, evaluating at $a_b = 0$ and rearranging yields that $\frac{\partial u_B}{\partial \chi} \geq 0$ only if $P(0) \geq \underline{P}(q)$. $\underline{P}(q)$ is increasing because $\Pi(q)$ is strictly decreasing. □

A.6 Proposition 6

Proof. Uniqueness of symmetric equilibrium is immediate because agents are symmetric and all constraints are convex. Existence follows from Assumption 2.

Optimal Financier Portfolio. It is without loss to assume that each financier invests in at most one submarket, say χ. Financiers strictly prefer risky assets to the safe technology if $\hat{x}(\chi)/p > 1$, and invest in the safe technology only if $\hat{x}(\chi)/p = 1$. To characterize optimal debt issuance, rewrite the solvency constraint as $b_F = \gamma\,[x_l(\chi)a_F + s_F]$, where $\gamma \in [0,1]$. Then optimal borrowing satisfies $\gamma^* = 1$ if $\hat{x}(\chi)/p > 1/q$, $\gamma^* = 0$ if $\hat{x}(\chi)/p < 1/q$, and $\gamma \in [0,1]$ otherwise. In a symmetric equilibrium, financiers choose the same γ when indifferent. Since $p \leq \bar{P}(\chi)$ and $\bar{P}(\chi) \in [\hat{y}, \hat{Y})$ for all $\chi \leq \bar{\chi}$, we have that $\hat{x}(\chi)/p > 1$ only if$\phi(\chi) \geq \underline{\phi}(\chi) \equiv \frac{p-\hat{y}}{p-\hat{y}+\chi(\hat{Y}-p)}$.

Optimal Bank Portfolio. Assumption 2 implies that $q > \frac{1}{\hat{Y}}$, hence the borrowing constraint binds. We have already ruled out equilibria with trade in submarkets $\chi > \bar{\chi}$. By Proposition 5, the borrowing constraint implies incentive-compatibility *conditional on* $a_b = \underline{a}$ for any k. Hence banks shirk on the path of play if and only if $a_B = k > \underline{a}$. By Proposition 2, there does not exist an equilibrium with $p > \bar{P}(\chi)$. Hence $p = \bar{P}(\chi)$ in at least one submarket in any equilibrium with shirking, and, in any such submarket, banks are indifferent between (i) $a_b = \underline{a}$ and $e = 1$, and (ii) $a_b = k$ and $e = 0$ Since the ex-post deviation is unobservable, no bank can borrow more than what is permitted by (11). It follows that banks do not differ in debt b_B, investment k, and asset-sale promise $\underline{a}$. By the proof of Proposition 5, $k(\chi) = \frac{w_B}{1-q\tilde{\Delta}-q\chi(p-\Delta)}$, $b_B(\chi) = (\tilde{\Delta} + \chi(p-\Delta))k(\chi)$, and $\frac{\partial u_B}{\partial \chi} > 0$ if and only if $p > \underline{P}(q)$, and $\frac{\partial u_B}{\partial \chi} = 0$ if $p = \underline{P}(q)$. Hence the optimal share of assets to be sold is $\chi^* = \chi^e$ if $p > \underline{P}(q)$, and $\chi^* \in [0, \chi^e]$ if $p = \underline{P}(q)$, and there exists a unique active submarket in any symmetric equilibrium. Banks who exert effort set $a_B = \underline{a}$, banks who shirk set $a_B = k(\chi^*)$. Hence aggregate asset market supply is $\phi(\chi^*)\chi^* k(\chi^*) + (1-\phi(\chi^*))k(\chi^*)$, and is decreasing in $\phi(\chi^*)$.

Constructing the cutoffs. Assume that all banks exert effort. Given χ, the market-clearing condition $\chi k(\chi) = \frac{w_F}{p-\gamma q Y_l}$ yields market-clearing price

$$\hat{p}(\omega|\chi,\gamma) \equiv \frac{\omega\,(1 - q(\Delta - m/\pi_l) + q\Delta\chi) + \gamma\chi q Y_l}{\chi(1+q\omega)}, \tag{19}$$

