Bank Size, Leverage, and Financial Downturns *

Chacko George †

Abstract

I construct a macroeconomic model with a heterogeneous banking sector. Banks differ in their ability to transform deposits from households into loans to firms. Bank size differences emerge endogenously in the model, and in steady state, the induced bank size distribution matches two facts: bigger banks borrow more on the interbank lending market than smaller banks, and bigger banks are more leveraged than smaller banks. I use the model to evaluate the impact of increasing concentration in US banking on the severity of potential downturns, and find that economies with higher concentration can experience deeper downturns and longer recoveries.

Keywords: Financial crisis, concentration

JEL Classifications: E02, E44, E61, G01, G21

Opinions expressed in this paper are those of the author and not necessarily those of the FDIC.

*Thanks to Matthias Kehrig, Olivier Coibion, Andrew Glover, Jacek Rothert, and David Kendrick for all their guidance, and to Rudiger Bachmann, Francois Gourio, Chris House, and Ariel Zetlin-Jones for all their helpful comments and suggestions. Thanks also to the seminar participants at the Federal Reserve Bank of Dallas and the University of Texas for their questions and suggestions.

†FDIC, Center for Financial Research, 550 17th St NW, Washington, DC 20429, email: cgeorge@fdic.gov.

1 Introduction

The Great Recession of 2007-2009 began in the housing sector but ultimately resulted in a decrease in employment and output in many sectors of the economy. Many accounts of this episode feature the financial sector as playing a central role in transmitting weakness from one part of the economy to another. According to the narrative, the initial weakness in the housing sector decreased the value of mortgages and their derivatives, decreasing the value of the asset side of their balance sheets. Financial institutions then engaged in a deleveraging process which reduced their investments in other sectors of the economy, driving an economy-wide recession.

If the financial sector played the transmission role that the narrative suggests, then the industrial characteristics of the banking sector - the number of banks, their size, the functions they engage in, and the way they interact - should matter a great deal for macroeconomic volatility. This paper explores the relationship between one such characteristic, concentration in terms of asset size, and the depth of recessions. At the most basic level, I ask: can a change in the distribution of bank sizes affect the depth of recessions? If so, how?

I address this question theoretically by building a macroeconomic model with a heterogeneous banking sector and an interbank lending market. Banks are endowed with the ability to transform deposits from households into loans to firms, but they differ in this ability endowment. From these ability differences, bank size differences emerge endogenously in the model. In steady state, the induced bank size distribution qualitatively matches two stylized facts in the data: bigger banks borrow more on the interbank lending market than smaller banks, and bigger banks are more leveraged than smaller banks.

In order to study downturns, I introduce an exogenous shock that destroys the value of net worth of all banks. In a downturn, the average ability of borrowing banks falls, worsening the extensive margin, and with it, the misallocation of funds across banks. At the same time, borrowing constraints tighten as bank net worth falls, reducing size on the intensive margin. In the typical case, both effects combine to prolong and deepen downturns. Bank heterogeneity in the model enables us to examine the difference in the impact of downturns on individual banks. The differences qualitatively match what we see in the data: the collapse in interbank lending disproportionately hurts large banks, large banks de-lever and lose more in downturns, and firm investment financed by large banks falls more than investment financed by small banks. It will also turn out that asset size concentration will primarily depend on the concentration in the underlying intermediation ability distribution.

After calibrating the model economy to match the pre-crisis economy and banking sector of 2007, I induce a shock which results in a drop in output similar to that seen in 2007/08 financial crisis. I then consider the counterfactual: if the banking sector in 2007 had only been as concentrated as it was in 1992, would the same shock produce a smaller drop in output? I adjust banking sector size concentration by adjusting the underlying intermediation ability distribution. I find that the answer is yes, the same shock produces a 40% smaller drop in output in the less concentrated economy. Further, output recovers to its steady state level almost twice as quickly.

Why worry about the size distribution of banks? First, the financial crisis showed us that there may be good reasons to think that big banks may matter a lot for the macroeconomy purely because of their size. The knock-on effects from the failures of large banks and the potential moral hazard problems generated by large bank bailouts have been pointed out as inherently de-stabilizing to the system. More recently Bremus et al. (2013) points out that, just because of their size, the largest banks are so large that their individual failures can influence the macroeconomy.

Second, there are good reasons to think that big banks operate in qualitatively different ways than small banks. Big banks are more leveraged, lose more in downturns, and rely more on interbank lending to finance their day-to-day operations. There is some evidence to suggest that big banks may benefit from economies of scale.

Rezitis (2006) considers the production of assets from deposits, labor, and capital in the Greek banking sector, and finds that large banks enjoy higher productivity than small banks. Hughes and Mester (2013) views bank productivity as the quantity of assets and deposits generated for a given level of labor hours and capital, and finds that banks enjoy economies of scale in this type of production. Kovner et al. (2014) finds that bank holding company size increases with noninterest expense, suggesting that large bank holding companies have lower operating costs. On the other hand, because of their systemic importance, big banks may receive a "too-big-to-fail" implicit guarantee of government bailouts, allowing them to operate with riskier balance sheets. Some of the apparent scale benefits may actually come from a "too-big-to-fail" funding advantage for larger banks. Recent work (e.g. Tsesmelidakis and Merton (2012) and Jacewitz and Pogach (2014)) has uncovered empirical evidence for this advantage.

Last, the size distribution has transformed in recent decades. Over the last 40 years, the concentration of the banking system has increased dramatically, as shown in figure 1. Higher concentration means that big banks make up a larger share of the banking sector than in the past, and therefore the actions taken by big banks, and the interactions between big and small banks, matter more for outcomes in the banking sector.

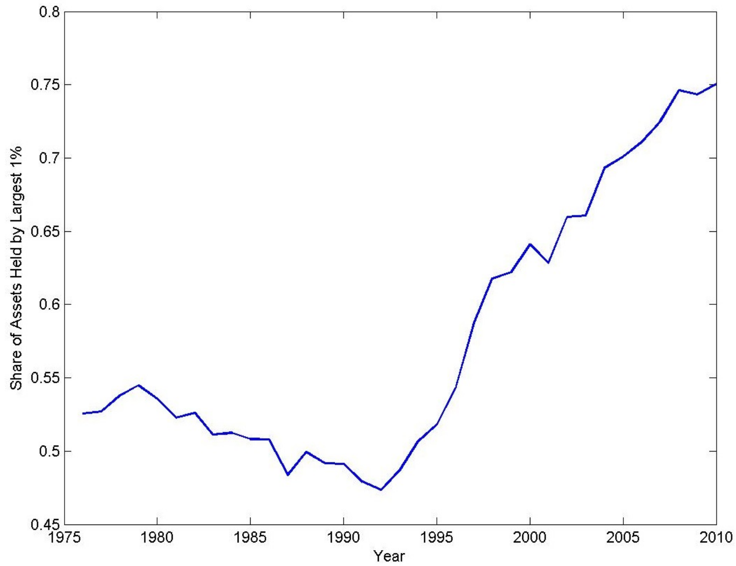
Whether this trend has affected the economy negatively is still a topic of debate. One view in the literature is that concentration makes the economy less prone to crises because large banks are also more diversified, and are therefore better able to maintain the credit they provide to firms in the event of a downturn. For example, Beck et al. (2007) perform a reduced-form, cross-country study using a global dataset of banks, finding that concentration is associated with fewer crises (though they do not find evidence specifically supporting the diversification theory). At the same time, other research supports the idea that large banks are less disciplined by competition than smaller banks, make poorer lending choices, and lose more in downturns. The model of Boyd and De Nicolo (2005) predicts that banks in less competitive environments charge higher interest rates to firms, which induces firms to take on greater risk and default more often. Supporting this hypothesis, De Nicolo et al. (2004) performs a cross-country study using a dataset of large banks, and finds that higher concentration is associated with higher fragility of the largest five banks in a given country. In some ways, this division in the research still persists. Corbae and D'Erasmo (2010) builds a two-tiered model of the banking sector, where a few dominant banks interact with a competitive fringe. They find evidence that an increase in banking concentration (implemented as an increase in entry costs) has offsetting effects: bank exit decreases, increasing stability, while interest rates increase, decreasing stability as in Boyd and De Nicolo (2005).

Mechanically, this paper follows a strand of literature in macroeconomics that considers the impact of information frictions in the financial sector on financial crises. The model in this paper builds on the model of Gertler and Kiyotaki (2011), who incorporate an interbank lending market in a macroeconomic model with a banking sector composed of identical banks. The agency problem in that model keeps aggregate investment below its efficient level in normal times, and amplifies the adverse effects of financial shocks in downturns. Boissay et al. (2013) consider the effects of this agency problem with a heterogeneous banking sector, and are able to create an economy in which banking crises arise endogenously.[1]

My paper adds two elements to this strand of literature. First, my focus is different: I am

[1]This paper is also related to a broader literature considering the quantitative effects of financial frictions in macroeconomic models. Carlstrom and Fuerst (1997) implement a friction generated by costly monitoring of borrowers, and focus on the effects of a shock to net worth, as in this paper. Jermann and Quadrini (2012) consider the effects of a friction generated by the possibility of default by firms, similar to the friction considered here, and find that the addition of this friction is important in explaining dynamics during the crisis. Christiano et al. (2010) find a more general result, weighing the quantitative effects of several frictions in the literature, finding their addition to generally improve the output of this class of models. For a good review of this literature, see Christiano et al. (2010) and Christiano and Ikeda (2011).

Figure 1: Concentration in the US Banking Sector



Note: Share of assets held by largest 1% of banks calculated for all commercial banks from 1976 to 2010. Data from call reports (forms FFIEC 031 and 041), retrieved from Federal Reserve Bank of Chicago (2013).

more interested in the connection between banking industry characteristics and downturns, and as such, I restrict my attention to the analysis of industry trends. Second, my model is different: leverage, or the assets banks hold per dollar net worth, will differ with size. This margin will generate a connection between the depth of downturns and the variance of the size distribution - the larger the highest bank is relative to the smallest, the more output declines in response to a given decrease in liquidity, the deeper the recession.

The paper proceeds as follows. I begin by presenting some stylized facts about the banking sector. The second section presents the model framework, describing its solution and discussing its general properties. The third section presents the counterfactual exercise described above, namely, to consider how the economy in 2007 may have responded differently if the banking sector had only been as concentrated as it was in 1992. A fourth section concludes.

2 Banking Facts

All banks are not the same, and size appears to be a key determinant of differences in bank behaviors. In this section, I document three trends in banking which relate to size and play a central role in my model. I then revisit concentration in the US banking sector, along with two related trends that will be important for quantitative analysis.

First, big banks make a lot more loans and deposits than small banks. The first row of table 1 shows the differences in the volume of loans performed by banks of different sizes calculated from Federal Reserve Economic Data, where large is defined as the largest 25 banks by asset size. With respect to commercial and industrial loans in isolation or all loans and leases in general, the largest banks lend twice as much as the smallest. On the deposit side, banks also perform more intermediation activity than small banks. Again, large banks receive twice as many deposits as small banks.

Second, bigger banks tend to be more leveraged than smaller ones. Table 1 shows the inverse of the Tier 1 leverage ratio, calculated for all bank holding companies by the Federal Reserve Bank of New York (2009). This measure is the ratio of the bank's risk-weighted assets, where assets are weighted by their credit risk, to the bank's tier 1 capital, which consists of a bank's equity and its retained earnings. Larger values of this ratio imply that less of a bank's asset holdings are funded through the bank's equity. In this sense, leverage is a measure of the bank's vulnerability to downturns. Both before and after the onset of the financial crisis, large banks were more leveraged than smaller banks. In the run up to the financial crisis, leverage increased for all banks, and in the midst of the crisis, leverage decreased substantially. This partly reflects the tightening of borrowing constraints (margins) banks faced when the value of many assets was deemed uncertain.[2]

Third, big banks tend to rely more on short-term liquidity markets than small banks do. Banks rely on interbank lending and short-term debt to fund their day-to-day operations. With respect to interbank lending, there are several studies that indicate that when interbank liquidity markets stop functioning, existing credit relationships are interrupted, causing damage to the real economy. Ivashina and Scharfstein (2010) finds that banks that relied less on interbank liquidity reduced their lending to nonbank borrowers in the wake of the financial crisis. Puri et al. (2011) find that banks that were affected worse by liquidity shortages rejected more potential borrowers than banks that weren't.

The importance of these markets depends on bank size. With respect to interbank lending, Furfine (1999) finds that net borrowers of Fed funds tend to be larger in asset size

[2] See Brunnermeier (2009) for a detailed explanation of this mechanism and its consequences.

