The Dissertation Committee for Chacko George
certifies that this is the approved version of the following dissertation:

Three Essays in Macroeconomics

Committee:

Matthias Kehrig, Supervisor

Olivier Coibion

Andrew Glover

David Kendrick

Pablo D'Erasmo

Jacek Rothert

Three Essays in Macroeconomics

by

Chacko George, B.S.E.E., B.A, M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2014

Dedicated to everyone I have ever met, but mostly to Shila.

Acknowledgments

Thanks to...

my advisor, Matthias Kehrig, for working tirelessly to get me ready for job market, for showing me the value of engaging with other economists, and for showing me what it takes to put together a serious research project...

...to Oli Coibion, who helped me to focus on what was important, to get real about what I was doing, and to put my ideas into perspective...

...to Andy Glover, who encouraged me to think more rigorously about my research...

...to David Kendrick and Jacek Rothert, who helped me discover an interest in financial regulation...

...to Pablo D'Erasmo for guiding my research through his own work, and for serving on my committee...

...to Dean Corbae, Tom Wiseman and Casey Mulligan, for introducing me to the world of professional economics...

...to the University of Texas for shaping my mind for a decade...

Thanks to my mother for always making me feel cared for, loved, and special...

Thanks to my father for always being interested in my work, for showing me the value of research, for helping me see how satisfying it is to be curious about the world, and for always showing me that the path taken by others is not necessarily the best one...

To my sister Maria for advising me as an adult almost as much as I advised her when we were children...

To my uncle Thomas, my aunt Prabha, Chandu and Vichu, for helping me feel like Austin was my home, and always making high achievement seem possible...

To Florian for turning me on to an interesting idea that I hope to work on further, and guiding me through my first co-authoring experience...

To my many friends in the UT econ department for making grad school seem a little less like punishment for past sins...

To my wife Shila for putting up with me through this process, for her daily love and support, and for showing me the value of diligence and hard work by her own example.

Three Essays in Macroeconomics

Publication No. ______________

Chacko George, Ph.D.
The University of Texas at Austin, 2014

Supervisor: Matthias Kehrig

This dissertation consists of three essays on topics in macroeconomics. In the first chapter, I construct a macroeconomic model with a heterogeneous banking sector and an interbank lending market. 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 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. I use the model to evaluate the impact of increasing concentration in US banking on the severity of potential downturns. I find that if the banking sector in 2007 was only as concentrated as it was in 1992, GDP during the Great Recession would have declined by 40% less it did, and would have recovered twice as fast.

In the second chapter, my co-author and I investigate the impact of firm capacity constraints on aggregate production and productivity when the economy is driven by aggregate and idiosyncratic demand shocks. We are motivated by three observed regularities in US GDP: business cycles are asymmetric, in that large absolute changes in output are more likely to be negative than positive; capacity and capital utilization are procyclical, and increase the procyclicality of measured productivity; the dispersion of firm productivity increases in recessions.

We devise a model of demand shocks and endogenous capacity constraints that is qualitatively consistent with these observations. We then calibrate the model to aggregate utilization data using standard Bayesian techniques. Quantitatively, we find that the calibrated model also exhibits significant asymmetry in output, on the order of the regularities observed in GDP.

The third chapter explores the role of distance in equilibrium selection. I consider a model economy with multiple steady state equilibria where a high productivity and a low productivity technology are available for use in production. The high productivity technology requires a fixed set up cost for production. Sectors are linked by localized production complementarities. I consider selection under a learning rule in which agents imitate their most successful neighbor. As distance between neighbors decreases, the possible profits from industrialization increase, and the likelihood that the learning rule process converges to a steady state matching the H equilibrium increases. The result suggests that, in the presence of localized technology spillovers, there may be important gains to economic growth from infrastructure development.

Table of Contents

Acknowledgments **v**

Abstract **vii**

List of Tables **xii**

List of Figures **xiii**

Chapter 1. Bank Size, Leverage, and Financial Downturns **1**

1.1 Introduction 2
1.2 Stylized Facts 7
1.3 Model 13
1.3.1 Banks 14
Additional Assumptions 20
1.3.2 Households 21
1.3.3 Firms 23
1.3.4 Capital Goods Producers 24
1.3.5 Other Conditions 25
1.3.6 Bank Solution 26
1.3.7 Solution Properties 30
1.3.8 Aggregating to Ability Types 36
1.3.9 Steady State Properties 39
1.4 Homogeneous and Heterogeneous Banking Sector Comparison 45
1.4.1 Calibration 45
1.4.2 Crisis Response 47
1.5 Quantitative Effects of Rising Concentration 53
1.6 Conclusion 58

Chapter 2. Aggregate Implications of Capacity Constraints under Demand Shocks (with Florian Kuhn) 61

2.1 Introduction . . . 62
2.2 Three stylized facts relating to asymmetry of business cycles . . . 63
2.3 Model . . . 68
2.3.1 Final Goods Firms . . . 69
2.3.2 Intermediate Goods Firms . . . 70
Firm Costs and Factor Demands . . . 72
Firm Supply . . . 72
Firm Expected Profit . . . 74
2.3.3 Solution . . . 74
2.4 Estimation . . . 76
2.4.1 Data . . . 77
2.4.2 Estimation Procedure . . . 78
2.4.3 Results . . . 79
2.5 Conclusion . . . 82

Chapter 3. Technology Spillovers and Local Interactions 84

3.1 Introduction . . . 84
3.2 Model - Stage Game . . . 89
3.2.1 Firms . . . 90
3.2.2 Households . . . 93
3.3 Solution Properties . . . 93
3.3.1 Existence of Multiple Equilibria . . . 98
3.3.2 Neighborhood Size and Profits . . . 101
3.4 Imitation and Dynamics . . . 103
3.4.1 Imitation as Selection Process . . . 104
3.4.2 Nearest Neighbors Criterion . . . 107
3.4.3 Effects of Increasing Neighborhood Size . . . 110
3.5 Conclusions . . . 111
3.5.1 Future Work . . . 111
3.5.2 Empirical Exercises . . . 113

Appendices 115

Appendix A. Bank Size, Leverage, and Financial Downturns **116**
A.1 Equilibrium Conditions 116
A.1.1 Full System 118
A.1.2 Steady State 120

Appendix B. Aggregate Implications of Capacity Constraints **123**
B.1 Solution 123
B.2 Estimation Details 125
B.2.1 Moment Conditions 125
B.2.2 Estimation Procedure 127
B.2.3 Results 128

Index **137**

Vita **138**

List of Tables

1.1 Large and Small Bank Differences: Intermediation Volume and Leverage . . 9
1.2 Increasing Concentration and Related Trends 13
1.3 Parameters for Heterogeneous and Homogenous Banks Comparison 48
1.4 Parameters and Induced Distribution for Rising Concentration Analysis . . 55

2.1 Strong Recessions are more Severe than Strong Expansions 66
2.2 Parameters 79
2.3 Quantitative Performance 80

B.1 Assumed Priors 128

List of Figures

1.1 Concentration in the US Banking Sector 4
1.2 Typical Bank Size Distribution 12
1.3 Timing 26
1.4 Ability Cutoff 40
1.5 Interbank Borrowing and Investment 41
1.6 Leverage and Size 42
1.7 Borrowing and Financial Friction 44
1.8 Impulse Responses: Heterogeneous and Homogeneous Banking Sectors . . . 50
1.9 Impulse Responses: High and Low Concentration 57

2.1 GDP and TFP measures 67

B.1 Posterior Distributions: Parameter Estimates 129

Chapter 1

Bank Size, Leverage, and Financial Downturns

I construct a macroeconomic model with a heterogeneous banking sector and an interbank lending market. 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 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.