which is strictly increasing in ω if $q\hat{Y} \neq p$ and locally independent of ω otherwise. Define the *lowest* and *highest* individually optimal financier leverage given p and q to be $\bar{\gamma} = \mathbb{1}(q\hat{Y} > p)$ and $\underline{\gamma} = 1 - \mathbb{1}(q\hat{Y} < p)$. Note that $\bar{\gamma}(p) \neq \underline{\gamma}(p)$ if and only if $q\hat{Y} = p$. There does not exist an equilibrium with trade in good assets if $\underline{P}(q) > \bar{P}(\chi)$ or $\chi > \bar{\chi}$. Hence any equilibrium without shirking must have $\chi^* \leq \chi^e$, and cutoffs satisfy

$$\hat{p}(\underline{\omega}|\chi^e, \bar{\gamma}) = \underline{P}(q) \quad \text{and} \quad \hat{p}(\bar{\omega}|\chi^e, \underline{\gamma}) = \bar{P}(\chi^e) \tag{20}$$

where $\underline{\omega} \geq \bar{\omega}$ and $\underline{\omega} = \bar{\omega}$ if and only if $\hat{\chi} \leq \bar{\chi}$. The remaining characterization is as follows.

(i) The construction of $\underline{\omega}$ implies that $p^* = \underline{P}(q)$ and $\phi^* = 1$ for all $\omega < \underline{\omega}$. Since $\frac{\partial u_B}{\partial \chi} = 0$ if $p = \underline{P}(q)$, asset market quantities are determined by financier demand: $\chi^* = \frac{a_F^*}{k(\chi)}$, $k^* = \frac{w_B + q(p-\Delta)a_F}{1-q(\Delta - m/\pi_l)}$, and $a_F^* = \frac{w_F}{p - \gamma^* q Y_l}$. So $\frac{\partial k^*}{\partial w_B} > 0$. If $q\hat{Y} \neq p$, then $\gamma^* \in \{0,1\}$ and $\frac{\partial k^*}{\partial w_F} > 0$. If $q\hat{Y} = p$, then $\gamma \in [0,1]$ and k^* is locally independent of w_F. Since all banks exert effort, expected output is increasing if k is. Since $\hat{Y} > \underline{P}(q)$ for all $\chi \leq \chi^e$, financiers strictly prefer to borrow if q is sufficiently close to one. If $\gamma^* = 1$, then $a_F^* = \frac{w_F}{p - qY_l}$ and $\frac{\partial k^*(\chi^*)}{\partial w_F} > \frac{\partial k^*(\chi^*)}{\partial w_B}$ if and only if $Y_l - \Delta > \frac{(1-q)\underline{P}(q)}{q}$. If $y_h > Y_h$, then $\Delta > Y_l$, and so condition is satisfied for q sufficiently close to one. Hence changing the wealth distribution to $(w_F + \epsilon, w_B - \epsilon)$ from (w_F, w_B) strictly increases investment and output.

(ii) Now let $\omega \in (\underline{\omega}, \bar{\omega}]$. This interval is measure zero if $\bar{\chi} \leq \hat{\chi}$. Hence let $\bar{\chi} > \hat{\chi}$. By the arguments above, then $p^* = \hat{p}(\omega|\bar{\chi})$, 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. Since all agents exert effort, expected output increases if k increases.

(iii) By construction of the cutoff, $\chi = \chi^e$ and $p^* = \bar{P}(\chi^e)$ for all $\omega > \bar{\omega}$. Hence k^* is independent of w_F. Financiers can leverage less if the fraction of good assets is low. Asset market-clearing can be written as $\phi^*(\chi^e)\chi^e k(\chi^e) + (1 - \phi(\chi^e))k(\chi^e) = \frac{w_F}{p^* - q\gamma^*(y_l + f(\chi^e)(Y_l - y_l))}$, where $f(\chi^e) = \frac{\phi(\chi^e)\chi^e}{\phi(\chi^e) + (1-\phi)}$ is the fraction of good *assets*. Then

$$\phi^* = \frac{(\bar{P}(\chi^e) - q\gamma^* y_l)k(\chi^e) - w_F}{q\gamma^*(Y_l - y_l)\chi^e k(\chi^e) + (\bar{P}(\chi^e) - q\gamma^* y_l)(1-\chi^e)k^*(\chi)}, \tag{21}$$