Table 1: Large and Small Bank Differences: Intermediation Volume and Leverage

Volume of Intermediation

Measure	Large Banks	Small Banks	Ratio Large to Small
All Loans	2314	1210	1.91
Commercial and Industrial Loans	497	244	2.00
Total Deposits	2516	1318	1.90

Leverage

Measure	Large Banks	Small Banks
Assets/Equity		
Average - All Years	13.1	11.0
1992	17.0	13.0
2007	10.0	9.50
Assets/Tier 1 Capital		
Average - All Years	15.4	12.7
1996	14.8	11.9
2007	17.0	14.3

Note: Top panel: loans, deposits, and interbank loans in units of billions of US dollars for large and small commercial banks, where large is defined as the largest 25 banks by asset size. Averages calculated from monthly data over the period April 1988 Q1 to 2013 Q3. Recessions are defined by NBER recessions. Data are also displayed in 1992 and 2007, the reference years considered in the quantitative analysis, to show the effects of industry trends.
Bottom panel: ratio of total assets to equity for large and small commercial banks, where large is defined as banks with asset size larger than $20 B, for quarterly data over the period April 1985 to September 2013. Next, ratio of risk-weighted assets to tier 1 capital for large and small banks, where large is defined as all banks with asset size larger than $500 B, for quarterly data from 1996 Q1 to 2013 Q2.
Sources: (*Label - Series Name*)
Top Panel: All Loans - Loans and Leases from Bank Credit, Large/Small Domestically Chartered Banks; C+I Loans - Commercial and Industrial Loans, Large/Small Domestically Chartered Banks; Total Deposits - Deposits, Large/Small Domestically Chartered Banks
Bottom Panel: Assets/Equity Large - inverse of Total Equity/Total Assets, Banks with Total Assets over $20 B [EQTA5]; Assets/Equity Large - Total Equity/Total Assets, Banks with Total Assets less than $20B [EQTA1-4]
Data from Board of Governors of the Federal Reserve System, retrieved from Federal Reserve Bank of St. Louis (2013).
Assets/Tier 1 Capital - inverse of leverage ratio, from Federal Reserve Bank of New York (2009).

than net lenders of funds. Cocco et al. (2009) uncovers a similar result in the Portuguese interbank lending market, finding that larger banks borrow more often and borrow more when they do. With respect to repo, a form of securitized lending banks use in a similar way as interbank loans, Fecht et al. (2011) find that, in auctions for repo, banks that bid for repo were larger than non-bidders.

Industry-wide, banking sector concentration has increased, which has changed the relationship of large banks to small banks over time. A primary objective of this paper is to understand whether concentration has made downturns worse or better. The answer, it turns out, will depend on how concentration manifests itself.

Increasing concentration is typically characterized by an increase in the share of assets held by the largest banks, but this increase in share can arise in a number of ways. In the US data, the increase occurred for two reasons. First, banks have become more dispersed by asset size - the size of the biggest and smallest banks have become more extreme relative to the mean. Second, there are relatively more big banks in the banking sector today - the bank size distribution has become more skewed to the right.

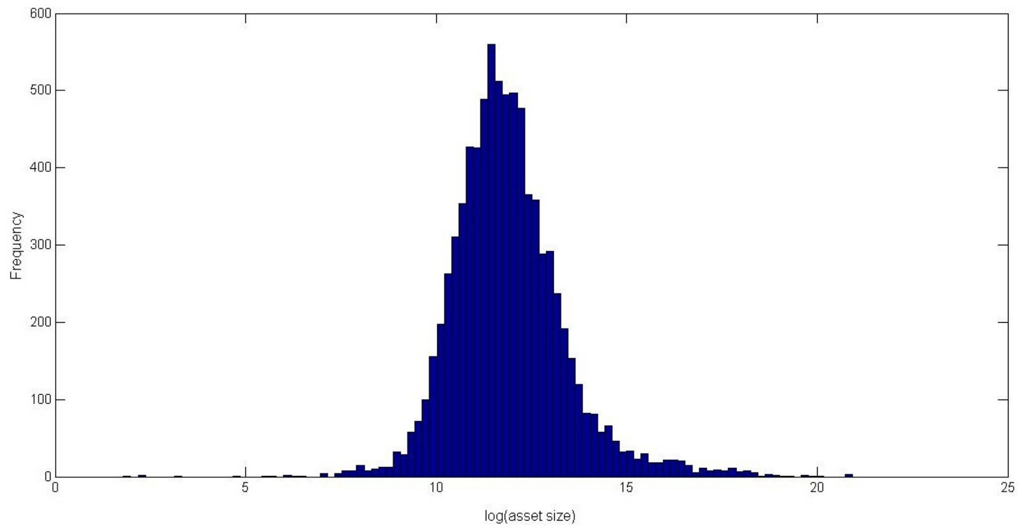
Though it has changed over time, the distribution has maintained two typical characteristics: it has a large mass of small banks, so the left side of the distribution is heavy, and it has a long right tail, representing a few very large banks. Figure 2 is a histogram of banks in 2007 by the logarithm of their asset size. Janicki and Prescott (2006) considers similar bank size distributions over time, and conclude that the distribution is best captured by a lognormal distribution with a Pareto tail.

In answering this question, I examine the banking sector in two reference years, one to represent a low concentration year, another to represent a high concentration year. I choose 1992 as the reference low concentration year for several reasons.[3] I choose 2007 as the high concentration year for comparison to the financial crisis.

Table 2 gives several measures of concentration. This increase corresponds to an increase in the share of total assets held by the top banks. The overall upward trend is robust to changes in the definition of concentration; as we see in the table, whether measured by the top 1% share, top 10% share, or the GINI coefficient, concentration has increased.

[3] In figure 1, we saw that the share of assets held by the largest 1% of banks slightly decreases over the 1976-1992 period, and then increases until 2010. This apparent trend break in concentration is accompanied by trend breaks in dispersion, or the standard deviation relative to the mean, and skewness/kurtosis, which for the bank size distribution are synonymous with an increase in the fatness of the right tail. In both cases, these roughly move with concentration over the period - between 1976 and 1992, these measures decrease, while between 1992 and 2010, these measures increase. Historically, this roughly coincides with the end of the savings and loan crisis, which took place during the mid-1980s and early 1990s. During this period, many smaller commercial banks faced competitive pressure from savings and loans, leading to a reduction in the number of small banks.

Figure 2: Log Bank Size Distribution



Note: Frequency plot of banks by their log (asset size) in 2007, where asset size is measured in thousands of US dollars. Calculated from call reports FFIEC 031 and 041, retrieved from the Federal Reserve Bank of Chicago (2013).

As the banking sector has become more concentrated, it has also become more dispersed and heavier in the right tail. Dispersion is quantified in two ways in the table: first, in the coefficient of variation, or the ratio of the standard deviation to the mean, and second, in the interquartile range, or the difference in size between the banks at the 25^{th} and 75^{th} percentiles in size.[4] Both measures have increased, indicating that banks have become more dispersed over time. The number of large banks has also increased relative to the number of smaller banks. This is indicated by an increase in skewness of the size distribution over the period. Throughout the period, the bank size distribution has a long right tail; therefore, the increase in skewness indicates an increase in the mass of the right tail of the size distribution.

3 Model

I consider an infinite-horizon economy composed of a continuum of islands $a \in [0, 1]$. On each island, there are many households that supply labor and save funds, many firms that

[4]Because banks are generally larger today than in the past, the variance has increased over time, the interquartile range (measured in units of dollars of assets) has increased much more than the coefficient of variation (a unitless quantity).

Table 2: Increasing Concentration and Related Trends

Measure	1992	2007
Concentration		
Top 1% Share of Assets	0.47	0.72
Top 10% Share of Assets	0.81	0.91
GINI Coefficient	0.85	0.92
Dispersion		
Coeff. of Variation	7.65	15.6
Interquartile Range	97063	180933
Right Tail		
Skewness	33.7	39.5

Note: Measures of concentration, dispersion, and skewness calculated from the distribution of total assets across banks in 1992 and 2007. Assets and interquartile range measured in thousands of US dollars. (Interquartile range in 2007 adjusted to 1992 dollars using series GDP Implicit Price Deflator in the United States, from OECD, retrieved from Federal Reserve Bank of St. Louis (2013).) Data from call reports FFIEC 031 and 041, retrieved from Federal Reserve Bank of Chicago (2013).

produce consumption goods from capital goods and labor, and many banks that intermediate funds from households and lend them to the firms. Households and firms will be identical across islands, but banks will not.

Following Gertler and Kiyotaki (2011) and Gertler and Karadi (2011), households and bankers on each island will belong to the same family, or economy-wide household. This induces two desirable features in the model. First, households will own banks, and receive dividends that bankers pay out. Second, households will insure one another against consumption risk across islands. This setup is maintained to ensure that we can represent the household side of the model with a single representative household. All the heterogeneity in the model will be isolated in the banking sector.

Banks will differ in a single dimension, intermediation ability, defined as the quantity of assets they can produce for each dollar of liabilities. Ability differences will generate differences in investment demands across islands, which will be satisfied through the reallocation of funds through an interbank lending market. Reallocation will be imperfect, however, and this imperfection will generate size differences between banks. In order to abstract away any distributional consequences other than what I am interested in, I will specify the model in a way that the size distribution does not endogenously evolve over time.

I will denote a single island by a, and an ability type by κ. I will refer to the quantity x

on island a with $x(a)$. The mass of any set of islands A will be given by the measure $\mu(A)$.

3.1 Banks

There are many potentially infinitely lived, risk-neutral banks on each island. These banks raise deposits from households on the same island, raise interbank loans from banks on all other islands, and invest in the firms on the same island. All banks on the same island are identical, but banks on different islands are not.

Banks in this model are intermediaries, that is, they take deposits from households and invest them as loans to firms. Banks will be heterogeneous in their ability to perform this task. We can think of intermediation ability as a productivity term in a simple bank production function, so that if banks are producing assets S valued at price Q from net worth N:

$$QS = \kappa N \tag{1}$$

After they receive deposits but before they borrow interbank loans, all the banks on an island a receive a random, iid ability draw $\kappa(a)$. Draws are assumed to be distributed with cumulative distribution function $\kappa(a) \sim F(\cdot)$, with bounded support $[\underline{\kappa}, \overline{\kappa}]$.

Banks raise deposits in an economy-wide deposit market. Each bank offers a common deposit rate R_t to households for a one-period deposit. This means that for each dollar deposited in period t, the household will be repaid R_t dollars in period $t+1$. I denote the total deposits raised by all banks on each island in time t by d_t. Deposits are written without an island identifier because they raise identical quantities of deposits.[5]

Banks on any island are able to borrow or lend to banks on any other island. Since there are many banks on each island, instead of characterizing loans between individual banks, I will be interested in the net loans made between all banks on one island with all banks on another. Ability and amount borrowed are observable, so we can characterize all interbank loans with a contract specifying the interest rate and the borrowing quantity, $(R_{bt}, b_t(a))$, where a is the island where the borrower banks live.[6]

[5]Banks receive their ability draw after they raise deposits, and ability draws are iid. Therefore, if the banks on any given island a raise $d_t(a)$, and banks on another island raise $d_t(a')$, then $d_t(a) = d_t(a')$. I refer to this common quantity as $d_t = d_t(a) = d_t(a')$.

[6]Technically, an interbank loan contract is bilateral. Therefore, we should specify an interbank lending rate for each lender-borrower pair, or $(R_{bt}(a, a'), b_t(a, a'))$, where a, a' are the island where the borrower and lender banks live, respectively. It will turn out that the only rates we observe in equilibrium will be both borrower and lender independent, however. The rate is lender-independent because there are a continuum of potential lenders; since lenders compete with each other for borrowers across all islands, and loan contracts are identical to borrowers, only one lending rate will prevail. The borrowed amount will be

Banks lend a mixture of capital and consumption goods to the firms on the same island. Firms can use these loans to buy new capital goods in period t, and then use the old and any new capital to produce consumption goods in period $t+1$. The firm then gives the bank its output (less wages) along with any undepreciated capital goods, also at time $t+1$.

Banks are potentially infinitely-lived, but face an incentive constraint that I will present below. To prevent the bank from saving its way out of this constraint, a constant proportion σ of banks on each island exit every period. Upon exit, a bank transfers its earnings to the household on its island. Second, new banks enter to replace the old banks. To ensure that these new banks have something to invest with, they receive a "start-up transfer" equal to a constant fraction ξ of the total assets held by all banks on the island.

Banks carry wealth from period to period in the form of net worth, defined as the payoff from assets less deposits and interbank loans. The net worth of the island a representative bank at time t is given by:

$$n_t(a) = [Z_t + (1-\delta)Q_t(a)]\psi_t(\sigma+\xi)s_{t-1}(a) - R_{t-1}d_{t-1}(a) - R_{bt-1}b_{t-1}(a) \tag{2}$$

The first term gives the bank's returns on investments in the firm: Z_t is the economy-wide representative firm's gross profits from investments (per unit invested), $Q_t(a)$ is the price of capital on island a, and $s_{t-1}(a)$ is the units of capital held by the firm in period $t-1$. The second term gives the bank's repayments for deposits and interbank loans: $b_{t-1}(a)$ are the funds borrowed on the interbank market in the last period, and $d_{t-1}(a)$ are the deposits made by households in the last period.

ψ_t is the "quality of capital"; we can view this as an underlying asset value. This variable will be the key shock in this model. This type of shock behaves differently from a shock to total factor productivity. Most important, a shock to capital quality exogenously affects the value of net worth; because the price of capital is endogenous to this model, this shock will also exogenously decrease asset prices. This type of shock has been used in recent work, including Gertler and Kiyotaki (2011), because it mimics a financial crisis - a shock to capital quality mimics an unforeseen drop in asset values (e.g. housing crisis).