I induce a financial downturn in the model with an exogenous shock to bank equity. Similar to the Great Recession, financial downturns are marked by a sharp decrease in interbank lending, causing a misallocation of liquidity among banks and, as a result, a contraction of the credit banks provide to firms.

I carry out two quantitative exercises with the model. First, when calibrated to match the observed long-term concentration in the banking sector, the model shows that a typical adverse financial shock leads to a considerably sharper downturn compared to an economy with a homogeneous banking sector. Second, I use the model to evaluate the impact of increasing concentration in US banking on the severity of potential downturns.

I find that if the banking sector in 2007 was only as concentrated as it was in 1992, GDP during the Great Recession would have declined by 40% less it did, and would have recovered twice as fast.

1.1 Introduction

The financial crisis of 2008 began in the housing sector but eventually affected employment and output in many other sectors of the economy. The banking sector played a crucial role in this story - banks held mortgages and their derivatives, and when the value of these assets declined, financial institutions engaged in a deleveraging process which reduced their investments dramatically, driving a large and persistent downturn in the real economy.

This project aims to further our understanding of one possible mechanism for the transmission of financial crises. A strand of previous work[1] considers the transmission of crises through the liquidity banks provide to each other through interbank loans and other forms of short-term debt. Banks rely on other banks for funds to efficiently meet day-to-day investment demands. If these funds are hard to come by in times of weakness in the economy, otherwise unaffected banks may not be able to function as efficiently as in good times. This leads to a decrease in the investments they make and amplifies the real effects of the initial shock.

In this paper, I go a step beyond this existing literature and investigate the following questions: does bank size, and the distribution of bank sizes in the industry, affect the transmission of crises in this way? If so, how?

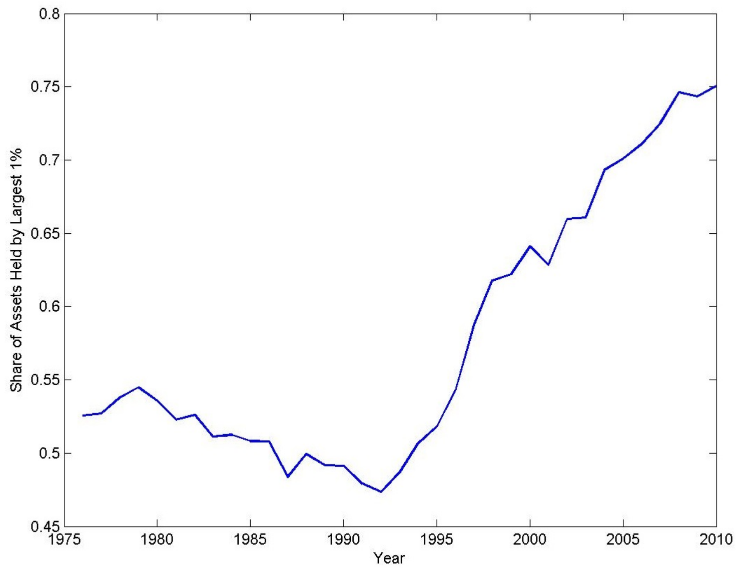
[1]Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) consider this mechanism through interbank lending specifically. He and Krishnamurthy (2012) and Boissay et al. (2013) consider similar effects, but liquidity problems arise between households and banks instead of between banks.

Why think about size? First, the financial crisis showed us that there may be good reasons to think that, on their own, big banks de-stabilize the economy in a crisis - in particular, 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. 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. Last, a key trend in US financial markets in recent decades has been a remarkable rise in the concentration of the banking system, as shown in figure 1.1. 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 entire banking sector.

I build a model in which households, banks, and firms are divided across a continuum of islands. Banks take deposits from households and lend them to firms on the same island, and also have access to a national interbank market where they can lend to or borrow from other banks. Banks will face a moral hazard problem when taking funds from other banks - they can divert a fraction of their funds (deposits and interbank loans) for their private purposes. Crucially, banks will differ in the productivity, or intermediation ability, with which they transform deposits into investments in firms.

In equilibrium, the moral hazard problem will limit the amount each bank can borrow from other banks; as a result, rather than efficiently allocating all funds to the most able bank, the interbank market will misallocate funds, directing some to less able banks. More able banks will borrow more on the interbank lending market and invest more in the firms on their island. The model generates an endogenous asset size distribution of banks

Figure 1.1: 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).

which matches two features of the data: (1) large banks will tend to be net borrowers on the interbank market, while small banks will tend to be net lenders, and (2) larger banks will be more leveraged than smaller banks.

I focus on financial downturns in the model by inducing a shock that destroys the value of net worth of all banks. The net effect on aggregate credit and investment will depend on two factors: the average ability of borrowing banks, or "extensive margin, and size of the investments/borrowing of each borrowing bank, or "intensive margin. 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. Moreover, the richness of the model enables us to examine the difference in the impact of downturns on individual banks. In particular, downturns in the model mirror downturns 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.

The dispersion of bank sizes will end up mattering for the depth of downturns after financial shocks. The more dispersed the bank sizes, the more banks rely on interbank lending for efficiently allocating funds. Downturns interrupt interbank lending, and they therefore have a bigger impact in an economy with a more dispersed banking sector.

I quantify these effects in two ways. First, I compare the effects of financial shocks in the model with heterogeneous banks to a model with a homogenous banking sector. I calibrate both models to the US economy in 2007, just before the financial crisis. Introducing realistic heterogeneity in the banking sector yields significant amplification and propagation of financial shocks. Recessions in the heterogeneous model lead to a 15% deeper drop in output.

Next, I use the model to answer another question: if the US banking sector had not experienced the long-term increase in concentration that it did, would potential downturns have been any less severe? If so, by how much?

Over the last three decades, increases in concentration have been accompanied by increases in the dispersion of bank sizes and the skewness of the size distribution, or relative frequency of big banks to small banks. I calibrate the model to roughly match the changes in concentration, dispersion, and skewness of the bank size distribution, from their lowest values in 1992 to the pre-crisis economy of 2007. I then consider the effect of a financial shock which produces a drop in output similar to that seen in 2007/08 financial crisis in both economies. If the banking sector in 2007 had maintained the lower concentration, dispersion, and skewness of the banking sector in 1992, the same financial shock would have produced a 40% smaller drop in output, and would have recovered twice as fast.

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 (2010), 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.[2]

My paper differs from this related work in two ways. First, my focus is different: I am 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.

Concentration in the banking sector has been steadily increasing over the last two decades, and a great deal of research has been done to consider its effects on the likelihood and severity of banking sector crises. One view is that concentration makes the economy

[2]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).

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 even of a downturn. 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 supporting the diversification theory). On the other hand, another strand of literature maintains 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. De Nicolo et al. (2004) performs a cross-country study using a dataset of large banks, and finds that higher concentration is associated with a higher fragility of the largest five banks. 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).