where k^* is linear in w_B so that ϕ^* is a function of ω. If $q\left(f^*\hat{Y} + (1-f^*)\hat{y}\right) \neq \bar{P}(\chi^e)$, then $\gamma^* \in \{0,1\}$, and ϕ^* is strictly decreasing in ω, and $\mathbb{E}\mathbf{Y}(z)$ is strictly decreasing in w_F because k^* is constant in w_F. If $q\left(f^*\hat{Y} + (1-f^*)\hat{y}\right) = \bar{P}(\chi^e)$, then f^* is a constant because changes ω are absorbed by changes in financier leverage holding ϕ^* fixed. Hence ϕ^*, k^* and $\mathbb{E}\mathbf{Y}(z)$ are constant in w_F. Since leverage is bounded below by 0, such adjustments are feasible for a finite interval of financier wealth. For all ω, ϕ^* is bounded below by $\underline{\phi}$ (as defined above in the optimal financier portfolio) since financiers cease to buy otherwise. Output variance increasing in the share of bad assets because their payoffs have a higher variance.

(iv) Follows from the construction of the cutoffs above, since the upper and lower bound on the asset price coincide at χ^e under the stated conditions.

□

A.7 Corollary 1

Proof. If $\chi^e = \bar{\chi}$, then $\bar{P}(\chi^e) = \hat{y}$ and $\bar{\omega} = \frac{\bar{\chi}(\hat{y}-\bar{\gamma}qY_l)}{1-q(\Delta-m/\pi_l)-q\bar{\chi}(\hat{y}-\Delta)}$, where $\bar{\gamma}$ is the highest degree of leverage consistent with private optimality (as defined in the proof of Proposition 6). It is sufficient to show that $\frac{\partial\bar{\omega}}{\partial Y_h} > 0$ given $\bar{\gamma} = 1$. Note that $\bar{\omega} = \frac{B}{1-qA}$, where $A = \frac{\hat{y}(\hat{Y}-\hat{y})-m(\hat{y}-y_l)}{\hat{Y}-\hat{y}}$, and $B = \frac{(\hat{Y}-\hat{y}-m)(\hat{y}-qY_l)}{\hat{Y}-\hat{y}}$. The result follows from $\frac{\partial A}{\partial Y_h} > 0$, $\frac{\partial B}{\partial Y_h} > 0$. □

A.8 Proposition 7

Proof. Let $k(\lambda) = \lambda w_B$, and observe that banks must sell assets to achieve this level of investment if $k(\lambda) > \frac{w_B}{1-q(\Delta-m/\pi_l)}$. Define $\chi(\lambda)$ to be the fraction of assets that must be sold to invest $k(\lambda)$. Then $\chi(\lambda) = (\lambda(1-q(\Delta-m/\pi_l))-1)\,/\,(\lambda(p-\Delta)q)$, and $\lim_{\lambda\to(1-q(\Delta-m/\pi_l))^{-1}}\chi(\lambda) = 0$. Now suppose for a contradiction that no bank shirks. Market-clearing requires $\frac{w_F}{p-qY_l} = \chi(\lambda)\lambda w_B$. The equilibrium price is $p(\lambda) = \frac{w_F}{\chi(\lambda)\lambda w_B} + qY_l$, and $\lim_{\lambda\to(1-q(\Delta-m/\pi_l))^{-1}} p(\lambda) = \infty > \bar{P}(\chi)$ for all $\chi \leq \chi^e$. Hence there cannot exist an equilibrium without shirking for λ sufficiently close to close to $(1-q(\Delta-m/\pi_l))^{-1}$. □

A.9 Proposition 8

Proof. Since the objective function of each generation and all payoffs are as in the static model, optimal portfolios and effort decisions are as in the static model. Proofs of the individual statements are as follows:

(i) Since $\bar{P}(\chi^e) < \hat{Y}$, it is without loss to assume that financiers' solvency constraint binds for q sufficiently large. If $\omega < \underline{\omega}$, the proof of Proposition 6 shows that $k^* = \frac{w_B+q(\underline{P}(q)-\Delta)a_F}{1-q(\Delta-m/\pi_l)}$ and $p = \underline{P}(q)$. By the definition of $\underline{P}(q)$, it follows that $k^* = \frac{w_B}{1-q(\Delta-m/\pi_l)} + \left(\frac{\hat{Y}-\Delta}{\hat{Y}-(\Delta-m/\pi_l)}\right)a_F$. Since the bank borrowing constraint binds in equilibrium, $w'_B(z) = \nu_B(z)w_B - (Y_z-\hat{Y})\rho(1-\tilde{m})a_F$, where $\nu_B(z) = \frac{Y_z-(\Delta-m/\pi_l)}{1-q(\Delta-m/\pi_l)}$. Since the financier solvency constraint binds, $a_F = \frac{w_F}{p-qY_l}$ and $w'_F(z) = \nu_F(z)w_F$, where $\nu_F(z) = \frac{Y_z-Y_l}{\underline{P}(q)-qY_l}$. Given $p = \underline{P}(q)$, $\frac{\partial u_B}{\partial\chi} = 0$, and $u_B = \mathbb{E}_z\nu_B(z)w_B$ and $u_F = \mathbb{E}_z\nu_F(z)w_F$. It is then easy to $E_z\nu_F(z) \geq \mathbb{E}_z\nu_B(z)$ if and only if $\underline{P}(q) \leq \underline{P}(1)$, and $E_z\nu_F(z) = \mathbb{E}_z\nu_B(z)$ if $q = 1$. This implies that financiers can take on more leverage if $q = 1$, i.e. $q = 1 \Rightarrow \frac{1}{\underline{P}(q)-qY_l} > \frac{1}{1-q(\Delta-m/\pi_l)}$. Since $\omega'(z) = \frac{\nu_F(z)\omega}{\nu_B(z)-(Y_z-\hat{Y})\left(\frac{m/\pi_l}{\underline{P}(q)-qY_l}\right)\omega}$, it follows that $\omega'(h)$ is strictly increasing in ω and $\omega'(h) > \omega$ if $q = 1$. The existence of $q^C < 1$ follows from the continuity of $\nu_F(h) - \nu_B(h)$.

(ii) The definition of leverage ratios λ_F and λ_B are given in the proofs of Propositions 5 and (6), suitably modified for the case where all banks exert effort, $\chi = \bar{\chi}$, and $p = \bar{P}(\bar{\chi})$. $\underline{a} \in (\underline{\omega}, \bar{\omega}]$ is non-empty if and only if $\bar{\chi} > \hat{\chi}(q)$. Hence let $\chi^e = \bar{\chi}$. Then $w'_B(z) =$

$Y_z(k - \underline{a}) - b_b + P(\bar{\chi})\underline{a}$, where $\underline{a} = \bar{\chi}k$ and $b_B = [(\Delta - m/\pi_l) + (P(\bar{\chi}) - \Delta)\bar{\chi}]\, k = [(1-\bar{\chi})y_l + \bar{\chi}P(\bar{\chi})]\, k$. Hence $w_B'(z) = (Y_z - y_l)(1-\bar{\chi})k$. Given $\omega < \underline{\omega}$, financiers not invest in storage. By the optimal portfolio in the proof of Proposition 6, $w_F'(z) = (Y_z - \gamma Y_l)a_F = w_F'(z) = (Y_z - \gamma Y_l)\bar{\chi}k$, where the last equality follows from market clearing condition $a_F = \bar{\chi}k$. Hence $\omega'(z) = \frac{r_F(z)}{r_B(z)}\frac{\bar{\chi}}{1-\bar{\chi}}$. To construct $\bar{\omega}$, observe that $\bar{P}(\bar{\chi}) = \hat{y}$ is a constant, and so market clearing can be stated as $\lambda_F w_F = \bar{\chi}\lambda_B w_B$. Hence $\bar{\omega} = \bar{\chi}\frac{\lambda_B}{\lambda_F}$. Rearranging the inequality $\omega'(h) > \bar{\omega}$ then yields χ^0. Note that since $\omega \leq \bar{\omega}$, $\gamma = 1$ as $q \to 1$. It is also easy to verify that $\underline{P}(1) \geq 1$. Hence for $\bar{\omega} > \underline{\omega}$, we require that $\bar{P}(\bar{\chi}) > 1$. It follows that $\epsilon_B^0 > 1$ if $q = 1$, and the denominator is positive if the numerator is.

(iii) Observe that banks who shirk sell more assets than banks who exert effort. Hence the fraction of investment purchased by financiers is strictly greater if $\omega > \bar{\omega}$ than if $\omega \leq \bar{\omega}$. Moreover, γ^* is weakly decreasing in ω. The result then follows from $y_h \geq Y_h$.