We can now revisit the balance sheet of the bank with the new variable descriptions in hand. Banks balance assets against the *intermediated* value of liabilities and equity, which in this case is the sum of net worth, deposits, and interbank borrowing. We can summarize

borrower-dependent because of the financial friction I describe below. Then the debt contract will have the property that $(R_{bt}(a,a'), b_t(a,a')) = (R_{bt}, b_t(a))$.

the balance sheet of the bank with a flow of funds constraint:

$$Q_t(a)s_t(a) = \kappa_t(a)\left[n_t(a) + d_t + b_t(a)\right] \tag{3}$$

In every period, banks are supposed to repay depositors and banks they borrowed from in the previous period. Instead of repayment, however, banks can choose to default on their loans, in which case they take a fraction $\theta \in [0, 1]$ of the total funds $Q_t(a)s_t(a)$ and exit forever. If a bank chooses to default, it does so at the end of a period, and sells its shares to another bank on the island.[7] On the other hand, there will be no friction between banks and firms on the same island - banks will be able to enforce full repayment of loans made to firms.

We can now state the bank's maximization problem. Bank managers choose a quantity of deposits, loans to firms, and interbank loans today to maximize the expected present value of future dividends. Households are bank owners, and the banker internalizes this when calculating the expected value of dividends, discounting it by $\Lambda_{t,t+i}$, the stochastic discount factor of the economy-wide representative household.[8] At time t, the bank formulates a plan for the path of deposits, assets, and interbank loans. The plan is state-contingent, so that it chooses a different level of deposits d_t for each value of aggregate shock it faces in that period, ψ_t, and a different level of $s_t(a)$ and $b_t(a)$ for each level of aggregate shock and ability shock $\kappa_t(a)$ it receives in that period.

$$\begin{aligned} V_t(d_{t-1}, s_{t-1}(a), b_{t-1}(a)) = & \underset{\{d_{t+i},(s_{t+i}(a_i))_a,(b_{t+i}(a))_a\}_{i=0}^{\infty}}{max} E_t \sum_{i=1}^{\infty}(1-\sigma)\sigma^{i-1}\Lambda_{t,t+i}n_{t+i}(a_{t+i}) \\ & s.t.\ V_t \geq \theta Q_t(a)s_t(a) \end{aligned} \tag{4}$$

Note that $n_t(a)$ in (4) is a state variable here. At the time the bank makes its deposit demand decision, it only knows the expected value of its net worth. However, at the time it makes its asset and interbank loan decisions, it knows its net worth precisely.

The constraint in the above problem is an incentive constraint imposed by the bank's ability to default. In equilibrium, the continuation value of staying on the equilibrium path has to exceed the dollar value of assets the bank can run away with.

[7] I interpret this "running away" as capturing a bankruptcy cost: if a borrower bank decides to declare bankruptcy, creditors can capture some, but probably not all, of the repayments they are owed.

[8] As mentioned in the setup, households insure one another against differences in consumption across islands.

We can rewrite the above problem in a Bellman equation:

$$V_t = \max_{d_t} E_t \max_{s_t(a),b_t(a)} E_{t,a}\Lambda_{t,t+1}\left[(1-\sigma)n_{t+1}(a) + \sigma \max_{d_{t+1}}\left(\max_{s_{t+1}(a),b_{t+1}(a)} V_{t+1}\right)\right]$$
$$s.t.\ V_t \geq \theta Q_t(a)s_t(a) \tag{5}$$

3.1.1 Additional Assumptions

I make two additional assumptions to ensure a tractable solution to the model. First, even if ability draws are independent across periods, the interbank loans from the previous period will make the expected returns from assets, or the ratio of net worth to capital, unequal. In this case, the history of previous draws will matter for bank returns today. To prevent this, I follow Gertler and Kiyotaki (2011) and allow individual banks to arbitrage these return differences away, before their new ability type is realized:

Assumption 1. *At the beginning of a period, banks can move between islands to equalize expected returns on assets.*

Consider two islands, a_L and a_H. a_L has low expected returns when it enters the period, that is, the representative bank has high interbank debt obligations, and a_H has high expected returns. When given the opportunity, an individual bank on a_L decides to move to a_H. It currently holds assets (investments) in firms on a_L, interbank debt to banks on a_H, and deposits (economy-wide). Before it moves, it trades its assets with another bank on a_L for more interbank debt to a_H. It then takes its net worth and deposits and moves; this reduces the individual bank's interbank debt of a_L to a_H but maintains the total asset held in a_L firms. This process continues until returns are equalized.

The second assumption has to do with the capital goods that carry over from previous periods. Intermediation ability also applies to the undepreciated capital that banks carry between periods - banks will reallocate (or re-intermediate) the undepreciated portion of the capital stock among firms on the same island. Re-intermediation of capital will leave some islands with higher production capacity than others. To ensure that we can maintain a simple law of motion for capital, I assume that these changes to production capacity can be realized as gains in the output of firms in the same period:

Assumption 2. *When an island's existing capital $k_t(a)$ is intermediated, any gains (losses) in the production possibilities of that capital are realized as an increase (decrease) in output of all island firms $y_t(a)$ in the same period.*

3.2 Households

Households on each island a are infinitely-lived, supply up to one unit of labor per period, save in the form of deposits, and consume consumption goods. The household makes its deposit and labor supply decisions before the banks on its island realize their ability. Because households across islands are members of the same family and insure one another against consumption shocks, we can represent the decisions of each household with an economy-wide representative household.

Households on each island save by making riskless one-period deposits D_t in the economy-wide deposit market[9] at the interest rate R_t. Deposits made in the last period are repaid at the beginning of this period.

I make another simplifying assumption on labor:

Assumption 3. *Workers can supply labor on any island.*

With the above, wages will be identical across islands. Let's call this economy-wide wage W_t. The household also owns the bank, and receives dividends from exiting banks every period Π_t. Then the household's budget constraint is

$$C_t = W_t L_t + \Pi_t + R_{t-1} D_{t-1} - D_t \tag{6}$$

The household's maximization problem becomes

$$\begin{gathered} \underset{(C_t, L_t, D_t)_{t=0}^{\infty}}{max} E_t \sum_{i=0}^{\infty} \beta^i \left[ln(C_{t+i} - \gamma C_{t+i-1}) - \frac{\chi}{1+\varphi} (L_{t+i})^{1+\varphi} \right] \\ s.t.\ C_t = W_t L_t + \Pi_t + R_{t-1} D_{t-1} - D_t \text{ (for each } t) \end{gathered} \tag{7}$$

where $\beta \in (0,1)$ is the discount factor and $\gamma \in [0,1)$ is a habit formation parameter[10]. φ is the inverse Frisch elasticity of labor supply.

Taking first order conditions, we first work out a condition for aggregate deposits:

$$R_t (E_t \Lambda_{t,t+1}) = 1 \tag{8}$$

[9] We could alternatively restrict the household to only making deposits in the banks on their island; with the timing assumptions below, nothing about the model changes. I maintain this formulation for ease of explanation.

[10] These preferences exhibit habit formation when $\gamma \in (0,1)$, a feature that is included for comparison to other models in the literature. The model is not fundamentally different when we turn off habit formation, i.e. set $\gamma = 0$.

and a condition for aggregate labor:

$$W_t E_t(u_{Ct}) = \chi(L_t)^{\varphi} \tag{9}$$

where $u_{Ct} \equiv (C_t - \gamma C_{t-1})^{-1} - \beta\gamma(C_{t+1} - \gamma C_t)^{-1}$ is the marginal utility of consumption and $\Lambda_{t,t+1} \equiv \beta \frac{u_{Ct+1}}{u_{Ct}}$ is defined as the household's stochastic discount factor.

3.3 Firms

There are many identical, competitive firms on each island. Immediately after the bank's ability is realized, the firm uses the capital it held from the last period to produce output. It makes its loan repayment to the bank, and the bank then makes its loan to the firm for next period production.

Coming into period t, there is capital $k_t(a)$ on island a. The firm's problem is then reduced to one of choosing how much labor to input. The representative firm on island a chooses $l_t(a)$ to maximize:

$$y_t(a) = A_t k_t(a)^{\alpha} l_t(a)^{1-\alpha} \tag{10}$$

where A_t is the (economy-wide) total factor productivity of the firm.[11] Since labor is mobile, firms on every island face the same wage W_t. Firms will choose labor to equate the economy-wide wage with the marginal product of labor:

$$W_t = (1-\alpha)\left(\frac{y_t(a)}{l_t(a)}\right) = (1-\alpha)\frac{Y_t}{L_t} \tag{11}$$

where the second equality follows from the fact that the ratio of capital to labor is constant across islands. Thus, to find the optimal capital/labor ratio on each island, we only need to solve the economy-wide representative firm's problem:

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{12}$$

Even though the ratio of net worth to capital is identical across islands by assumption 1, the capital on each island is unknown. As a result, the size distribution of firms across islands in equilibrium will be indeterminate. (In the solution below, I will show that despite this, we can still pin down the size distribution of firms in a meaningful way.)

[11] This will be constant and equal to 1 throughout the paper, though TFP shocks can be induced in the model through this channel.

After production happens, capital depreciates, leaving $(1-\delta)k_t(a)$ on the island. As part of the repayment to banks, firms transfer ownership of this capital to banks. Banks then decide on the loan package $s_t(a)$, which comprises this undepreciated capital and (possibly) of cash (basic goods) for new investment, which I denote by $i_t(a)$. This implies that

$$s_t(a) = (1-\delta)k_t(a) + i_t(a) \tag{13}$$

3.4 Capital Goods Producers

When firms decide to expand their existing capital stock, they travel to a central (economy-wide) market for new capital. The market is perfectly competitive, but all producers face adjustment costs. Capital goods producers sell new capital to firms for the price Q_t^i. These producers then choose I_t to maximize

$$E_t \sum_{\tau=t}^{\infty} Q_\tau^i I_\tau - \left(1 + f(\frac{I_\tau}{I_{\tau-1}})\right) I_\tau \tag{14}$$

From the first order condition for this profit maximization problem, the new capital price should satisfy

$$Q_t^i = 1 + f\left(\frac{I_t}{I_{t-1}}\right) + \left(\frac{I_t}{I_{t-1}}\right) f'\left(\frac{I_t}{I_{t-1}}\right) - E_t\Lambda_{t,t+1}\left(\frac{I_{t+1}}{I_t}\right)^2 f'\left(\frac{I_{t+1}}{I_t}\right) \tag{15}$$

I'll add another assumption to ensure that the law of motion for capital has a simple form. With this assumption, banks will always find it optimal to reinvest their capital stock.

Assumption 4. *Once installed, existing capital cannot be used on any other island.*

3.5 Market Clearing Conditions

When we combine the household budget constraint with the equation for firm output and new capital, we can write an economy-wide resource constraint for basic goods:

$$Y_t = C_t + \left(1 - f\left(\frac{I_t}{I_{t-1}}\right)\right) I_t \tag{16}$$

With the assumptions above, we can write a simple law of motion for capital:

$$K_{t+1} = \psi_{t+1}\left(I_t + (1-\delta)K_t\right) \tag{17}$$

A period is divided into two parts: before and after intermediation ability is realized.

Figure 3: Timing

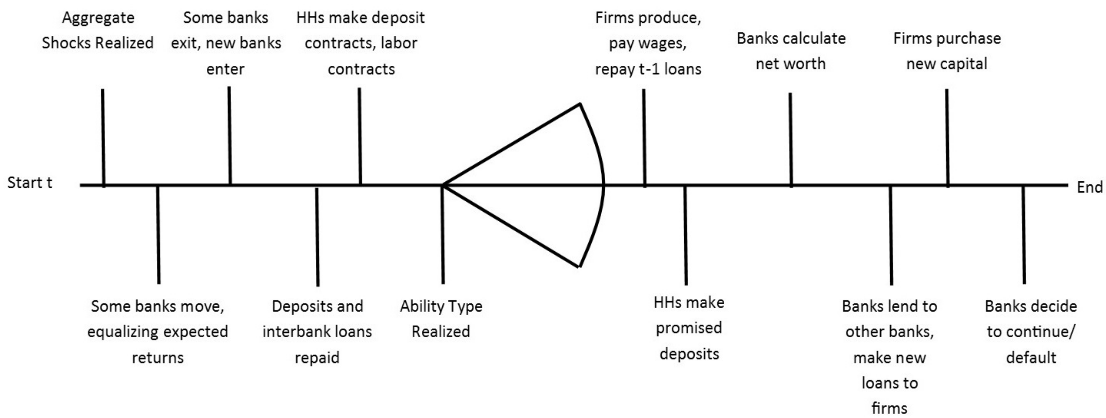


Figure 3 gives the timing of the model. First, the aggregate shocks ψ_t and A_t are realized. Banks then move to equalize expected returns, and some banks exit exogenously. Banks then pay dividends to households, and new banks enter in their place. Deposits and interbank loans are then repaid, and households make their deposit and labor supply decisions.

The second half of the period begins when intermediation ability types are realized. Firms borrow from banks, and use household labor and existing capital to produce output. They then pay profits Z_t to banks and wages W_t to workers. Those firms that want to expand their capital stock for next period's production use the loans to purchase capital from the capital goods producers (in the central new capital market). Simultaneously, banks borrow from other banks. Banks then make their plan for the future, deciding in the process whether to default before the next period.

As mentioned above, the size distribution of firms across islands will be indeterminate in equilibrium. In what follows, I will first define and solve the model aggregated to the level of ability type. To make things easier, I will first characterize the solution to the bank's maximization problem.

3.6 Bank Optimization

The bank on island a maximizes net worth, which itself is a linear function of assets, deposits, and loans. Because of this linearity, we'll guess and verify (details are left to the appendix) that the bank's value function will also be linear. The problem then boils down to solving for the coefficients of a linear value function.