The paper proceeds as follows. I begin by presenting some stylized facts about the banking sector. The next section presents the model framework, describing its solution and discussing general properties. Then, I lay out the two quantitative exercises described above and briefly analyze them. A third section concludes.

1.2 Stylized 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 properties of financial institutions which vary systematically with 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. Because the data exhibits a strong time trend, I will also display values for 1992 and 2007, the two reference years I consider in the quantitative analysis.

First, bigger banks intermediate a lot more funds than small banks. The first row of table 1.1 shows the differences in the volume of loans performed by banks of different sizes calculated from Federal Reserve Bank of St. Louis (2013), 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.[3] 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 1.1 shows the inverse of the tier 1 leverage ratio, calculated for all bank holding companies by the Federal Reserve Bank of New York (2009). The measure in the graph is the ratio of the bank's risk-weighted assets, where assets are weighted by their regulator-determined credit risk, to the bank's tier 1 capital, which consists of a bank's equity and its retained earnings. Expressed in this way, 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

[3]Literature since the financial crisis focusing on systemic risk, and more recently Bremus et al. (2013), point out the degree to which this is true. A key component of these argument is that the largest banks are so large that they can individually influence macroeconomic outcomes. The difference in systemic importance is another key difference between banks of different sizes - for example, that because they are systemically important, big banks will be able to rely on implicit guarantees of government bailouts in the event of a crisis. (See Admati and Hellwig (2013) for a good discussion of this.) I do not consider this point here, but plan to address it in future work.

Table 1.1: 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
Total Deposits	2516	1318	1.9

Leverage

Measure	Large Banks	Small Banks
Assets/Equity		
Average	13.1	11
1992	17	13
2007	10	9.5
Assets/Tier 1 Capital		
Average	15.4	12.7
1996	14.8	11.9
2007	17	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 bn, 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 bn, 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 $ 20B [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).

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.[4]

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, the real economy suffers. 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.

Big banks tend to rely more on short-term liquidity markets than small banks do. With respect to interbank lending, Furfine (1999) finds that net borrowers of Fed funds tend to be larger in asset size than net lenders of funds. Cocco et al. (2009) finds 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.

Finally, I document a few more facts about the increasing concentration observed in the banking sector over time. A primary objective of this paper is to understand whether concentration has made downturns worse or better. The answer to this question will depend crucially on how concentration manifests itself.

Increasing concentration is typically characterized by an increase in the share of assets held by the largest banks. However, for nearly three decades in the US, this trend

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

has been accompanied in the data by two other relevant trends. First, banks have become more dispersed by asset size, that is, 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, that is, the bank size distribution has become more skewed to the right.

Though it has changed over time, the distribution has maintained three typical characteristics: it has a large mass of small banks, so the the left side of the distribution is heavy, and it has a long right tail, representing a few very large banks. Figure 1.2 is a histogram of banks in 2007 by their log asset size. Janicki and Prescott (2006) considers similar bank size distributions over time, and concludes that the distribution is best captured by a lognormal distribution with a Pareto tail. [5]

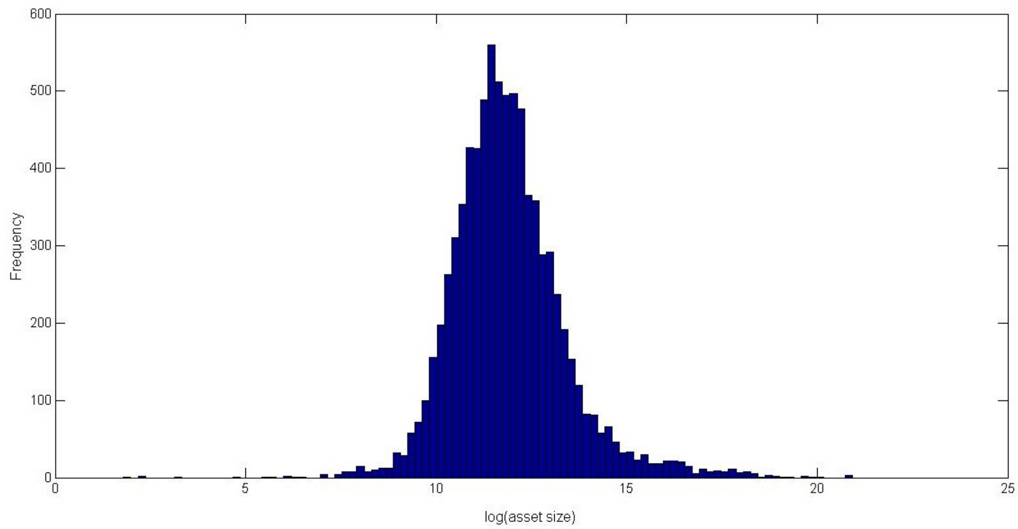
In answering this question, I choose to examine the banking sector in two reference years, one to represent a low concentration sector, another to represent a high concentration sector. I choose 1992 as the reference low concentration year for several reasons.[6] I choose 2007 as a reference high concentration year for comparison to the financial crisis.

Table 1.2 shows the relevant facts about the size distribution in each of the reference years. 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

[5]For computational convenience, I will model the bank size distribution with a truncated Pareto distribution in the quantitative section.

[6]In figure 1.1, 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 1.2: Typical 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 Federal Reserve Bank of Chicago (2013).

we see in the table, whether measured with the top 1% share, top 10% share, or the GINI coefficient, concentration has increased.

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 25th and 75th percentiles in size.[7] 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 the sample skewness over

[7]Because banks are generally larger today than yesterday, 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 1.2: 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	0.845	0.92
Dispersion		
Coeff. of Variation	7.65	15.6
Interquartile Range	97063	180933
Right Tail		
Skewness	33.7	39.49

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 US, 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).

the period. The bank size distribution typically has a large mass on the left side and a long right tail. Therefore, the increase in the skewness indicates an increase in the mass of the right tail of the size distribution.

1.3 Model

I consider an infinite-horizon economy comprised of a continuum of islands $a \in [0, 1]$. On each island, there are many households that supply labor and save funds, many firms which produce consumption goods from capital goods and labor, and many banks which 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 (2010) 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: households will own banks, and receive dividends that bankers pay out. Second, households will insure each other against consumption risk across islands. This setup is maintained to ensure that we can represent the household side of the model with a single, economy-wide representative household. All the heterogeneity in the model will be isolated in the banking sector.

Banks will differ in a single dimension, the quantity of assets they can invest in with a given dollar of deposit; I call this intermediation ability. 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 $\mu(A)$.

1.3.1 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 intermediate between households and firms, and differences between banks arise purely through differences in intermediation ability. After they receive deposits but before they borrow interbank loans, all the banks on an island a receive a random ability draw $\kappa(a)$. Draws are assumed to be distributed with cumulative distribution function $\kappa(a) \sim F(\cdot)$. I make several assumptions on the distribution of draws:

Assumption 1. The random ability draws for each island a in each period t, $\kappa(a)$, are independent and identically distributed over time, where $\kappa(a) \sim F(\cdot)$ has the following properties:

- $E(\kappa(a)) = 1$
- Draws have bounded support, so that $\kappa(a) \in [\underline{\kappa}, \overline{\kappa}]$.
- The distribution of draws F admits a density function, denoted $p(\cdot)$.