□

A.10 Corollary 2

Proof. Follows from Propositions 6 and 8, and observing that χ^0 is continuous in q. □

A.11 Corollary 3

Proof. Under the stated conditions, $w_F'(z) = (\phi^*\bar{\chi}Y_z + (1-\phi^*)y_z)k^*$, where $y_l < Y_l$. By Proposition 6, $\frac{\partial k}{\partial w_F} = 0$ and $\frac{\partial \phi^*}{\partial w_F} < 0$ given $\omega > \underline{\omega}$. □

B Value Functions and Computational Algorithm

B.1 Value Functions

I first 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

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

Here $\Gamma(\cdot)$ 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 $\mathbf{w}$ by $\chi^*(\mathbf{w})$ and the fraction of bad banks by $\phi(\mathbf{w})$.

Definition 3. *Conditional on* $\mathbf{w}$*, banks value a good asset at* $\tilde{Y}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z))Y_z$ *and a bad asset at* $\tilde{y}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z))y_z$*. The value of the risk-free asset is* $\tilde{\mu}_B(\mathbf{w}) \equiv \mathbb{E}_z \mu_B(\mathbf{w}'(z))$*. The pledgeable return is* $\Delta(\mathbf{w}) = \frac{\tilde{Y}_B(\mathbf{w}) - \pi_h \mu_B(\mathbf{w}'(h))y_h}{\pi_l \mu_B(\mathbf{w}'(l))}$.

Proposition 9. *The borrowing constraint is*

$$b_B \leq (\Delta(\mathbf{w}) - m/(\pi_l \mu_B(\mathbf{w}'(l)))k + (P(\chi) - \Delta(\mathbf{w}))a_B,$$

and banks shirk at any asset price if

$$\chi > \bar{\chi}(\mathbf{w}) = \left(\frac{\Delta(\mathbf{w}) - m/(\pi_l \mu_B(\mathbf{w}'(l)) - y_l}{\Delta(\mathbf{w}) - y_l} \right)$$

The upper and lower bounds on the asset price are, respectively,

$$\bar{P}(\chi) = \frac{1}{\tilde{\mu}_B(\mathbf{w})} \left(\tilde{Y}_B(\mathbf{w}) - \frac{m}{1-\chi} \right) \quad \textit{and} \quad \underline{P}(q, \mathbf{w}) = \frac{\tilde{Y}_B(\mathbf{w}) + \Delta(\mathbf{w})\Pi(q, \mathbf{w})}{\tilde{\mu}_B(\mathbf{w}) + \Pi(q, \mathbf{w})}$$

where $\Pi(q, \mathbf{w}) = \frac{q\tilde{Y}_B(\mathbf{w}) - \tilde{\mu}_B(\mathbf{w})}{1 - q(\Delta(\mathbf{w}) - m/(\pi_l \mu_B(\mathbf{w}'(l)))}$.

Proof. Banks are risk-neutral conditional on the state-contingent marginal values defined in Definition 3. Hence the derivations from the static model apply. □

Now turn to financiers. Their value function is

$$\begin{aligned} V_F(w_F, \mathbf{w}) = & \max_{a_F \geq 0, b_F \geq 0, s_F \geq 0} \quad \mathbb{E}_z \beta \Big[(1-\psi)w'_F(z) + \psi V_F(w'_F(z), \mathbf{w}') \Big] \\ \text{where} \quad & s_F + pa_F = w_F + qb_F && \text{(Budget constraint)} \\ & b_F \leq x_l(\mathbf{w})a_F + s_F && \text{(Solvency constraint)} \\ & w'_F(z) = s_F + x_z(\mathbf{w})a_F - b_F && \text{(Individual law of motion)} \\ & \mathbf{w}'(z) = \Gamma(\mathbf{w}, z) && \text{(Aggregate law of motion)} \end{aligned}$$

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 4. *Given* $\mathbf{w}$*, financiers' expected value of a risky asset purchased on asset markets is* $\tilde{x}_F(\mathbf{w}) \equiv \mathbb{E}_z \mu_F(\mathbf{w}'(z))x_z$*, and the expected value of the risky asset is* $\tilde{\mu}_F(\mathbf{w}) \equiv \mathbb{E}_z \mu_F(\mathbf{w}'(z))$*.*

Proposition 10. *Financiers are willing to purchase risky assets if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} \geq p$*. Financiers borrows as much as possible if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} > \frac{p}{q}$ *and are indifferent if* $\frac{\tilde{x}_F(\mathbf{w})}{\tilde{\mu}_F(\mathbf{w})} = \frac{p}{q}$*.*

Proof. Financiers are risk-neutral conditional on the state-contingent marginal values defined in Definition 4. Hence the derivations from the static model apply. □

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.