First, guess that in every period a bank maximizing a linear function of shares, deposits, and interbank loans will maximize its expected net worth. Specifically, this function will take the following form:

$$V_t = \nu_{st} s_t(a) - \nu_{bt} b_t(a) - \nu_t d_t \tag{18}$$

In what follows, I restrict attention to equilibria with interior solutions to the above guess, so that banks can only hold nonzero quantities of all choice variables - only these equilibria will have positive volumes of interbank lending.

Next, use the guess to set up a Lagrangian for the bank's problem. First, substitute the flow of funds constraint into the guess to reduce the number of choice variables to two:

$$\begin{aligned} V(d_t, s_t(a), b_t(a)) &= \nu_{st} s_t(a) - \nu_t d_t - \nu_{bt} b_t(a) \\ &= \frac{\kappa(a)\nu_{st}}{Q_t(a)}(n_t(a) + d_t + b_t(a)) - \nu_t d_t - \nu_{bt} b_t(a) \\ &= \frac{\kappa(a)\nu_{st}}{Q_t(a)} n_t(a) + (\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_t) d_t + (\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_{bt}) b_t(a) \end{aligned} \tag{19}$$

Note again that $n_t(a)$, the bank's net worth, is a state variable. Next, substitute the guess into the incentive constraint, and use it to construct a Lagrangian:

$$\begin{aligned} L(d_t, b_t(a)) = &\left((1 + \lambda_t(a)) \frac{\kappa(a)\nu_{st}}{Q_t(a)} - \lambda_t(a)\kappa(a)\theta \right) n_t(a) \\ &+ \left((1 + \lambda_t(a))(\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_t) - \lambda_t(a)\kappa(a)\theta) \right) d_t \\ &+ \left((1 + \lambda_t(a))(\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_{bt}) - \lambda_t(a)\kappa(a)\theta \right) b_t(a) \end{aligned}$$

$\lambda_t(a)$ is the multiplier on the incentive constraint in period t for the representative bank from island a.

We obtain the following first order conditions for d_t and $b_t(a)$:

$$\nu_t = \nu_{bt} \tag{20}$$

$$\kappa(a)\theta \frac{\lambda_t(a)}{1 + \lambda_t(a)} = \kappa(a) \frac{\nu_{st}}{Q_t(a)} - \nu_{bt} \tag{21}$$

Using the above, we can also rearrange the borrowing constraint:

$$b_t(a) \leq \phi_t(a) n_t(a) - d_t(a) \tag{22}$$

where

$$\phi_t(a) = \frac{\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \kappa(a)\theta}{\nu_{bt} - \frac{\kappa(a)\nu_{st}}{Q_t(a)} + \kappa(a)\theta} \tag{23}$$

$\phi_t(a)$ will be a useful quantity throughout the paper. It is related to the leverage ratio, denoted $L_t(a)$, or the ratio of shares held to net worth; $L_t(a) = \kappa(a)(1 + \phi_t(a))$. The leverage ratio $L_t(a)$ is an increasing and convex function of κ. Convexity arises because intermediation ability makes asset purchases easier for banks in two ways. First, intermediation ability decreases the cost of assets directly through its effect on the flow of funds constraint. Second, ability decreases the value to running away by increasing the returns to borrowing on the interbank market, which in turn loosens the borrowing constraint.

We can use the form of the linear solution to characterize the price of capital on each island. Recall that, after production happens, the bank receives ownership of the existing capital stock, which it can then re-lend to the firms as part of a loan package $s_t(a)$. Islands that are not borrowing constrained have $\lambda_t(a) = 0$, which implies that they are indifferent between lending on the interbank market and re-lending to firms. Faced with indifference between these choices, I will impose that the bank will always reinvest the existing capital stock on its island. The following assumption states this formally.

Assumption 5. $s_t(a) \geq (1-\delta)k_t(a) \ \forall a$

It will be convenient to group islands based on the borrowing activity of their banks (recall that all banks on a single island are identical) . If the banks on island a are borrowing constrained we will say that the island is part of the set B, and if banks on island a are not borrowing constrained, we say that the island is part of the set L.

Further, if banks on island a are not borrowing constrained, $\lambda_t(a) = 0$. Then from equation (21) it follows that

$$\frac{Q_t(a)}{\kappa(a)} = \frac{\nu_{st}}{\nu_{bt}} \equiv Q_t^n \ \ \forall a \in L \tag{24}$$

Assumption 4 also implies that the price of capital on each island is free to adjust to a different value on each island. Here we make another guess that we can verify later:

$$Q_t(a) = \kappa(a)\frac{\nu_{st}}{\nu_{bt}} = \kappa(a)Q_t^n \ \ \forall a \in L \tag{25}$$

where L is the set of ability types for which banks are not borrowing constrained. This price is related to the shadow cost of re-lending existing capital - if a bank on island a were given an extra unit of basic good, it could either lend it on the interbank market, earning the bank ν_{bt}, or it could turn it into $\kappa(a)\frac{1}{Q_t(a)}$ units of capital, earning it $\kappa(a)\frac{\nu_{st}}{Q_t(a)}$ worth of value.

We can also characterize the price of capital on islands with banks that are borrowing constrained. The FOC for $b_t(a)$ also tells us that constrained banks view assets as more valuable than interbank lending, implying that these banks also desire higher investment. If these banks were indifferent between investing in assets and lending on the interbank market, the price $Q_t(a)$ would get very large, as banks would demand ever higher returns per unit capital.

No firm on any island will ever allow the price of capital $Q_t(a)$ to exceed the price for new capital in the economy-wide new capital market, Q_t^i. To see this, imagine a bank that wants to lend a unit of existing capital to a firm, but demands repayment $Q_t(a) > Q_t^i$ for a unit of this capital. Rather than agreeing to these terms, the firm instead threatens to go to another bank on the island to get a loan, which it could then use to buy new capital at the lower price Q_t^i. Since there are many potential lenders on the island, the threat is credible, and the bank should drop the price of the existing capital to Q_t^i as well. Thus, the price of capital on borrowing constrained islands gets pinned down:

$$Q_t(a) = Q_t^i \ \forall a \in B \tag{26}$$

where B is the set of islands on which banks are borrowing constrained.

Some sets of parameters will not admit equilibria with positive lending. I make two restrictions on parameters to prevent these cases. I will choose parameters so that these restrictions hold in steady state, and in the quantitative analysis below, I will choose the level of the shock to be small enough to ensure that the restrictions hold in the impulse responses.

It will turn out that intermediation ability will have a simple relationship with borrowing behavior. Banks with ability high enough will borrow, and be borrowing constrained, while low ability banks will lend. I show this in the next proposition.

Proposition 1. *Consider an equilibrium with a strictly positive volume of interbank lending. In each period t, there exists a κ_t^* such that:*

- *for all banks with $\kappa(a) < \kappa_t^*$, $b_t(a) = -(n_t(a) + d_t - Q_t^n k_t(a))$, that is, the bank will lend all of its net worth and deposits less the cost of refinancing the entire existing capital stock on its island.*

- *for all banks with $\kappa(a) > \kappa_t^*$, $b_t(a) = \overline{b_t}(a)$, that is, the bank will borrow up to its borrowing constraint and use the funds to purchase assets.*

Proof. Consider the term for borrowing in equation (19), and let us first assume

$$(\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_{bt}) > 0$$

In this case, the bank gets positive value for every dollar it borrows, and it will choose to borrow as much as it can. This bank will therefore run into its borrowing constraint, i.e. until $b_t(a) = \overline{b_t}(a)$. Because all constrained banks face capital prices $Q_t(a) = Q_t^i$, this implies that $(\frac{\kappa\nu_{st}}{Q_t^i} - \nu_{bt}) > 0$, or after rearranging, $\kappa > \frac{\nu_{bt}}{\nu_{st}}Q_t^i$.

If $(\frac{\kappa\nu_{st}}{Q_t(a)} - \nu_{bt}) \leq 0$, the bank gets positive value for every dollar it lends (negative value for every dollar it borrows). This implies that the bank will lend as many funds as it can. Assumption 5 requires that the bank re-invest its entire existing capital stock; thus, the funds the bank actually lends out on the interbank lending market will be limited to those left over after this re-investment. These remaining funds are $b_t(a) = -(n_t(a)+d_t-Q_t^n k_t(a))$. Further, we know that capital prices on these islands are smaller than the new capital market price, so $Q_t(a) \leq Q_t^i$. Since $(\frac{\kappa\nu_{st}}{Q_t(a)} - \nu_{bt}) \leq 0$ implies that $\kappa \leq \frac{\nu_{bt}}{\nu_{st}}Q_t(a)$, it must also be that $\kappa \leq \frac{\nu_{bt}}{\nu_{st}}Q_t(a) \leq \frac{\nu_{bt}}{\nu_{st}}Q_t^i$.

Call $\kappa^* \equiv \frac{\nu_{bt}}{\nu_{st}}Q_t^i$. Then, for any bank with ability $\kappa > \kappa^*$, borrowing and investing is strictly more profitable than lending on the interbank market. For any bank with ability $\kappa \leq \kappa^*$, lending is (weakly) more profitable than borrowing. □

The above proposition shows that an island a will fall into one of the two sets introduced above: either banks on the island borrow from the interbank lending market, in which case the island is in the set B, or banks on the island lend to the interbank lending market, in which case the island is in the set L. Moreover, the proposition tells us that we can summarize these sets with one object, the ability cutoff.

It will also turn out that lower ability interbank lenders will lend less than higher ability interbank lenders. I show this in the next proposition.

Proposition 2. *Consider two banks on different islands a, a' with types $\kappa(a)$ and $\kappa(a')$ and $b_t(a), b_t(a') < 0$. If $\kappa(a) < \kappa(a')$, then $b_t(a) < b_t(a')$.*

Proof. The cost of refinancing the entire existing capital stock $k_t(a)$, net of the benefit derived in the form of net worth, is given by

$$\frac{Q_t(a)}{\kappa(a)}k_t(a) - Q_t(a)k_t(a)$$

Because of assumption 2, any differences in refinancing costs are reflected in output in the same period. This allows banks to lend the consumption good equivalent of the differences.

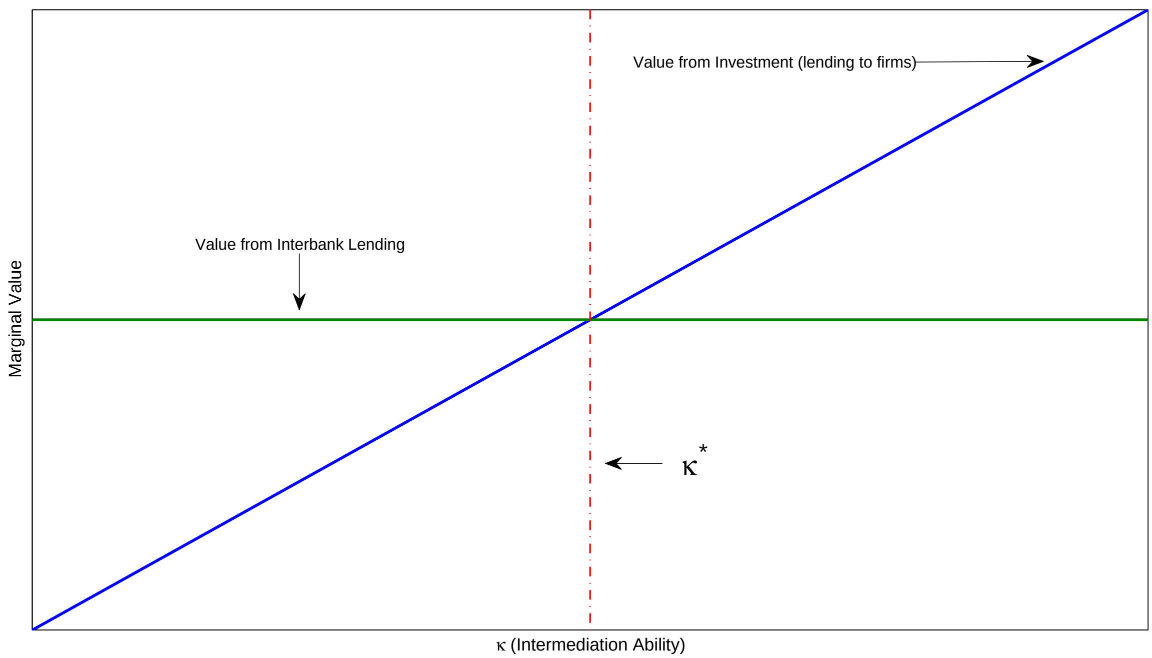
For unconstrained (lending) islands, this simplifies to

$$Q_t^n k_t(a)(1 - \kappa(a))$$

This is a decreasing function of κ. Thus, for interbank lenders the cost of refinancing the capital stock is decreasing in κ. By the flow of funds constraint, the amount left over for lending to other islands is then increasing in κ. $\square$

Since banks on islands with the same ability draw $\kappa(a)$ will make the same choices for d_t and $b_t(a)$, this result extends to all banks with the same ability κ.

Figure 4: Ability Cutoff



Note: Marginal value from interbank borrowing (green dash) and marginal value from holding shares (blue solid) versus intermediation ability. The intersection of the two lines marks the ability cutoff κ^*. Above the cutoff, banks act as net borrowers on the interbank lending market, while below the cutoff, banks act as net lenders.