Banks with higher ability get better returns per dollar invested in firms[8]. I introduce ability into the model as a productivity term in a simple bank production function, so that if banks are producing assets S valued at price Q from liabilities and net worth L:

$$QS = \kappa L \tag{1.1}$$

Banks in this model are intermediaries, taking deposits from households and investing them as loans to firms. I implement size heterogeneity in this model through heterogeneity in productivity differences in this intermediation. [9]

[8]Ability differences, and their correlation with size, can reflect a number of different forces: bigger banks provide more services (e.g. consulting services, business contacts) that firms value, but don't manifest as differences in loan interest rates. These services reduce costs for firms, and the firms realize higher returns on projects as a result. Bigger banks may also have more productive investment hunters, so that per hour of bank employee labor, bigger banks find a higher number of investment opportunities.

[9]There is some evidence to suggest that such differences exist in reality. 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. Altunbas et al. (2001) 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.

In this model, big banks are big because they are better. As we will see later, it will turn out that they are also more leveraged because they are better. This represents a decidedly neutral, benign stance regarding big banks. For example, I could have instead assumed that managers at big banks prefer risk more than managers at small banks, as has been suggested since 2008.[10] I take this stance for two reasons. First, it is easier to implement this stance in the model. Second, any negative consequences of concentration that my model generates will represent a lower bound for more general cases where large banks prefer risk.

Banks raise deposits in an economy-wide deposit market. At the time they visit the deposit market, banks are identical, so that all island representative banks offer the same deposit rate R_t to raise deposits D_t from all households. Every bank will thus demand the same quantity of deposits, which means the banks on each island will hold $d_t \equiv d_t(a) = \mu(a)D_t$.

Ability increases bank demands for investment. Because deposits are made before ability is realized, deposits will not be allocated according to ability, and therefore a reallocation of funds after ability is realized can improve the profits of all banks. This reallocation is the primary function of the interbank lending market.

Banks on any island are able to borrow or lend to banks on any other island. Since there are many banks on each islands, 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,

[10] Perhaps the most widely discussed explanation for this is that too big to fail policies, or the implicit guarantee of a government bailout in the event of a crisis, influences the behavior of bank managers and executives. See Admati and Hellwig (2013) for a good explanation of this.

$(R_{bt}, b_t(a))$, where a is the island where the borrower banks live.[11]

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. In order 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; in order to ensure that these banks have something to invest with, these banks 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:

$$n_t(a) = [Z_t + (1-\delta)Q_t(a)]\psi_t(\sigma+\xi)s_{t-1}(a) - R_{t-1}d_{t-1}(a) - R_{bt-1}b_{t-1}(a) \tag{1.2}$$

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

[11] More precisely, an interbank loan contract is bilateral, that is, we should specify $(R_{bt}(a,a'), b_t(a,a'))$, where a, a' are the island where the borrower and lender banks live, respectively. However, since there are a continuum of potential lenders, lenders compete with each other for borrowers across all islands. All potential lenders then offer the same interest rate, since otherwise a borrower would go to another lender. The borrowed amount will be borrower-dependent because of the financial friction I describe below. Then the debt contract will have the property that $(R_{bt}(a,a'), b_t(a,a')) = (R_{bt}, b_t(a))$.

price of capital on island a, and s_{t-1} 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} is the funds borrowed on the interbank market last period, and d_{t-1} is the deposits made by households last period.

ψ_t is a shock to the quality of capital, and the key source of uncertainty in this model; it typically takes value 1. This kind of shock is crucial to the model, because in order to precipitate a "financial crisis", we need a way to exogenously affect the value of net worth. In addition, because the price of capital is endogenous in this model, this shock serves as an exogenous trigger for the type of asset price dynamics that were an important feature of the recent downturn.

Without intermediation ability, banks would balance the dollars invested in assets with the dollars taken as liabilities and equity, i.e. deposits, interbank loans, and net worth. With intermediation ability, banks balance assets against the intermediated value of liabilities and equity. We can summarize the balance sheet of the bank with a flow of funds constraint:

$$Q_t(a)s_t(a) = \kappa_t(a)\left[n_t(a) + d_t(a) + b_t(a)\right] \tag{1.3}$$

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 very end of a period, and sells its shares to another bank on the island. Depositors and potential lender banks know this, and though default will not occur in equilibrium, it will impose a limit on the amount of interbank loans any bank can borrow in a period.[12] On the other hand, there will be

[12] I interpret this "running away" as capturing a bankruptcy cost: if a borrower bank decides to declare

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.[13] 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.

$$V_t(d_{t-1}, s_{t-1}(a), b_{t-1}(a)) = \max_{\{d_{t+i},(s_{t+i}(a_i))_a,(b_{t+i}(a))_a\}_{i=0}^{\infty}} E_t \sum_{i=1}^{\infty} (1-\sigma)\sigma^{i-1}\Lambda_{t,t+i} n_{t+i}(a_{t+i}) \tag{1.4}$$

$$s.t. \; V_t \geq \theta Q_t(a) s_t(a) \tag{1.5}$$

Note that $n_t(a)$ 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, but 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.

---

bankruptcy, creditors can capture some, but probably not all, of the repayments they are owed. Despite the perhaps unrealistic form of the constraint, it generates the more realistic property that better banks can borrow more.

[13] As mentioned in the setup, households insure each other against differences in consumption across islands.

We can rewrite the above problem in terms of a Bellman equation:

$$V_t = \max_{d_t} E_t \max_{s_t(a),b_t(a)} E_{t,a}\Lambda_{t,t+1}\left[(1-\sigma)n_{t+1}(a) + \sigma \max_{d_{t+1}}\left(\max_{s_{t+1}(a),b_{t+1}(a)} V_{t+1}\right)\right] \quad (1.6)$$

$$s.t.\ V_t \geq \theta Q_t(a)s_t(a) \quad (1.7)$$

Additional Assumptions

I make two additional assumptions to ensure a tractable solution to the model. First, I make the ability distribution in any period independent of the ability distribution in previous periods. 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 (2010) and allow individual banks to arbitrage these return differences away, before their new ability type is realized:

Assumption 2. At the beginning of a period, banks can move between islands to equalize expected returns on assets.

To see how this works, 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 interbank debt of a_L to a_H but maintains the total asset held in a_L firms. In equilibrium, this process will continue until returns are equalized.

Another 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 re-allocate, or re-intermediate, the undepreciated portion of the capital stock among firms on the same island. In equilibrium, banks on low ability islands will only invest enough to maintain the undepreciated capital stock. Re-intermediation of capital will leave some islands with higher production capacity than others. In order 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 3. 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.

If we think of re-intermediation as bankers giving management advice to firm owners, this assumption essentially requires that any extra productivity realized by the firm gets paid to the banker as a consulting fee in the same period that the advice was given.

1.3.2 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, 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[14] at the interest rate R_t. Deposits made last period are

[14] We could alternatively restrict the household to only making deposits in the banks on their island; with

repaid at the beginning of this period.

I make another simplifying assumption on labor:

Assumption 4. Workers can supply labor on any island.