An important difference with the static model that financiers with precautionary motives are less willing to buy low-quality assets at a given price than risk-neutral financiers. However, banks with precautionary motives are *more* likely to deviate than risk-neutral banks because double deviation to shirking and selling lowers banks' risk exposure. Hence there continues to be scope for shirking because the upper bound on the asset price falls. Since the precautionary motive decreases in own net worth, moreover, shirking is more likely when financiers are relatively wealthy. In addition, all agents take the evolution of aggregate net worth as given. Hence there is dynamic externality whereby individual banks and financiers do not internalize that their current portfolio choices may trigger shirking in the future.

B.2 Computational Algorithm and Equilibrium Value Functions

I now describe the computational algorithm. Given that we can focus without loss of generality on symmetric portfolios and z is i.i.d., the state variables are the aggregate wealth distribution and own net worth. Hence we can compute symmetric equilibrium using discrete grids for $\mathbf{w}_B$ and $\mathbf{w}_F$. Solving the model presents two challenges. The first is that there is 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). (An important difference to their algorithm is that, conditional on prices, the equilibrium law of motion for the wealth distribution can be computed in closed form here.)

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 coefficients $v_B^0(\cdot), v_F^0(\cdot)$ and an aggregate law of motion $\Gamma^0(\cdot)$.
3. For each point on the grid, compute the static equilibrium given $\{v_B^0(\cdot), v_F^0(\cdot), \Gamma^0(\cdot)\}$.
4. Use optimal bank and financier portfolios to update the law of motion to $\Gamma^0_{updated}(\cdot)$.
5. Iterate to convergence 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 6 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 $\mathbf{w}_F$. There are various kinks because optimal portfolios are highly non-linear: for example, there exist cut-off prices at which financiers stop issuing debt, or stop buying risk assets. The combination of these effects leads to non-linearities in the value functions. The left panel shows

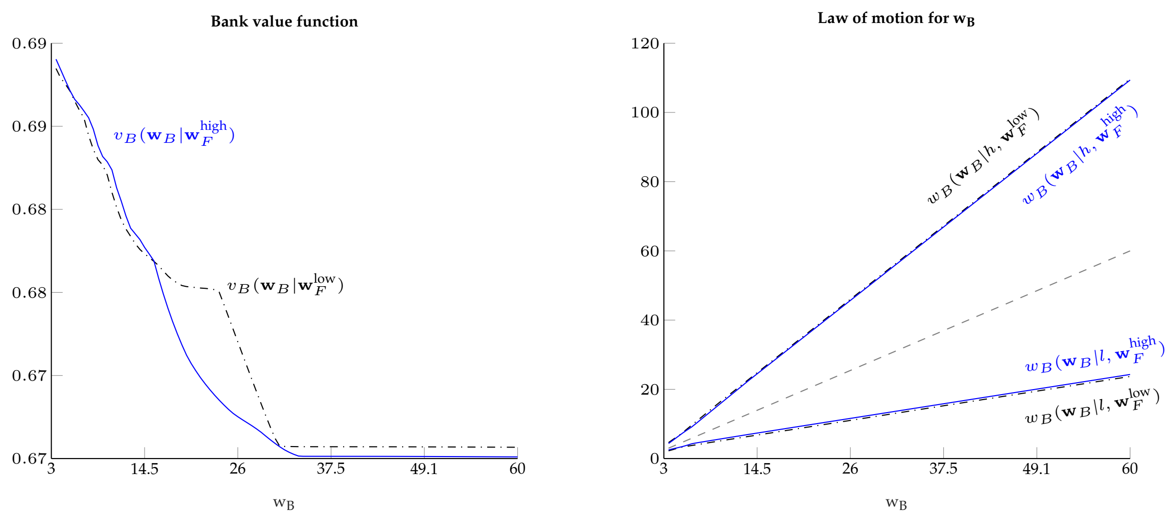

Figure 6: Bank value function coefficients and law of motion for bank wealth. $v_B(\cdot)$ and $w'_B(z)$ are plotted as a function of $\mathbf{w}_B$ conditional on two levels of financier wealth. Parameters as in Figure 4 below.