Figure 4 is a visualization of proposition 1. It diagrams the marginal value from lending on the interbank market (dash line) and the marginal value from investing in firms in a typical steady state. The value of lending is the same for all banks since all banks receive the same interest rate for loans of any size. The value from investing/lending to firms, on the other hand, increases with bank intermediation ability. In any equilibrium with positive lending, the two lines have to cross - if the interbank lending line was always under the investment line, no bank would be willing to lend, and if the interbank lending line was always above the investment line, no bank would be willing to borrow. The ability level at which these two lines cross is the ability cutoff κ^* - for any bank with ability above the cutoff, the marginal value from investing is higher than the value from lending, so it becomes profitable for the bank to borrow and invest.

3.7 Equilibrium in Ability Types

The quantity of capital installed on each island in period t depends on the quantity of capital installed in period $t-1$. Since each island gets a new ability draw every period, the only way to determine the capital on each island is to keep track of the history of all ability draws over time. This requirement makes computing an equilibrium very difficult. In this section, I introduce an alternative, aggregated equilibrium concept that is much easier to solve, but still preserves the heterogeneity across banks.

In this alternative equilibrium concept, I will solve for prices and total quantities for groups of islands with the same ability draw, e.g. instead of solving for the quantity of interbank loans made by banks on a single island, I will determine the sum total value of interbank loans made by all banks on every island with the same ability draw. Aggregating the model in this way will lend itself to a tractable solution concept which still preserves heterogeneity across banks. For the remainder of the paper, the quantity x on all islands with the same ability type (a sum across individual islands) is called $x(\kappa)$, that is, $x(\kappa) = \int_{A_\kappa} x(a)da$, where $A_\kappa = \{a : \kappa(a) = \kappa\}$.

For any island receiving ability draw κ, there will be a positive measure of islands with the same ability draw.[12] Since the distribution is assumed to be iid across periods, any positive measure of islands will be representative of the entire economy in the previous period; the net interbank loan repayment in this measure will be 0, and the total capital

[12] To be more precise, we can justify this with the following setup: specify the set of islands over the space $[0,1] \times [0,1]$, where a specific island is referred to by two coordinates, $a = (a_1, a_2)$. Now assume that in even periods, all islands with same first coordinate receive the same draw, and in odd periods, all islands with the same second coordinate receive the same draw. Then for any island with a given level of κ, there will be a positive measure of islands with the same level of κ. Since $[0,1] \times [0,1]$ has the same cardinality as $[0,1]$, this same argument will work for the model in the main text. For ease of exposition, I leave this setup out of the main text.

installed last period will just be a fraction of aggregate capital: $b_{t-1}(\kappa) = 0$, $s_{t-1}(\kappa) = p(\kappa)K_t$.

On each island, banks choose deposits d_t identically, before uncertainty is realized. Banks choose $s_t(a)$ and $b_t(a)$ differently across islands, but two islands with the same ability type will choose identical values of $s_t(a)$ and $b_t(a)$. This means that banks on all islands with the same ability type will either all be borrowing constrained or not. This implies, then, that the price of capital on islands with the same ability type will be identical, that is, $Q_t(a) = Q_t(a')$ for all a, a' with $\kappa(a) = \kappa(a')$. This can also be shown for the quantity $\phi_t(a)$ as well.

We can use these facts to write an equation for the aggregate net worth of all banks with the same ability type, which I label $n_t(\kappa)$, and the aggregate quantity of loans made by all banks with the same ability type, which I label $s_t(\kappa)$:

$$n_t(\kappa) = [Z_t + (1-\delta)Q_t(\kappa)]\psi_t(\sigma + \xi)p(\kappa)K_t - p(\kappa)\sigma R_{t-1}D_{t-1} \tag{27}$$

$$s_t(a) = \frac{\kappa(a)}{Q_t(a)}(1 + \phi_t(a))n_t(a) \tag{28}$$

The first term in equation (27) gives the representative bank's returns on investments in the firm: Z_t is the economy-wide representative firm's gross profits from investments (per unit invested), $Q_t(\kappa)$ is the price of capital on all islands with ability κ, and K_t is the capital installed by all firms in the economy at the beginning of period t. The second term gives the bank's repayments for deposits and interbank loans: b_{t-1} is the funds borrowed on the interbank market last period, and d_{t-1} is the deposits made by households last period.

Since the quantities $Q_t(a)$, $\phi_t(a)$, and $\kappa(a)$ are identical for any islands with the same ability type, we can also write down an ability type level flow of funds constraint:

$$s_t(\kappa) = \frac{\kappa}{Q_t(\kappa)}(1 + \phi_t(\kappa))n_t(\kappa) \tag{29}$$

We can use this to solve for the ability type aggregates for banks with the maximization problem

$$V_t = \max_{\{d_{t+i},(s_{t+i}(\kappa_{t+i}))_\kappa,(b_{t+i}(\kappa_{t+i}))_\kappa\}_{i=0}^{\infty}} E_t \sum_{i=1}^{\infty}(1-\sigma)\sigma^{i-1}\Lambda_{t,t+i}n_{t+i}(\kappa_{t+i}) \tag{30}$$

$$s.t.\ V_t \geq \theta Q_t(\kappa)s_t(\kappa) \tag{31}$$

On the firm side, since net worth is pinned down for each ability type, and the ratio $\frac{n_t(a)}{k_t(a)}$ is equal for each island by assumption 1, we can pin down the quantities $k_t(\kappa)$ and $i_t(\kappa)$ for each ability type. These will take the form

$$k_t(\kappa) = p(\kappa)K_t \tag{32}$$

$$i_t(\kappa) = s_t(\kappa) - (1-\delta)k_t(\kappa) \tag{33}$$

Second, even though firms on each island may have different sizes, when we aggregate them to the ability type level, all firms on islands with the same ability type will choose ability type aggregate labor $l_t(\kappa)$ to maximize:

$$y_t(\kappa) = A_t k_t(\kappa)^\alpha l_t(\kappa)^{1-\alpha} \tag{34}$$

With these in hand, we can describe the size distribution of firms across ability types. Since the household was already represented by an economy-wide single representative agent, we can define the aggregated equilibrium concept, equilibrium in ability types:

Definition 1. *A recursive competitive equilibrium in ability types consists of a sequence of economy-wide prices* $\mathbb{P}_t \equiv (R_{t+i}, R_{bt+i}, W_{t+i}, Z_{t+i}, Q^i_{t+i})_{i=0}^\infty$ *, a sequence of type-specific prices* $\mathbb{P}_{\kappa t} \equiv (Q_{t+i}(\kappa))_{i=0}^\infty$*,*

a sequence of economy-wide quantities $\mathbb{Q}_t \equiv (K_{t+i}, C_{t+i}, I_{t+i}, Y_{t+i}, L_{t+i}, L_{t+i}, D_{t+i})_{i=0}^\infty$*, a sequence of type-specific quantities* $\mathbb{Q}_{\kappa t} \equiv (b_{t+i}(\kappa), s_{t+i}(\kappa), d_{t+i}(\kappa), i_{t+i}(\kappa))_{i=0}^\infty$ *such that:*

(Individual Optimization) for each t*,*

- $(d_t, b_t(\kappa), s_t(\kappa)$ *maximizes the representative bank's expected value (30) subject to their flow of funds constraint (3) for each ability type* κ
- $(k_t(\kappa), l_t(\kappa))$ *maximizes the representative firm's profits (34) for each ability type* κ
- (C_t, L_t, D_t) *maximizes the economy-wide representative household's expected utility (7)*
- I_t *maximizes capital goods producer profits (14)*

(Market Clearing) and for each t*, these markets clear*

- *deposits:* $D_t = \int_\kappa d_t(\kappa)p(\kappa)d\kappa$
- *labor:* $L_t = \int_\kappa l_t(\kappa)p(\kappa)d\kappa$

- *interbank loans:* $\int_\kappa b_t(\kappa)p(\kappa)d\kappa = 0$
- *new capital, so that* $I_t = \int_\kappa i_t(\kappa)p(\kappa)d\kappa$
- *assets (for each ability type):* $s_t(\kappa) = (1-\delta)k_t(\kappa) + i_t(\kappa)$

The full set of equilibrium equations is given in the appendix.

3.8 Discussion

There are several qualitative properties of the model that are worth discussing here, mainly because they are important for understanding the dynamics in the next section, but also because they illustrate the model's mechanics.

First, bank asset size increases with intermediation ability. To see this, first note that interbank activity increases with the intermediation ability of a bank. From proposition 1 above in the previous section, the higher a bank's intermediation ability, the more it borrows from the interbank lending market if it is a borrower, and the more it lends to the interbank lending market if it is a lender.

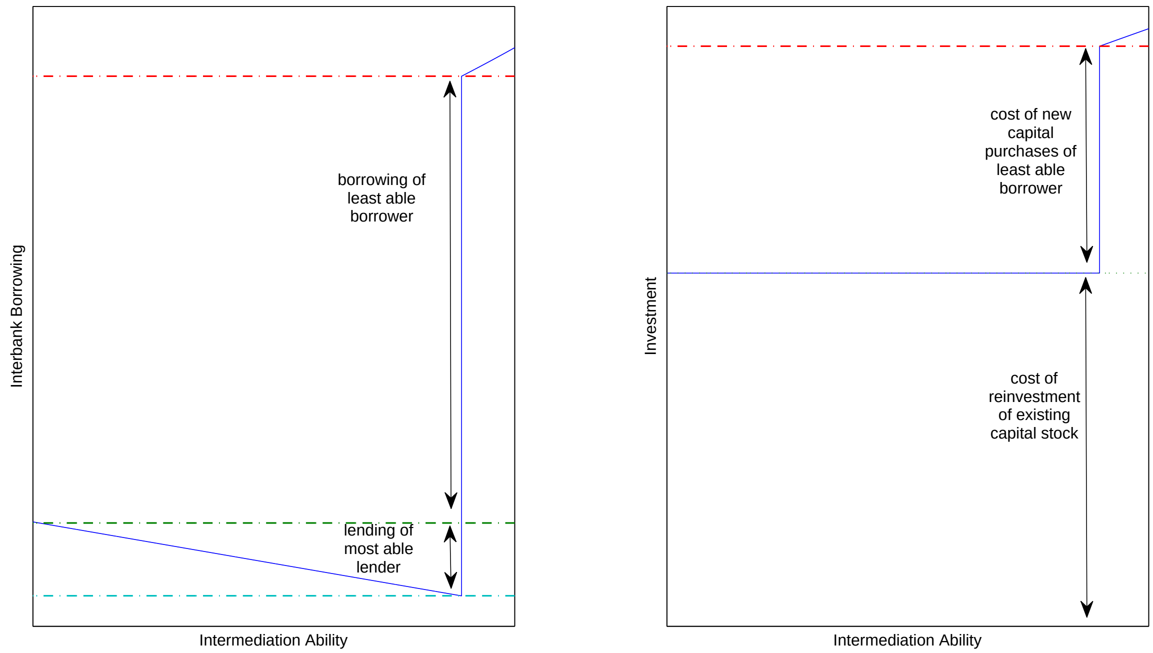
If a bank borrows from the interbank lending market, as it borrows more intensively, the liabilities side of their balance sheet becomes larger. Through the bank's "production function" (3), larger liabilities are transformed into larger assets. Banks which lend to the interbank market behave identically in terms of asset size. In the discussion in proposition 2, I note that all banks which lend hold exactly the same quantity of assets (equal to $Q_t^n k_t(\kappa)(1-\kappa)$). Thus, as intermediation ability increases, asset size either stays the same (when intermediation ability is below the cutoff) or increases (if intermediation ability is above the cutoff). Figure 5 diagrams this relationship.

Note here that borrowing banks are significantly larger than lending banks. The borrowing banks in this models are the real drivers of the economy in this model; they are responsible for the bulk of aggregate investment in the economy.

To further illustrate the mechanics of the model, let us consider how worsening financial frictions affect the interbank lending market in steady state. Recall that financial frictions are introduced into the model through the θ parameter in interbank borrowing constraint. The worse the financial friction, the higher the value of θ, the tighter the borrowing constraint for all banks.

If there were no financial friction in the model, no bank would be constrained in its borrowing. Depositors would make deposits in all banks as before, but once ability is realized, the interbank lending market would funnel all deposits to the most able bank.

Figure 5: Interbank Borrowing, Asset Size, and Intermediation Ability



Note: Interbank borrowing (left) and asset size (right) versus intermediation ability.