With the above, wages will be identical across islands - 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

$$C_t = W_t L_t + \Pi_t + R_{t-1} D_{t-1} - D_t \tag{1.8}$$

The household's maximization problem becomes

$$\max_{(C_t, L_t, D_t)_{t=0}^{\infty}} E_t \sum_{i=0}^{\infty} \beta^i \left[ln(C_{t+i} - \gamma C_{t+i-1}) - \frac{\chi}{1+\varphi} (L_{t+i})^{1+\varphi} \right] \tag{1.9}$$

$$s.t.\; C_t = W_t L_t + \Pi_t + R_{t-1} D_{t-1} - D_t \text{ (for each } t) \tag{1.10}$$

where $\beta \in (0, 1)$ is the discount factor and $\gamma \in [0, 1)$ is a habit formation parameter[15]. φ is the inverse Frisch elasticity of labor supply.

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

$$R_t(E_t \Lambda_{t,t+1}) = 1 \tag{1.11}$$

and a condition for aggregate labor:

$$W_t E_t(u_{Ct}) = \chi(L_t)^{\varphi} \tag{1.12}$$

the timing assumptions below, nothing about the model changes. I maintain this formulation for ease of explanation.

[15]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 $\gamma = 0$.

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.

1.3.3 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 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 the period, there is capital $k_t(a)$ on island a. The firm's problem then reduces to one of choosing how much labor to input. The representative firm on island a chooses $l_t(a)$ to maximize:

$$y_t(a) = A_t k_t(a)^\alpha l_t(a)^{1-\alpha} \tag{1.13}$$

where A_t is the (economy-wide) total factor productivity of the firm.[16] 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:

$$W_t = (1-\alpha)\left(\frac{y_t(a)}{l_t(a)}\right) = (1-\alpha)\frac{Y_t}{L_t} \tag{1.14}$$

where the second inequality also 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 need only solve the economy-wide representative firm's problem:

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

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{1.15}$$

Even though the ratio of net worth to capital is identical across islands by assumption 2, the capital on each island is not known. As a result, the size distribution of firms across islands in equilibrium will be indeterminate.

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 consists of this undepreciated capital and (possibly) of cash (basic goods) for new investment, which I denote by $i_t(a)$. This implies that, for the asset markets on each island to clear, it should be the case that

$$s_t(a) = (1-\delta)k_t(a) + i_t(a) \tag{1.16}$$

1.3.4 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

$$E_t \sum_{\tau=t}^{\infty} Q_\tau^i I_\tau - \left(1 + f(\frac{I_\tau}{I_{\tau-1}})\right) I_\tau \tag{1.17}$$

From the FOC for this profit maximization problem, the new capital price should satisfy

$$Q_t^i = 1 + f\left(\frac{I_t}{I_{t-1}}\right) + \left(\frac{I_t}{I_{t-1}}\right) f'\left(\frac{I_t}{I_{t-1}}\right) - E_t \Lambda_{t,t+1} \left(\frac{I_{t+1}}{I_t}\right)^2 f'\left(\frac{I_{t+1}}{I_t}\right) \tag{1.18}$$

1.3.5 Other Conditions

I'll add another assumption to ensure that the law of motion for capital has a simple form. (In equilibrium, this will also lead to more able lenders lending more than less able lenders.)

Assumption 5. Once installed, existing capital cannot be used on any other island.

With this assumption, banks will always find it optimal to reinvest their capital stock. This also implies that capital (asset) prices on each island are not necessarily the same. Denote the price of capital on island a by $Q_t(a)$.

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:

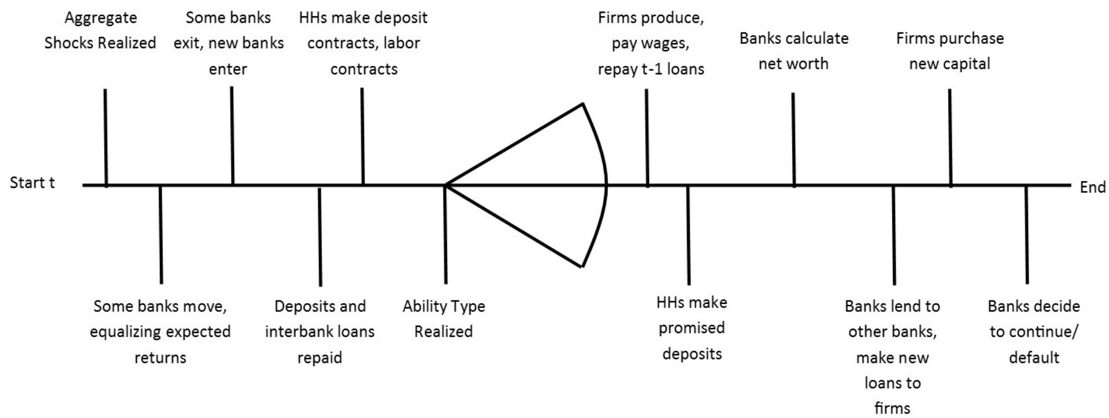
$$Y_t = C_t + \left(1 - f\left(\frac{I_t}{I_{t-1}}\right)\right) I_t \tag{1.19}$$

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

$$K_{t+1} = \psi_{t+1} \left(I_t + (1-\delta) K_t\right) \tag{1.20}$$

A period is divided into two parts: before and after intermediation ability is realized. Deposits are made in the first part, while firm loans, interbank loans, and investment purchases are made in the second. Figure 1.3 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.

Figure 1.3: Timing



The second half of the period begins when intermediation ability types are realized. Firms use household labor and capital to produce output, then pay profits Z_t to banks and wages W_t to workers. Firms borrow from banks, and those that want to expand their capital stock use the loans to purchase capital from the capital goods producers on the central island. Simultaneously, banks borrow from other banks. Banks then make their plan for the future, deciding in the process whether to default between periods.

1.3.6 Bank Solution

As mentioned above, the size distribution of firms across islands will be indeterminate in equilibrium. To get around this, I will define and solve a version of 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.

To solve the bank's problem, I'll follow a method similar to that presented in Gertler and Kiyotaki (2010). First, solve the bank's problem for the difference between the returns from investment and the deposit interest rate; there will be at least one such spread which clears the interbank lending market. (I will focus on situations where only one such spread exists.) Once this spread is known, solve the household and firm problems; only one combination of spread and deposit interest rate will solve the problems of these agents as well. We'll then solve for the remaining quantities by aggregating up to the ability type level.

The bank on island a maximizes net worth, which itself is a linear function of assets, deposits, and loans. Because of this property, the bank's value function will also turn out to be linear. The problem then boils down to solving for the coefficients of this value function.