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 7 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 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

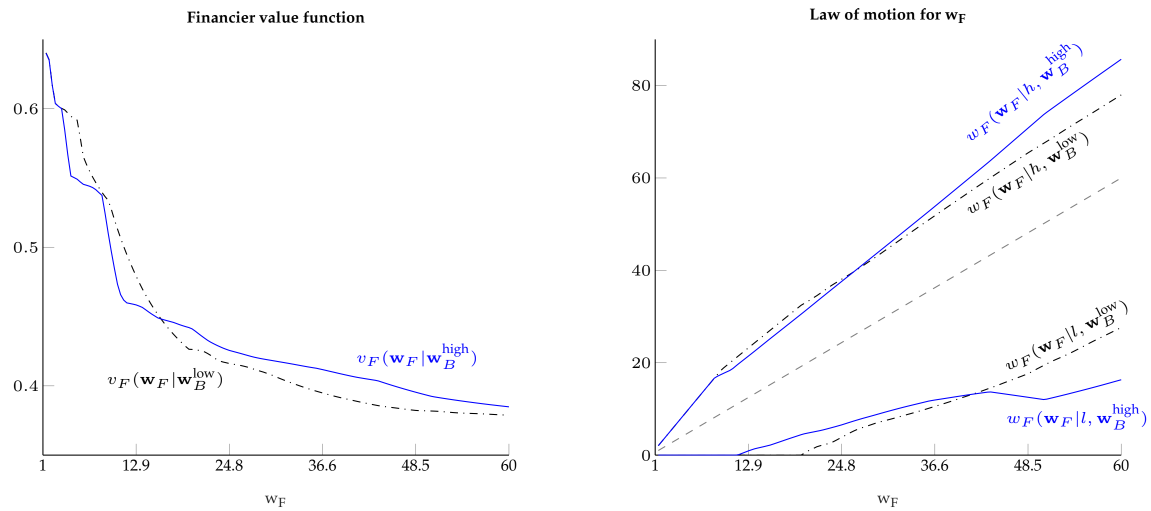

Figure 7: Financier value function coefficients and law of motion for financier wealth. $v_F(\cdot)$ and $w'_B(z)$ are plotted as a function of $\mathbf{w}_F$ conditional on two levels of bank wealth. Parameters as in Figure 4 below.

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 3). Eventually, the fraction of shirking banks reaches its maximum level and $\mathbf{w}'_F(l)$ again increases in $\mathbf{w}_F$.

Online Appendix. Not for Publication.

C 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 $\chi \in [0, \chi^e]$ 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.

C.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

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

By construction, banks who shirk obtain the same utility as banks who exert effort if $P(\chi) = \bar{P}(\chi)$. Conditional on χ, we can therefore write bank utility as

$$u_B(\chi) = \big(\bar{P}(\chi) + m\big)\, k(\chi) - b_B(\chi) = \left(\frac{\hat{Y} - (\Delta - m/\pi_l) - (\hat{Y} - \Delta)\chi}{1 - q(\Delta - m/\pi_l) - q\chi(\hat{Y} - \Delta) + \frac{qm\chi}{1-\chi}}\right) w_B$$

Individual optimality requires $\chi^* = \arg\max_{\chi \in [0,\chi^e]} u_B(\chi)$. (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 infra-marginal 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 de-lever.

Assumption 3 (Costly Signaling). $\frac{\partial k(\chi)}{\partial \chi} \geq 0 \; \forall \; \chi \in [0, \bar{\chi}]$.

Observation 2. *Assumption 3 holds if and only if* $\bar{\chi} \leq \frac{\hat{Y} - \Delta - m}{\hat{Y} - \hat{y} + \hat{Y} - \Delta}$.

Proof. $\frac{\partial k(\chi)}{\partial \chi} = k(\chi) \left(\frac{q}{1 - q(\Delta - m/\pi_l) - \chi q(\bar{P}(\chi) - \Delta)} \right) \left(\bar{P}(\chi) - \Delta + \chi \bar{P}'(\chi) \right)$ where $\bar{P}'(\chi) = -\frac{m}{(1-\chi)^2}$. $\bar{P}(\chi)$ and $\bar{P}'(\chi)$ are strictly decreasing in χ. Hence Assumption 3 holds iff $\frac{\partial k(\chi)}{\partial \chi} |(\chi = \bar{\chi}) \geq 0$. □

Assumption 3 does *not* imply that it is not optimal for the bank to signal: it may be privately beneficial borrow slightly less if this means the bank can sell all infra-marginal 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 3 holds.