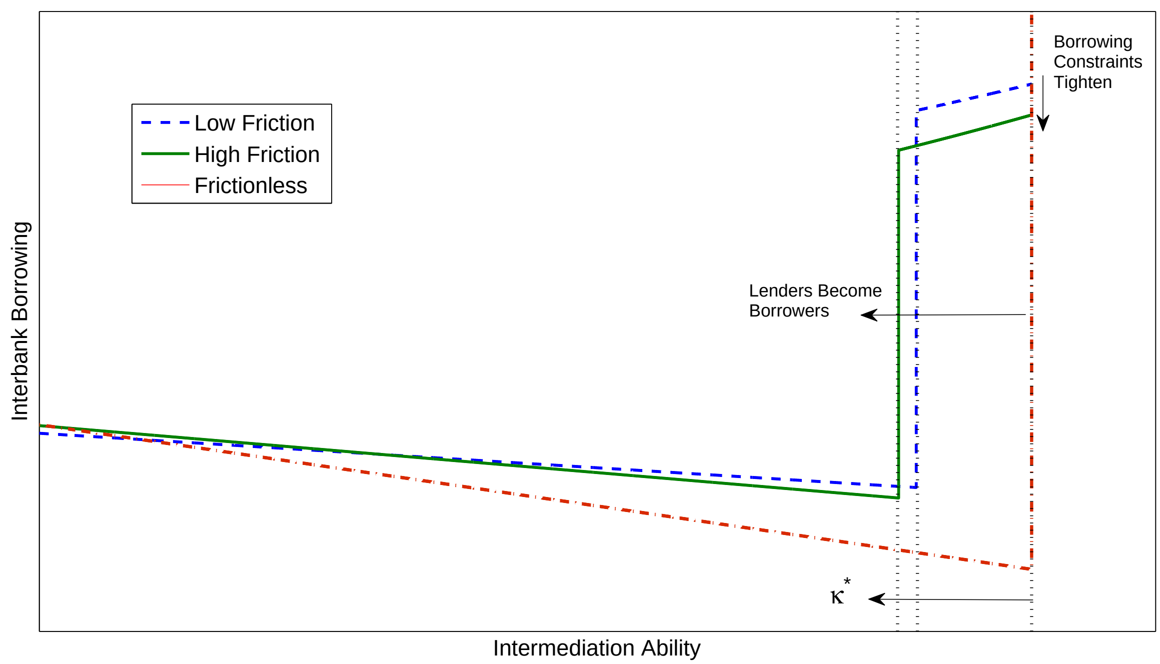
This bank would transform these loans into assets, which firms would use to purchase the highest amount of capital possible.

A positive value for the friction limits the maximum the most able bank can borrow through the interbank lending market. By limiting the maximum the bank can borrow, interbank loans that would go to the most able bank are redistributed to less able banks. All else equal, as shown in equation (23), less able banks face tighter borrowing constraints than more able ones. Thus, even though less able banks still invest in assets, because they can only borrow significantly less, economy-wide investment is lower than in the frictionless case.

An increase in the value of the friction pushes this redistribution further. We can decompose the effects of increasing the friction into two components: an "intensive" and an "extensive" effect. The intensive effect is a tightening of borrowing constraints: as the friction θ increases, the value of $\phi_t(a)$ decreases, so any borrower that continues to be a borrower will face a tighter constraint on the level of their borrowing, and thus will not be able to invest as much. The extensive effect is the pure redistribution of interbank loans to less able banks: as the friction increases and demand for interbank loans drops, the price of borrowing decreases (relative to the value of lending to firms). This decrease in price causes some banks that were lenders on the interbank market to become borrowers and invest.

Figure 6 diagrams the relationship between interbank borrowing and intermediation in steady state for different levels of financial friction. In the frictionless case (red dot-dash), only banks with the highest ability level borrow, while all other banks lend; this can be seen in the diagram as negative borrowing (lending) by banks with ability less than the maximum, and a straight vertical line at the maximum. The remaining lines diagram the same relationship for low friction (solid green) and high friction (dashed blue) cases. Increasing the friction results in an intensive effect, a decrease in the level of borrowing for all borrowing banks, and an extensive effect, an increase in borrowing by less able banks.

Figure 6: Borrowing and Financial Friction



Note: Steady state interbank borrowing in model with low friction (dash), high friction (solid), and frictionless (dash-dot) by banks of different abilities. Ability cutoff κ^* is largest for the frictionless case, then the low friction, then the high friction case.

The net effect of a given increase in the level of the financial friction on economy-wide investment depends on the level of the friction. The intensive effect tends to decrease aggregate investment because of the brake it puts on the borrowing taken by individual banks, while the extensive effect tends to increase aggregate investment as more banks are engaged in borrowing and investing (recall that borrowing banks are responsible for most of aggregate investment). We can establish that the intensive effect tends to be much larger than the extensive effect when the initial level of the friction is small, for example near the frictionless case. The opposite seems to be true when the initial value of the friction is large.

4 Quantitative Effects of Rising Concentration

In this section, I use the model to analyze the impact of changes to the asset size distribution that resulted from the increase in US banking sector concentration from 1992 to 2007. I calibrate the model to match the bank asset size distribution in 2007, and ask the counterfactual: if the banking sector were only as concentrated as it was in 1992, would downturns be less severe? If so, by how much?

Concentration is defined here as the share of assets (or more generally, output) held by the largest firms in an industry. An increase in concentration can happen for many reasons. For example, either an increase in the asset size of the largest bank in the industry or a decrease in the asset size of the smallest bank would cause the share of all assets held by the top 1 percent of banks to increase.

The concentration increase in the US banking sector took a specific form. First, the variance of bank asset sizes increased. Second, the skewness of asset sizes (asymmetry of the distribution toward higher values) increased. These two changes together will necessarily lead to an increase in concentration; an increase in variance puts more weight in the tails of the distribution, while an increase in skewness will ensure that the right tail gets more of the extra weight than the left tail. In what follows, I produce an increase in concentration in the model just as we see in the data, through an increase in the variance and skewness of the size distribution.

4.1 Calibration

The model requires the choice of nine parameters, the adjustment costs function, and the distribution of ability types. Five of the parameters control the standard preference and technology shocks from from the literature: the discount rate β, the habit parameter γ, the utility weight of labor χ, the share of capital in production α, and the depreciation rate δ. These parameters are drawn from Gertler and Kiyotaki (2011) and Christiano et al. (2005) and are given in Table 3 below.

The parameter φ, the inverse elasticity of labor supply, is chosen so that the Frisch elasticity is 10. This choice is made to induce realistic labor responses in a model with no other labor market frictions. The parameter ξ - the start-up transfer to new bankers - governs the average spread between the interbank lending rate and the average return on assets. The parameter σ, the exogenous probability of exit by a bank, is chosen so that the average bank survives for approximately 15 years. The parameter θ, the financial friction parameter, is chosen (along with the intermediation ability distribution) so that the average leverage ratio across all banks in the model matches the data.

Adjustment costs take the quadratic form:

$$f\left(\frac{I_t}{I_{t-1}}\right) = \frac{c_I}{2}(\frac{I_t}{I_{t-1}} - 1)^2 \tag{35}$$

The distribution of ability types can take any shape, but will always be bounded. In what follows, the distribution of ability types will take the form of a bounded Pareto distribution.[13] This distribution has a pdf, given by

$$p(x) = \frac{\rho \underline{\kappa}^{\rho} \overline{\kappa}^{\rho}}{\overline{\kappa}^{\rho} - \underline{\kappa}^{\rho}} x^{-\rho-1} \tag{36}$$

where $\underline{\kappa}$ and $\overline{\kappa}$ are the lower and upper bounds of the distribution. In accordance with the assumptions in the model, the parameters of the distribution will always be chosen so that the distribution has a mean of 1.

In accordance with the assumptions in the model, the parameters of the distribution will always be chosen so that the distribution has a mean of 1. The level of θ, the bounds of the distribution $\underline{\kappa}$ and $\overline{\kappa}$, and the shape parameter ρ will be adjusted to match the size concentration (measured by the share of assets held by the top 10 percent of banks) and the average leverage ratio in 1992 and 2007.

Table 4 shows how the calibrated model performs. Concentration is measured in the first column, by the share of assets held by the top 10 percent of banks. Both model rows show concentration measures for the steady state asset size distribution of the model calibrated to the 1992 and 2007 banking sectors. The model generates significantly less concentration than we see in the data, but significantly more leverage. This trade-off exists for a wide range of parameters.

[13] Janicki and Prescott (2006) suggests that the tail of the bank size distribution is best modeled with a Pareto distribution, while the remainder is best modeled with a lognormal distribution.

Table 3: Calibration - Parameters

Parameter		Value	Target/Source
Inverse Elasticity of Labor Supply	φ	0.33	Gertler/Kiyotaki (2011)
Start-up Transfer	ξ	0.002	Gertler/Kiyotaki (2011)
Probability of Bank Exit	σ	0.982	Average Bank Age - 2007
Discount Factor	β	0.99	Gertler/Kiyotaki (2011)
Habit Parameter	γ	0.8	Christiano, Eichenbaum, Evans (2005)
Depreciation Rate	δ	0.025	Christiano, Eichenbaum, Evans (2005)
Effective Capital Share	α	0.36	Christiano, Eichenbaum, Evans (2005)
Utility Weight of Labor	χ	5.584	Gertler/Kiyotaki (2011)
Adjustment Cost Parameter	c_I	1.5	Gertler/Kiyotaki (2011)
Low Concentration Scenario			
Parameter		Value	Target/Source
Friction	θ	0.3	Average Leverage Ratio - 1992
Min Ability	$\underline{\kappa}$	0.81	Asset Size Distribution - 1992
Max Ability	$\overline{\kappa}$	2.15	Asset Size Distribution - 1992
Shape	ρ	7	Asset Size Distribution - 1992
High Concentration Scenario			
Parameter		Value	Target/Source
Friction	θ	0.3	Average Leverage Ratio - 2007
Min Ability	$\underline{\kappa}$	0.835	Asset Size Distribution - 2007
Max Ability	$\overline{\kappa}$	2.8	mean(productivity distribution) = 1
Shape	ρ	6	Asset Size Distribution - 2007

Note: Parameters chosen for quantitative analysis. Low and High Concentration scenario parameters chosen to match characteristics of the asset size distribution in 1992 and 2007, respectively.

Table 4: Calibration - Performance

	Top 10% Share	Coeff of Variation	Avg Leverage
Model - 1992 (Low Concentration)	36%	1.1	6.6
Data - 1992	81%	7.65	6.6
Model - 2007 (High Concentration)	58%	2.1	14
Data - 2007	91%	15.6	6.6

Note: Performance of calibrated model. Model - 1992 and Model - 2007 rows show concentration measures for steady state asset size distribution of model calibrated to each of those years, respectively. Top 10% measures concentration as the percentage share of all assets held by top 10% of banks.

4.2 Crisis Response

The 2008 financial crisis, and banking crises more broadly, have their roots in more fundamental weaknesses in the wider economy. I want to focus on how the banking sector propagates and amplifies weaknesses. To this end, I initiate a downturn in the calibrated steady state economies with an exogenous shock to capital quality. This shock represents an unexpected fall in the underlying value of all assets in the economy. (With respect to the 2008 recession, this shock could capture the unexpected fall in underlying collateral values of mortgage-backed securities.) Since banks in the model derive their net worth partly from the capital investments they've already made, this shock to capital translates to a shock to the net worth of all banks.

Specifically, I first calculate the non-stochastic steady state of the model, and then use the software package Dynare[14] to calculate a second-order approximation of the model around the steady state. I then initiate a crisis in the model with a 5 percent shock to the capital quality ψ_t - the level is chosen to initiate a drop in output similar to that seen during the 2008 crisis. The shock has a persistence of 0.8, so that it returns to its steady state value after five years. I calculate the impulse responses to the shock in the approximated model, again using Dynare.

The impulse responses for these cases are given in figure 7. The high concentration response is the dotted line, while the low concentration response is the solid line. Notably, output decreases more, in percentage terms, in the high concentration economy. Output in the high concentration economy falls to a minimum that is 1.66 times the minimum of

[14] For more details regarding this software, see ?.

the low concentration economy; thus, if the high concentration case represents the economy just before the financial crisis, the economy would have experienced a downturn that was only 60 percent as deep.

In both scenarios, the initial fall in capital quality produces the same qualitative changes in the model. First, the shock destroys capital on impact, immediately reducing the stock of assets, K, invested in the economy. We can see this in the panel for capital - in both cases, there is an initial drop in the first period. Output Y falls, and because the marginal productivity of labor falls upon impact, demand for labor falls. Wages fall as well, so the labor supply curve shifts inward, and employment falls. The decrease in wages and employment forces households to save less today, leading to lower supply of deposits in both cases.

Second, on the bank side, because capital is held as assets by banks, the decline in capital quality leads to a significant decrease in bank net worth. This is seen in the bottom left panel. A decline in capital quality decreases net worth in two ways. First, a decline in capital quality destroys the value of assets held by the bank from last period, which directly decreases net worth. Second, the decline in capital quality reduces the price of capital, which indirectly decreases the value of the bank's assets.