First, guess that in every period a bank guided by a some linear function of shares, deposits, and interbank loans will maximize its expected net worth; we'll verify this guess later. Specifically:

$$V_t = \nu_{st} s_t(a) - \nu_{bt} b_t(a) - \nu_t d_t \tag{1.21}$$

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 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:

$$\begin{aligned} V_t &= \nu_{st} s_t(a) - \nu_t d_t - \nu_{bt} b_t(a) \\ &= \frac{\kappa(a)\nu_{st}}{Q_t(a)}(n_t(a) + d_t + b_t(a)) - \nu_t d_t - \nu_{bt} b_t(a) \\ &= \frac{\kappa(a)\nu_{st}}{Q_t(a)} n_t(a) + (\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_t) d_t + (\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \nu_{bt}) b_t(a) \end{aligned} \tag{1.22}$$

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:

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

where $\lambda_t(a)$ 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 $b_t(a)$:

$$\nu_t = \nu_{bt} \tag{1.23}$$

$$\kappa(a)\theta\frac{\lambda_t(a)}{1+\lambda_t(a)} = \kappa(a)\frac{\nu_{st}}{Q_t(a)} - \nu_{bt} \tag{1.24}$$

Using the above, we can also rearrange the borrowing constraint into a form that will prove useful later. If $\bar{b}_t(a)$ is the maximum a bank will borrow, then

$$b_t(a) \le \bar{b}_t(a) = \frac{\frac{\kappa(a)\nu_{st}}{Q_t(a)} - \kappa(a)\theta}{\nu_{bt} - \frac{\kappa(a)\nu_{st}}{Q_t(a)} + \kappa(a)\theta} n_t(a) - d_t(a) \equiv \phi_t(a)n_t(a) - d_t(a) \tag{1.25}$$

$\phi_t(a)$ is an important quantity. It is related to the leverage ratio (denoted $L_t(a)$), or the ratio of shares held to net worth; $L_t(a) = \kappa(a)(1+\phi_t(a))$. This means $L_t(a)$ is a convex, increasing function of κ. This property arises because intermediation ability affects banks in two ways: first, intermediation ability decreases the cost of shares directly through its effect on the flow of funds constraint; second, ability increases the continuation value relative to the "running away" value, loosening the borrowing constraint.

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

$$V_t = (1 + \lambda_t(a))\nu_{bt} n_t(a) \tag{1.26}$$

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

$$\nu_{st} s_t(a) - \nu_t d_t - \nu_{bt} b_t(a) = E_{t,a'} \Lambda_{t,t+1}(1-\sigma) n_{t+1}(a') + \sigma(1 + \lambda_{t+1}(a'))\nu_{bt+1} n_{t+1}(a') \tag{1.27}$$

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, $s_t(a)$:

$$\nu_{bt} = R_{bt} E_{t,a'} \Lambda_{t,t+1} \Omega^{a'}_{t+1} \tag{1.28}$$

$$\nu_t = R_t E_{t,a'} \Lambda_{t,t+1} \Omega^{a'}_{t+1} \tag{1.29}$$

$$\nu_{st} = E_{t,a'} \Lambda_{t,t+1} \psi_{t+1} (\Omega^{a'}_{t+1} (Z_{t+1} + (1-\delta) Q^{a'}_{t+1}) \tag{1.30}$$

where

$$\Omega^{a'}_{t+1} = 1 - \sigma + \sigma(1 + \lambda_{t+1}(a'))\nu_{bt+1} \tag{1.31}$$

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 (1.4):

$$V_t = E_{t,a'} \sum_{i=1}^{\infty} (1-\sigma)\sigma^{i-1} \Lambda_{t,t+i} n_{t+i}(a') \tag{1.32}$$

To see this, consider the effect of d_t on the original bank objective function. 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

$$\begin{aligned} \frac{dV_t}{dd_t} = (1-\sigma)\frac{\partial n_{t+1}}{\partial d_t} + \sigma(1-\sigma)\frac{\partial n_{t+1}}{\partial d_t}\frac{\partial s_{t+1}}{\partial n_{t+1}}\frac{\partial n_{t+2}}{\partial s_{t+1}} \\ + \sigma^2(1-\sigma)\frac{\partial n_{t+1}}{\partial d_t}\frac{\partial s_{t+1}}{\partial n_{t+1}}\frac{\partial n_{t+2}}{\partial s_{t+1}}\frac{\partial s_{t+2}}{\partial n_{t+2}}\frac{\partial n_{t+3}}{\partial s_{t+2}} + ... \end{aligned} \tag{1.33}$$

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:

Proposition 1. A set of choices $(d^*_{t+i}, b^*_{t+i}(a))_{i=0}^{\infty}$ that maximizes the linear value function guess (1.21) for each i will also maximize the bank objective (1.32).

1.3.7 Solution Properties

The linear solution has several useful properties. I will first use it to characterize the capital prices on each island a, then use that to characterize bank behavior in the interbank lending market. I will also mention two restrictions on parameters that ensure equilibrium existence.

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. I will assume that, given indifference between these choices, the bank will always reinvest the existing capital stock on its island.

Assumption 6. $s_t(a) \geq (1-\delta)k_t(a)\ \forall a$

Further, from the FOC for $b_t(a)$ in the previous section, we can see that if the banks on island a are not borrowing constrained, it must be that the price of capital on its island satisfies

$$\frac{Q_t(a)}{\kappa(a)} = \frac{\nu_{st}}{\nu_{bt}} \equiv Q_t^n \tag{1.34}$$

Assumption 5 also implies that the price of capital on each island is free to adjust to a different value on each island. Make another guess which we can verify later:

$$Q_t(a) = \kappa(a)\frac{\nu_{st}}{\nu_{bt}} = \kappa(a)Q_t^n\ \ \forall a \in L \tag{1.35}$$

where L is the set of ability types for which banks are not borrowing constrained. This price essentially represents 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 value ν_{bt}, or it could turn it into $\frac{1}{Q_t(a)}$ units of capital, earning it $\frac{\nu_{st}}{Q_t(a)}$ value. The previous equation tells us that this shadow cost increases in intermediation ability, that is, re-investing the existing capital stock is more costly for more able banks. To put it another way, banks can demand higher repayments per unit capital lent to the firms on its island.

What about the value of existing 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, as in the previous case, the price $Q_t(a)$ would get very large, as banks would demand ever higher returns per unit capital.

I maintain that firms will never allow the price $Q_t(a)$ to exceed the price for new capital, Q_t^i. This amounts to preventing banks from extracting too much from firms. 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 agree to these terms, the firm instead threatens to go to another bank on the island which will give it a cash loan instead, which it could then use to buy new capital at the lower price Q_t^i. Since the firm is small, 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:

$$Q_t(a) = Q_t^i \ \forall a \in B \tag{1.36}$$

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.

First, we need to restrict $\underline{\kappa}$, the lowest possible intermediation ability, to ensure that banks on all islands can fund reinvestment of their entire capital stock. Renegotiation of the remainder of the capital stock means that each share of this portion costs $\frac{Q_t(\kappa)}{\kappa}$. For high productivity islands, this is a boon; the cost of these shares is smaller than that of the fixed capital. For low productivity islands, this is a burden; their poor managerial skill causes them to lose some of their net worth through this process. Thus, if we require that the entire existing capital stock will be reinvested, we must ensure that the ability of the worst bank is not so low that it cannot cover the cost. This is accomplished if we ensure that $\underline{\kappa}$ satisfies

$$(1-\underline{\kappa})Q_t^n(1-\delta)K_t \leq Z_tK_t - R_{t-1}D_{t-1} + D_t \tag{1.37}$$

In steady state, this equation becomes

$$(1-\underline{\kappa})Q^n(1-\delta) \leq Z + (1-\frac{1}{\beta})\frac{D}{K} \tag{1.38}$$

Second, the borrowing constraint equation only acts as an upper bound on borrowing if the denominator of $\phi_t(a)$ is positive, that is, $\nu_{bt} > \frac{\kappa\nu_{st}}{Q_t^i} - \kappa\theta$. Though equilibria exist if the inequality is reversed, all banks in these equilibria are unconstrained, which would make the incentive constraint useless. I avoid this case by choosing $\overline{\kappa}$ and θ so that

$$\theta > \frac{\nu_{st}}{Q_t^i} - \frac{\nu_{bt}}{\overline{\kappa}} \tag{1.39}$$

In steady state, this equation becomes

$$\theta > \nu_s - \frac{\nu_b}{\overline{\kappa}} \tag{1.40}$$

Given the linear solution, it will turn out that high ability banks will borrow, and be borrowing constrained, while low ability banks will lend. To see this, first note that the flow of funds constraint reduces the number of bank choice variables to two, and since deposits are chosen before ability type is realized, the only choice the bank makes after ability is realized is the level of interbank loans.