Proposition 11. *Define* $MV(\chi) = \bar{P}(\chi) \cdot \chi \cdot k(\chi)$ *to be the market value of risky assets if no bank shirks. If Assumption* (3) *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 2, $\bar{P}(\chi) - \Delta + \chi \bar{P}'(\chi) \geq 0$. So $\left(\bar{P}(\chi) + \chi \bar{P}'(\chi) \right) k(\chi) > 0$. Since $k'(\chi) \geq 0$ it follows that $\frac{\partial MV(\chi)}{\partial \chi} > 0$. □

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.

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

Proposition 12. 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 individually 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. □

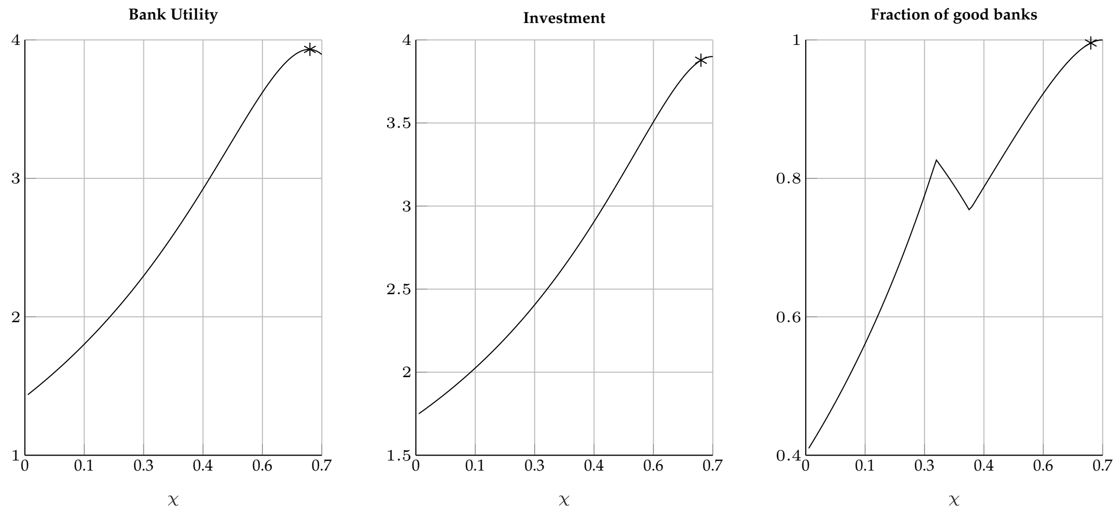

Figure 8: Optimal portfolio and equilibrium outcomes in the signaling equilibrium as a function of χ. The asterisk denotes the utility-maximizing choice χ^*. All parameters as in Figure 4.The non-monotonicity between 0.3 and 0.4 occurs because financiers are indifferent toward leverage in this region. Since $P(\chi)$ is decreasing in χ, reductions in χ lead to price increases. To keep financiers indifferent, ϕ must increase. (If financiers were not indifferent, then ϕ would be such that is strictly optimal to borrow if no one else does, and strictly optimal to not borrow if all financiers do)

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 features the same key economic mechanisms as the baseline model, but at the cost of additional complexity.

D Extension to multiple rounds of 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 (or placed in an escrow account), 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, and a "partial commitment" to a minimum level of asset sales arises endogenously.

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 $n = 0 \ldots N$ 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

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

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 borrowing constraint from the baseline model. Note that it is easy for $\tilde{P}$ to incorporate at least as much information as in the baseline model. To allow for double deviations to shirking and selling, 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.

E 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 asset 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

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

This problem is linear in L up to a binding solvency constraint. Hence banks will either choose to shirk on all assets or not at all. Moreover, whether or not the solvency constraint binds is determined by the same conditions as in the baseline model. If financiers use a random selection rule, there is thus no loss of generality in assuming that the bank either shirks on all assets or not at all.