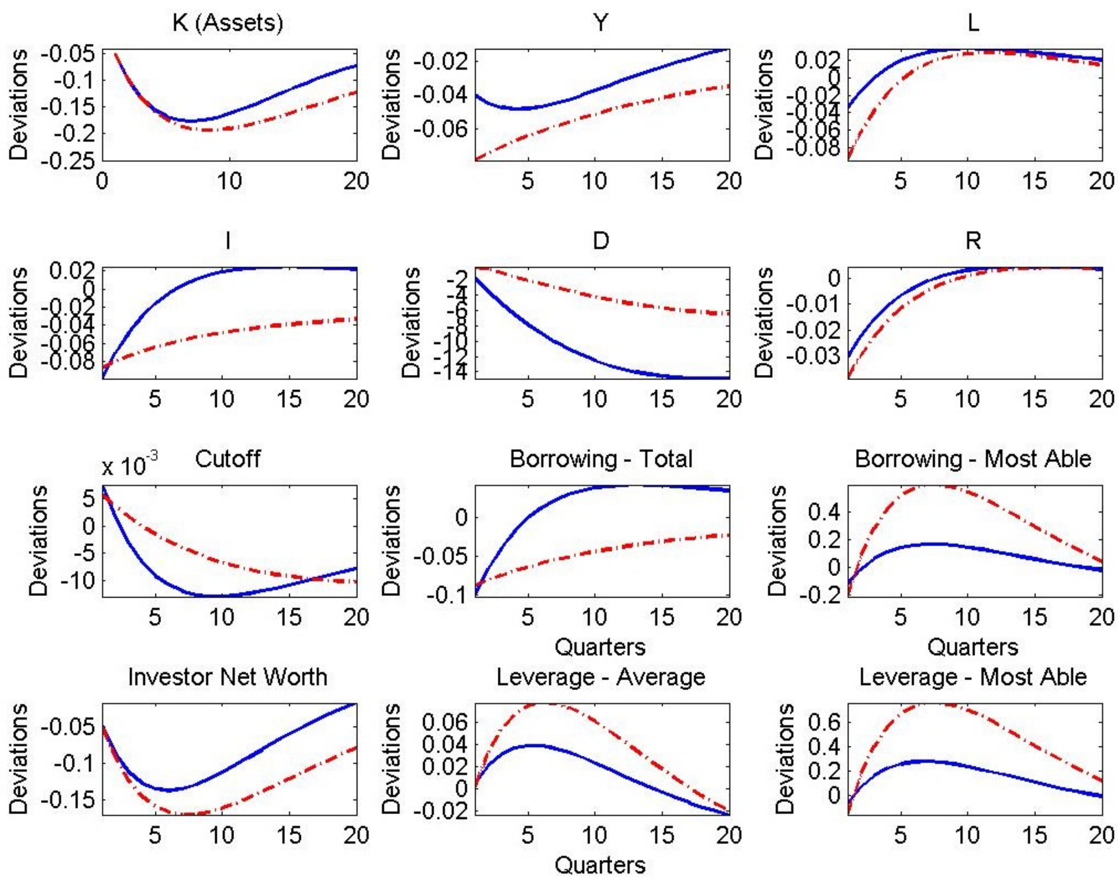
Because their net worth has decreased, borrowing banks must reduce their asset positions further to satisfy their borrowing constraint. As a result, investment I falls - this is depicted in the panel labeled I. In addition, adjustment costs help keep investment low for many periods after the initial shock.

In addition, because their net worth has decreased and because households supply fewer deposits, the supply curve of interbank loans shifts inward, and the quantity of loans made decreases. However, the interest rate on deposits and interbank loans, R, also falls. Banks are responsible for this, preventing households from dropping the supply of deposits even further than it was already.

Though banks de-lever and reduce their assets after their net worth decreases, the leverage ratio for every bank increases in a downturn, that is, leverage ratios are countercyclical. This is shown in the panel labeled average leverage. Leverage ratios are countercyclical because the value from holding assets increases more than the value of default on interbank loans; since capital quality shocks induce de-leveraging, the value from holding assets increases as the shock dissipates.

Dispersion can amplify the effects of financial frictions. The banking sector of the high concentration economy is more dispersed than the banking sector of the low concentration economy. Because more able banks are more leveraged, and because leverage increases

Figure 7: Impulse Responses: High and Low Concentration



Note: Impulse responses to a negative 5 percent financial shock for high concentration (dotted, red) and low concentration (solid, blue) banking sector. Responses are expressed in percentage deviations. Responses are calculated for a second-order approximation to the model using the software package Dynare.

at an increasing rate with intermediation ability, average leverage gets larger as dispersion increases. A larger portion of steady state investment is funded through interbank borrowing; if the net worth decreased by the same amount in both the heterogeneous and homogeneous bank models, investment decreases more in the heterogeneous case. In the figure, we actually see that net worth decreases more in the heterogeneous case, exacerbating this difference.

The decrease in the interbank lending rate causes less able banks to switch from lending to borrowing on the interbank lending market. This corresponds to a decrease in the ability cutoff, shown in the panel labeled cutoff. These new borrowers act as a stabilizing force in the interbank lending market, pushing up the interest rate and the volume of interbank lending above what it would be otherwise. These new borrowers are also less leveraged when they do borrow. Their entrance brings down the average leverage in the economy.

Heterogeneity among lenders also affects impulse responses. When faced with the same shock to net worth, more able lenders decrease their lending more than less able ones, making the average lender response in percentage terms larger in the heterogeneous banks case. Thus, the volume of interbank lending decreases by more in the homogeneous banks case.

Second, output in the high concentration economy returns to steady state much more slowly than in the low concentration economy. Though not pictured, output in the high concentration economy takes twice as many quarters to return to its steady state level than in the low concentration economy.

The difference on impact of output in the two economies is because of the initial difference in the response of employment. As shown in the top right panel, employment falls much farther in the high concentration than the low concentration economy. Since the high concentration economy is more dispersed, all banks know there is a high chance of being very productive tomorrow. This makes the value of net worth in the high concentration economy higher on average today than in the low concentration economy. Thus, as capital gets scarce, even though the marginal product of labor decreases in both economies, firms want capital more in the high concentration economy, and shift their capital/labor ratio toward capital more than in the low concentration economy. As a result, employment falls further initially, and output falls with it.

Net worth decreases more in the high concentration case because the net profits from firms decrease more, which occurs because output Y falls more. Thus, the relationship between investment today and net worth tomorrow produces a feedback effect that amplifies the shock.

The average leverage of borrowing banks is higher. These effects tend to worsen the de-leveraging borrowing banks do in response to a shock, as the more concentrated sector contains very leveraged banks.

In addition to dispersion, the mass of the right tail has increased over time. This has two effects. First, since the number of leveraged banks increases, more banks de-leverage, and the overall effect of de-leveraging on output worsens. Second, since the number of leveraged banks increases, the same changes to the interbank lending rate produce larger swings in the ability cutoff - for a given change in interest rate, there are more banks that respond to the change.

These leverage differences are responsible for the very slow recovery of the high concentration economy. Though both economies initially drop by similar amounts, total borrowing, investor net worth, and investment in the high concentration economy stays much lower than in the high concentration economy after a year or so. This occurs despite the fact that leverage, both of the average bank and the most able, is more strongly countercyclical in the high concentration than the low concentration economy.

Another contributing factor to the deeper downturn in the high concentration economy is the smaller extensive effect of banks switching from lending to borrowing. The net worth of banks in the high concentration economy decreases less than in the low concentration economy, while the volume of interbank borrowing decreases more. Both the supply and demand for interbank loans decreases. At first, this restricts demand, but because supply falls farther, there is a net positive effect on the value that banks can receive from interbank lending in both cases. This negative effect drives more banks into switching from lending to borrowing, pushing down the cutoff in the low concentration economy farther, and dampening the negative effects of de-leveraging relative to the high concentration economy.

As shown above, the banks that switch from lending to borrowing stabilize the interbank lending market, partly offsetting the negative effects of deleveraging - the extensive effect of switching opposes the intensive effect of deleveraging. As a result, even though amplification occurred for this parameterization, there are other parameterizations for which bank heterogeneity dampens downturns. Generally, amplification will occur for cases where cutoffs are already relatively low. The cutoff will not move much after a capital shock, and thus extensive switching effects are small.

5 Conclusion

In this paper, I construct a macroeconomic model with a heterogeneous banking sector, and show that heterogeneity has consequences for downturns. In particular, mean-independent changes in the distribution of bank sizes can mirror the effects of financial frictions, and as a result, changes to the bank size distribution qualitatively similar to those that occurred with the rising banking sector concentration observed in the US make potential downturns more severe today.

The model in this paper puts us in the unique position to consider targeted macroprudential policies in a general macroeconomic framework. One such policy taken up by the US Treasury and the Federal Reserve was embodied in broad targeted asset purchases like the TARP, which in some cases equated to the government taking up partial ownership positions in some banks. In the context of this model, this would lessen the borrowing constraint target banks face and essentially reduce the financial friction for the whole economy. If the economy is in really bad shape, however, the economy starts at a high level of friction, extensive effects can dominate. Loosening the friction will increase the interbank interest rate, drawing some banks into interbank lending. However, the banks to which this lending goes to are still so constrained that they don't expand their investment enough. They get an extra dollar of loan, but at a higher interest rate not offset by the change in the friction. In this case, offering a targeted program which doesn't affect the interbank market would improve welfare more than one which did affect the interbank market.

Despite the stylized assumptions in the model, the environment is rich, and I see three avenues worth pursuing in future work. First, the model has difficulty generating the extreme bank size inequality we see in the data. This problem is partly due to the sharp transition from lending to borrowing banks make as their ability increases. If banks calculated their net worth as a strictly convex, rather than linear, combination of deposits, assets and interbank loans, their value functions would also be convex in these three quantities, and banks would smoothly transition from lending to borrowing as ability increased. With this, we could more easily generate high inequality.

Second, dropping assumption 1 will cause bank size next period to depend on its size this period. This will induce an endogenously evolving size distribution, which may itself uncover some interesting dynamics. For example, if ability tomorrow doesn't depend on ability today, but size carries to next period, large banks can use their assets as a buffer against poor ability draws, but small banks would be more sensitive to such changes. As a result, the average value of interbank lending would not adjust as much as in the case of this model.

Third, if we modify the model by restricting the loans banks make to a subset of all banks, we can create an environment where network effects are important. Relationships matter in interbank lending (and other short-term debt, e.g. repo), with banks borrowing from a partner in one period only to lend to the same partner in the next. Bank A's failure will matter more for bank B if the two were lending partners. As a result, some network structures of lending relationships will transmit downturns better than others, just as a more concentrated banking sector seems to transmit crises better in this model.

This project is a small step in a larger agenda of analyzing the macroeconomic implications of industry-wide trends (not just size differences) in banking. So far, research in this field has not placed much emphasis on the characteristics of individual banks and the dynamic consequences from changes in those characteristics. Research in this vein may be especially informative for the conduct of unconventional monetary policy.

A Bank Value Function Derivation

In the main text, I guess that we can solve the bank's optimization problem with a linear value function. Here I verify that this solution does indeed solve the bank's problem in detail.

First, recall the guess that in every period a bank guided by some linear function of shares, deposits, and interbank loans will maximize its expected net worth:

$$V_t = \nu_{st} s_t(a) - \nu_{bt} b_t(a) - \nu_t d_t \tag{37}$$

If the guess is correct, the coefficients ν_{st} and ν_{bt} will be equal to the marginal value of shares and interbank loans. Shares produce profits from firms tomorrow, which increases the value of tomorrow's net worth. Deposits and interbank loans today reduce tomorrow's net worth through repayment. Because net worth is a linear function of each of these quantities, this marginal value will be constant. Calculating the value of these coefficients amounts to calculating the marginal effect of increasing these quantities on net worth.

Note that the above Lagrangian is also equal to the bank's maximized value written in terms of net worth. Substituting the FOCs into the Lagrangian, we get

$$V_t = (1 + \lambda_t(a))\nu_{bt} n_t(a) \tag{38}$$

If we iterate this one period forward and then plug this into the Bellman equation, we obtain:

$$\nu_{st}s_t(a) - \nu_t d_t - \nu_{bt}b_t(a) = E_{t,a'}\Lambda_{t,t+1}(1-\sigma)n_{t+1}(a') + \sigma(1+\lambda_{t+1}(a'))\nu_{bt+1}n_{t+1}(a') \quad (39)$$

We can obtain the value of each coefficient by taking the partial derivatives of both sides of the above equation with respect to each of the variables $b_t(a)$, d_t, $s_t(a)$:

$$\nu_{bt} = R_{bt}E_{t,a'}\Lambda_{t,t+1}\Omega^{a'}_{t+1} \quad (40)$$

$$\nu_t = R_t E_{t,a'}\Lambda_{t,t+1}\Omega^{a'}_{t+1} \quad (41)$$

$$\nu_{st} = E_{t,a'}\Lambda_{t,t+1}\psi_{t+1}(\Omega^{a'}_{t+1}(Z_{t+1} + (1-\delta)Q^{a'}_{t+1}) \quad (42)$$

where

$$\Omega^{a'}_{t+1} = 1 - \sigma + \sigma(1 + \lambda_{t+1}(a'))\nu_{bt+1} \quad (43)$$

As long as ability draws next period are independent of the draw this period, these coefficients do not depend on ability type or level of either of the choice variables. They also maximize the value of the original objective function (4):

$$V_t = E_{t,a'}\sum_{i=1}^{\infty}(1-\sigma)\sigma^{i-1}\Lambda_{t,t+i}n_{t+i}(a') \quad (44)$$

To see this, consider the effect of a decrease d_t on equation (4). d_t only affects the bank objective when the bank exits. A fraction $(1-\sigma)$ of banks exit from each island every period. An increase in d_t affects net worth in period $t+1$ directly, by increasing the repayment in that period. Thus, banks that exit in period $t+1$ will have lower net worth. The decrease in $t+1$ net worth decreases s_{t+1} through the flow of funds constraint, which turn reduces the net worth in period $t+2$; if banks exit then, they will also have lower net worth. The full effect of the change in deposits can then be written as

$$\begin{aligned}\frac{dV_t}{dd_t} = (1-\sigma)\frac{\partial n_{t+1}}{\partial d_t} + \sigma(1-\sigma)\frac{\partial n_{t+1}}{\partial d_t}\frac{\partial s_{t+1}}{\partial n_{t+1}}\frac{\partial n_{t+2}}{\partial s_{t+1}} \\ + \sigma^2(1-\sigma)\frac{\partial n_{t+1}}{\partial d_t}\frac{\partial s_{t+1}}{\partial n_{t+1}}\frac{\partial n_{t+2}}{\partial s_{t+1}}\frac{\partial s_{t+2}}{\partial n_{t+2}}\frac{\partial n_{t+3}}{\partial s_{t+2}} + ... \end{aligned} \quad (45)$$