Proposition 2. Consider an equilibrium with strictly positive lending. In each period t, there exists a κ_t^* such that:

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

- for all banks with $\kappa > \kappa_t^*$, $b_t(a) = \overline{b_t}(a)$, 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 (1.22), and let us first assume

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

In this case, the bank gets positive value for every dollar it borrows, and it will choose to borrow as much as it can, 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 $\kappa > \frac{\nu_{bt}}{\nu_{st}} Q_t^i$.

If $(\frac{\kappa \nu_{st}}{Q_t(a)} - \nu_{bt}) \leq 0$, the bank gets positive value for every dollar it lends (negative borrowing). By assumption 6, the bank will only lend funds left over after re-lending its existing capital stock, i.e. $b_t(a) = -(n_t(a) + d_t - Q_t^n k_t(a))$. We know that capital prices on these islands are smaller than the new capital market price, so $\kappa \leq \frac{\nu_{bt}}{\nu_{st}} Q_t(a) \leq \frac{\nu_{bt}}{\nu_{st}} Q_t^i$.

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 ability differences between banks in this model generate differences in borrowing and investment behavior. The above proposition tells us that these differences are easily summarized with one equilibrium object, the ability cutoff.

It will also turn out that lower ability interbank lenders will lend less than higher ability interbank lenders.

Proposition 3. Consider two banks on different islands 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 3, 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. □

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.

1.3.8 Aggregating to Ability Types

The size distribution of firms across islands is indeterminate. Rather than characterizing and solving an equilibrium at the island level, I will characterize the model at the ability type level. Aggregating the model in this way will lend itself to a tractable solution concept. In this section, the quantity x on all islands with the same ability type (a sum across 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.[17] 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 installed last period will just be a fraction of aggregate capital: $b_{t-1}(\kappa) = 0$, $s_{t-1}(\kappa) = p(\kappa)K_t$. Then the aggregate net worth for all islands with the same ability type will be

$$n_t(\kappa) = [Z_t + (1-\delta)Q_t(\kappa)]\psi_t(\sigma + \xi)p(\kappa)K_t - p(\kappa)\sigma R_{t-1}D_{t-1} \tag{1.41}$$

The first term 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 in period $t-1$. 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.

[17] 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, so that $\kappa(a_1, a_2) = \kappa(a_1, \widetilde{a}_2)$, and in odd periods, all islands with the same second coordinate receive the same draw. Then for any island with a particular κ, there will be a positive measure of islands with the same κ. Since $[0,1] \times [0,1]$ has the same cardinality as $[0,1]$, I leave this setup out of the main text for ease of exposition.

On each island, banks choose deposits d_t identically, before uncertainty is realized, and choose $s_t(a)$ with the solution (1.21). Across islands, this solution only changes when ability type changes, that is, banks on all islands with the same ability type will find the same values for the coefficients ν_{st}, ν_{bt}, and ν_t. Banks on different islands could still have different net worth, however; rewriting the flow of funds constraint for the bank with the solution, we see that

$$s_t(a) = \frac{\kappa(a)}{Q_t(a)}(1 + \phi_t(a))n_t(a) \tag{1.42}$$

But the quantities $Q_t(a)$, $\phi_t(a)$, and $\kappa(a)$ are identical for any islands with the same ability type. This implies that we can represent the ability type aggregate flow of funds constraint with the same form as the individual island, so that if $Q_t(\kappa)$ and $\phi_t(\kappa)$ are type common prices,

$$s_t(\kappa) = \frac{\kappa}{Q_t(\kappa)}(1 + \phi_t(\kappa))n_t(\kappa) \tag{1.43}$$

Thus, we can solve for the ability type aggregates for banks with the maximization problem

$$V_t = \max_{\{d_{t+i},(s_{t+i}(\kappa_{t+i}))_\kappa,(b_{t+i}(\kappa_{t+i}))_\kappa\}_{i=0}^{\infty}} E_t \sum_{i=1}^{\infty}(1-\sigma)\sigma^{i-1}\Lambda_{t,t+i}n_{t+i}(\kappa_{t+i}) \tag{1.44}$$

$$s.t.\ V_t \geq \theta Q_t(\kappa)s_t(\kappa) \tag{1.45}$$

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 2, we can pin down the quantities $k_t(\kappa)$ and $i_t(\kappa)$ for each ability type. These will take the form

$$k_t(\kappa) = p(\kappa)K_t \tag{1.46}$$

$$i_t(\kappa) = s_t(\kappa) - (1-\delta)k_t(\kappa) \tag{1.47}$$

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:

$$y_t(\kappa) = A_t k_t(\kappa)^{\alpha} l_t(\kappa)^{1-\alpha} \tag{1.48}$$

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 an 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 $\mathbb{Q}_{\kappa t} \equiv (b_{t+i}(\kappa), s_{t+i}(\kappa), d_{t+i}(\kappa), i_{t+i}(\kappa))_{i=0}^{\infty}$[18] such that:

(Individual Optimization) for each t,

- $(d_t, b_t(\kappa), s_t(\kappa)$ maximizes the representative bank's expected value (1.44) subject to their flow of funds constraint (1.3) for each ability type κ
- $(k_t(\kappa), l_t(\kappa))$ maximizes the representative firm's profits (1.48) for each ability type κ

[18]Note that type-specific quantities are functions $x_t(\kappa) : [0,1] \to \mathbb{R}$

- (C_t, L_t, D_t) maximizes the economy-wide representative household's expected utility (1.9)

- I_t maximizes capital goods producer profits (1.17)

(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): $s_t(\kappa) = (1-\delta)k_t(\kappa) + i_t(\kappa)$