The partial effects $\frac{\partial s_{t+i}}{\partial n_{t+i}}$ are captured by the quantity $(1+\lambda_{t+i}(a'))\nu_{bt+i}$. This means that the coefficient ν_t completely summarizes the direct effect of d_t on the island objective function. The same argument holds for $b_t(a)$. Because of this, we can conclude that

$$\frac{dV_t}{dd_t} = \nu_t \tag{46}$$

$$\frac{dV_t}{db_t} = -\nu_{bt} = \nu_t \tag{47}$$

Thus, if we solved the bank's optimization problem by taking first-order conditions directly on the objective function V_t with respect to each of the bank's choice variables, we would obtain the same first-order conditions as we obtain from the linear guess. Moreover, given that the marginal effects are constant, we can say that:

Proposition 3. *A set of choices $(d^*_{t+i}, b^*_{t+i}(a))_{i=0}^{\infty}$ that maximizes the linear value function guess (37) for each i will also maximize the bank objective (44).*

B Equilibrium Conditions

In this section, I list the full set of equations that describes equilibrium. Before proceeding, I make a few simplifications to the model. First, let's simplify the interbank lending market clearing condition. Note that clearing the interbank market automatically clears the market for deposits. Call B the set of ability types with representative banks that borrow, and L the set of all types that lend. The sum of all interbank lending by types in L should equal the sum of all interbank borrowing by types in B:

$$\begin{aligned}\int_{\kappa\in B} b(\kappa)p(\kappa)d\kappa &= \int_{\kappa\in B} \left(\phi_t(\kappa)n_t(\kappa) - d_t(\kappa)\right)p(\kappa)d\kappa \\ &= -\int_{\kappa\in L} b(\kappa)p(\kappa)d\kappa = \int_{\kappa\in L} \left(n_t(\kappa) + d_t(\kappa) - Q_t(\kappa)k_t(\kappa)\right)p(\kappa)d\kappa \end{aligned} \tag{48}$$

With this in hand, we can rewrite the interbank lending market condition in terms of aggregates:

$$\begin{aligned} D_t = &\int_{\kappa^*}^{\overline{\kappa}} p(\kappa)\phi_t(\kappa)((Z_t + (1-\delta)Q^i_t)K_t(\sigma+\xi) - \sigma R_{t-1}D_{t-1})d\kappa \\ &- \int_{\underline{\kappa}}^{\kappa^*} p(\kappa)((Z_t + (1-\delta)Q_t(\kappa))K_t(\sigma+\xi) - \sigma R_{t-1}D_{t-1})d\kappa \\ &+ \int_{\underline{\kappa}}^{\kappa^*} p(\kappa)\frac{Q_t(\kappa)}{\kappa}(1-\delta)K_t d\kappa \end{aligned} \tag{49}$$

Where D_t is aggregate deposits and K_t is the aggregate capital stock.

In order to pin down investment, we need to consider the market for new capital alone. We know that the aggregate investment and existing capital held by firms on borrower islands is equal to $I_t + (1-\delta)K_t \int_{\kappa^*}^{\overline{\kappa}} p(\kappa)$, and that capital is demanded in the form of interbank borrowing and deposits. Noting that borrowing islands will borrow the maximum possible, we can replace the borrowing constraint into the flow of funds equation:

$$\begin{aligned} s_t(\kappa) = i_t(\kappa) + k_t(\kappa) = \frac{\kappa}{Q_t(\kappa)}(n_t(\kappa) + d_t(\kappa) + b_t(\kappa)) \\ = \frac{\kappa}{Q_t^i}(1 + \phi_t(\kappa))n_t(\kappa) \ \forall \kappa \in [\kappa^*, \overline{\kappa}] \end{aligned} \tag{50}$$

Finally, integrate both sides over the set B to get the condition in terms of aggregates:

$$Q_t^i \left(I_t + (1-\delta)K_t \int_{\kappa^*}^{\overline{\kappa}} p(\kappa) \right) = \int_{\kappa^*}^{\overline{\kappa}} \kappa(1 + \phi_t(\kappa))n_t(\kappa) d\kappa \tag{51}$$

B.1 Equilibrium Equations

HH/Firms

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{52}$$

$$K_t = \psi_t(I_{t-1} + (1-\delta)K_{t-1}) \tag{53}$$

$$Y_t = C_t + (1 + f(\frac{I_t}{I_{t-1}}))I_t \tag{54}$$

$$1 = E_t \Lambda_{t,t+1} R_t \tag{55}$$

$$u_{Ct} = (C_t - \gamma C_{t-1})^{-1} - \beta\gamma(C_{t+1} - \gamma C_t)^{-1} \tag{56}$$

$$\Lambda_{t,t+1} = \beta \frac{u_{Ct+1}}{u_{Ct}} \tag{57}$$

$$Z_t = \alpha A_t (\frac{L_t}{K_t})^{1-\alpha} \tag{58}$$

$$Q_t^i = 1 + f(\frac{I_t}{I_{t-1}}) + (\frac{I_t}{I_{t-1}}) f'(\frac{I_t}{I_{t-1}}) - E_t \Lambda_{t,t+1} (\frac{I_{t+1}}{I_t})^2 f'(\frac{I_{t+1}}{I_t}) \tag{59}$$

$$\chi L_t^{\varphi} = (1-\alpha)\frac{Y_t}{L_t} E_t u_{Ct} \tag{60}$$

To drop habit formation from the model, set $\gamma = 0$. To drop adjustment costs from the model, set $f(\frac{I_t}{I_{t-1}}) = 0$ everywhere.

Exogenous Shock Processes

$$A_t = \rho_A A_{t-1} + \epsilon_{At} \tag{61}$$

$$\psi_t = \rho_\psi \psi_{t-1} + \epsilon_{\psi t} \tag{62}$$

Bank Optimization

$$Q_t^n = \frac{\nu_{st}}{\nu_{bt}} \tag{63}$$

$$Q_t(\kappa) = \kappa Q_t^n \ \ \forall \kappa \in L \tag{64}$$

$$\nu_t = \nu_{bt} \tag{65}$$

$$\lambda_t(\kappa) = \frac{(\kappa \nu_{st} - \nu_{bt} Q_t^i)}{\kappa \theta Q_t^i - (\kappa \nu_{st} - \nu_{bt} Q_t^i)} \ \ \forall \kappa \in B \tag{66}$$

$$\lambda_t(\kappa) = 0 \ \ \forall \kappa \in L \tag{67}$$

$$\Omega_t(\kappa) = 1 - \sigma + \sigma \nu_{bt}(1 + \lambda_t(\kappa)) \tag{68}$$

$$\phi_t(\kappa) = \frac{\kappa(\nu_{st} - \theta Q_t^i)}{\kappa \theta Q_t^i - (\kappa \nu_{st} - \nu_{bt} Q_t^i)} \ \ \forall \kappa \in B \tag{69}$$

$$\tag{70}$$

$$\nu_{bt} = E_t R_t \Lambda_{t,t+1}(1 - \sigma + \sigma \nu_{bt+1}) + E_t R_t \Lambda_{t,t+1} \sigma \nu_{bt+1} G_0 \tag{71}$$

$$\begin{aligned}\nu_{st} = E_t \Lambda_{t,t+1} \Psi_{t+1}[&(1 - \sigma + \sigma \nu_{bt+1} + \sigma \nu_{bt+1} G_0) Z_{t+1} \\ &+ (1 - \sigma + \sigma \nu_{bt+1})(1 - \delta) Q_{t+1}^i (\int_{\kappa_{t+1}^c}^{\overline{\kappa}} p(\kappa) d\kappa) + \sigma \nu_{bt+1} G_0 (1 - \delta) Q_{t+1}^i \\ &+ (1 - \sigma + \sigma \nu_{bt+1})(\int_{\underline{\kappa}}^{\kappa_{t+1}^c} p(\kappa) \kappa d\kappa)(1 - \delta) Q_{t+1}^n]\end{aligned} \tag{72}$$

$$G_0 = \int_{\kappa_{t+1}^c}^{\overline{\kappa}} p(\kappa) \lambda_{t+1}(\kappa) \, d\kappa \tag{73}$$

Securities Market

$$Q_t^i(I_t + (1-\delta)K_t \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa)d\kappa) = ((Z_t + (1-\delta)Q_t^i)(\sigma+\xi)K_t - \sigma R_{t-1}D_{t-1})G_1 \tag{74}$$

$$G_1 = \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa)\kappa(1+\phi_t(\kappa))\, d\kappa \tag{75}$$

Deposit Market

$$\begin{aligned} D_t = (Z_t + (1-\delta)Q_t^i)(\sigma+\xi)K_t - \sigma R_{t-1}D_{t-1})G_2 \\ - (Z_t(\sigma+\xi) - \sigma R_{t-1}D_{t-1})(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)d\kappa) \\ - (1-\delta)Q_t^n(\sigma+\xi)K_t(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)\kappa d\kappa) + Q_t^n(1-\delta)K_t(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)d\kappa) \end{aligned} \tag{76}$$

$$G_2 = \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa)\phi_t(\kappa)\, d\kappa \tag{77}$$

B.2 Steady State

HH

$$I = \delta K \tag{78}$$

$$C = [A(\frac{L}{K})^{1-\alpha} - \delta]K \tag{79}$$

$$\chi L^{\varphi} = (1-\alpha)A(\frac{L}{K})^{-\alpha}\frac{1-\beta\gamma}{1-\gamma}\frac{1}{C} \tag{80}$$

$$L = (\frac{Z}{\alpha A})^{\frac{1}{1-\alpha}}K \tag{81}$$

$$R = \frac{1}{\beta} \tag{82}$$

$$\Lambda = \beta \tag{83}$$

$$Q^i = 1 \tag{84}$$

Bank Optimization

$$Q^n = \frac{\nu_s}{\nu_b} \tag{85}$$

$$\lambda(\kappa) = \frac{(\kappa\nu_s - \nu_b)}{\kappa\theta - (\kappa\nu_s - \nu_b)} \ \ \forall\kappa \in B \tag{86}$$

$$\phi(\kappa) = \frac{\kappa(\nu_s - \theta)}{\kappa\theta - (\kappa\nu_s - \nu_b)} \ \ \forall\kappa \in B \tag{87}$$

$$1 = \nu_b(1 - \frac{\sigma}{1-\sigma}\overline{G_0}) \tag{88}$$

$$Z = \frac{\nu_s}{\beta\nu_b} - \frac{1-\delta}{\nu_b}\left((1 - \sigma + \sigma\nu_b)(\int_{\kappa^c}^{\overline{\kappa}} p(\kappa)d\kappa) + \sigma\nu_b\overline{G_0}\right) - \frac{(1-\delta)Q^n}{\nu_b}\left((1 - \sigma + \sigma\nu_b)(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)\kappa d\kappa)\right) \tag{89}$$

Securities Market

$$\frac{\sigma}{\beta}\frac{D}{K} = (Z + 1 - \delta)(\sigma + \xi) - \frac{\delta + (1-\delta)(\int_{\kappa^c}^{\overline{\kappa}} p(\kappa)d\kappa)}{\overline{G_1}} \tag{90}$$

Deposit Market

$$\frac{N^i}{K} = (Z + 1 - \delta)(\sigma + \xi) - \frac{\sigma}{\beta}\frac{D}{K} \tag{91}$$

$$\frac{D}{K} = \frac{N^i}{K}\overline{G_2} + Q^n(1-\delta)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)d\kappa\right) - \left(Z(\sigma+\xi) - \frac{\sigma}{\beta}\frac{D}{K}\right)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)d\kappa\right) - (1-\delta)Q^n(\sigma+\xi)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)\kappa d\kappa\right) \tag{92}$$

$$G_0 = \int_{\kappa^c_{t+1}}^{\overline{\kappa}} \frac{p(\kappa)(\kappa\nu_{st+1} - \nu_{bt+1}Q^i_{t+1})}{\kappa\theta Q^i_{t+1} - (\kappa\nu_{st+1} - \nu_{bt+1}Q^i_{t+1})}d\kappa \tag{93}$$

$$G_1 = \int_{\kappa^c_t}^{\overline{\kappa}} \frac{p(\kappa)\kappa\nu_{bt}Q^i_t}{\kappa\theta Q^i_t - (\kappa\nu_{st} - \nu_{bt}Q^i_t)}d\kappa \tag{94}$$

$$G_2 = \int_{\kappa^c_t}^{\overline{\kappa}} \frac{p(\kappa)\kappa(\nu_{st} - \theta Q^i_t)}{\kappa\theta Q^i_t - (\kappa\nu_{st} - \nu_{bt}Q^i_t)}d\kappa \tag{95}$$

$$\overline{G_0} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)(\kappa\nu_s - \nu_b)}{\kappa\theta - (\kappa\nu_s - \nu_b)} d\kappa \tag{96}$$

$$\overline{G_1} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa\nu_b}{\kappa\theta - (\kappa\nu_s - \nu_b)} d\kappa \tag{97}$$

$$\overline{G_2} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa(\nu_s - \theta)}{\kappa\theta - (\kappa\nu_s - \nu_b)} d\kappa \tag{98}$$