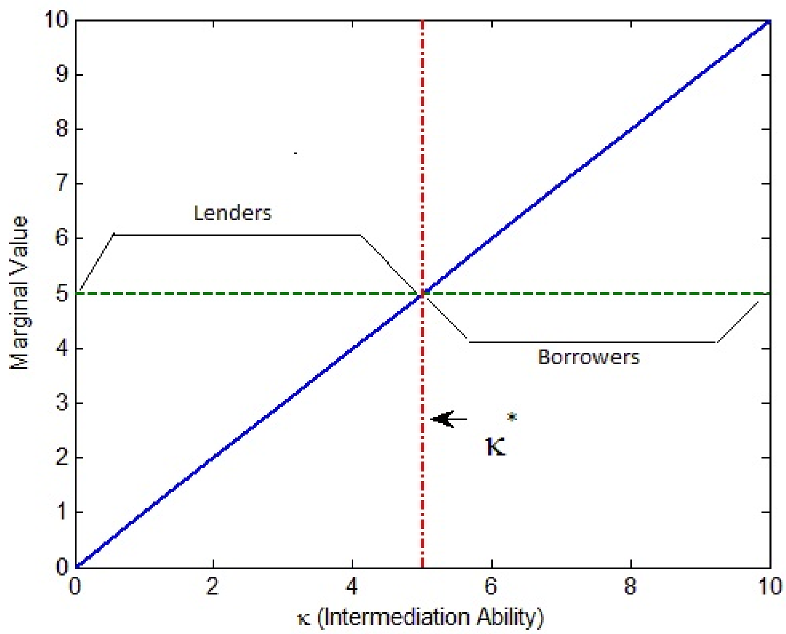
This system can be solved for any ability type distribution satisfying assumption 1. The full set of equilibrium equations is given in appendix A.

1.3.9 Steady State Properties

To illustrate the basic mechanism of the model, consider the case where ability is distributed according to a uniform distribution, and the economy is in steady state.

Figure 1.4 shows the marginal value from lending on the interbank market (dash line) and the marginal value from investing in firms. This is essentially a visualization of proposition 2. 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.

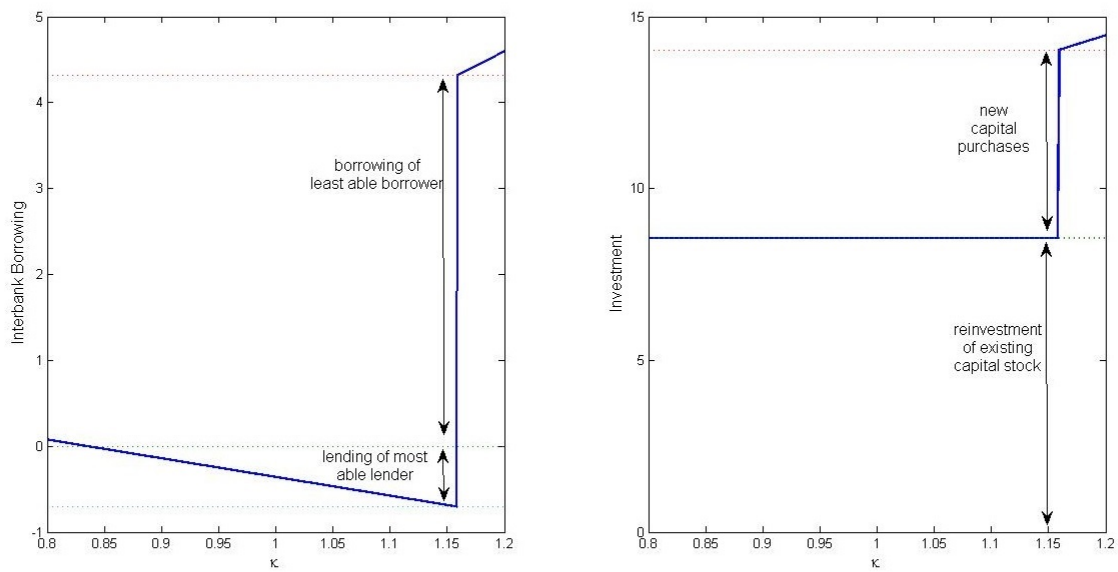
Figure 1.4: Ability Cutoff



Note: Marginal value from interbank borrowing (dash) and marginal value from holding shares (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.

Ability increases the intensity with which banks borrow or lend in the interbank market. Figure 1.5 shows interbank borrowing versus ability in a typical case of the model. All banks with ability smaller than the ability cutoff κ^* lend on the interbank market, and because higher ability banks are more efficiently able to refinance their existing capital stock,

Figure 1.5: Interbank Borrowing and Investment



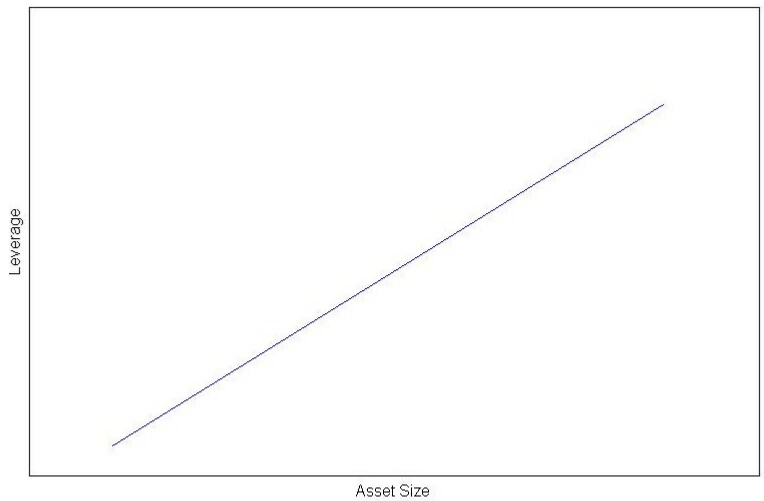
Note: Interbank borrowing (left) and investment (right) versus intermediation ability.

they have more dollars left over for lending, and therefore lend more on the interbank lending market than lower ability banks. (Since the vertical axis represents borrowing, lending is represented with a negative y-coordinate.)

Banks with ability higher than the ability cutoff decide to borrow, buy new capital, and expand the production capacity of the firms on their island. The difference in borrowing between banks just above and just below the cutoff can be divided into two components. The first component represents a reduction in lending - banks just below the cutoff make these loans. The second component represents borrowing by banks up to the point that their borrowing constraint binds.

Moreover, the cap on borrowing increases in κ at an increasing rate. The function is convex, implying that the average leverage in the economy is higher than the leverage of the average bank. This convex relationship between leverage and intermediation ability will ultimately generate a linear relationship between leverage and size, as shown in figure 1.6. This is one of the properties we wanted to replicate.

Figure 1.6: Leverage and Size



Note: Leverage, or the ratio of bank assets to bank net worth, for banks of different asset sizes in a typical steady state. As ability increases, leverage increases at an increasing rate, but as size increases, leverage increases linearly.

Interbank lenders lend different quantities based on their intermediation ability. This is a result of both assumptions 5 and 6. The cost of refinancing the capital stock is smaller for more able banks, and since all lending banks refinance their capital stock, more able banks have more funds left over to lend out to other banks.

The right panel of figure 1.5 shows the value of investments (measured in units of consumption goods) by banks of different abilities. All banks reinvest their existing capital stock - capital prices on lending islands adjust to ensure that this is optimal. Once ability increases above the cutoff, we see a jump in the quantity of investments as banks suddenly switch from a "lend and refinance" strategy to a "borrow and buy new capital" strategy. As ability increases, borrowing limits increase, and because the price of new capital is pinned down by the new capital market, this translates into higher investment.

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. This bank would transform these loans into assets, which firms would use to purchase the highest amount of capital possible. This borrowing curve would look something like the steepest, dashed curve in figure 1.7.

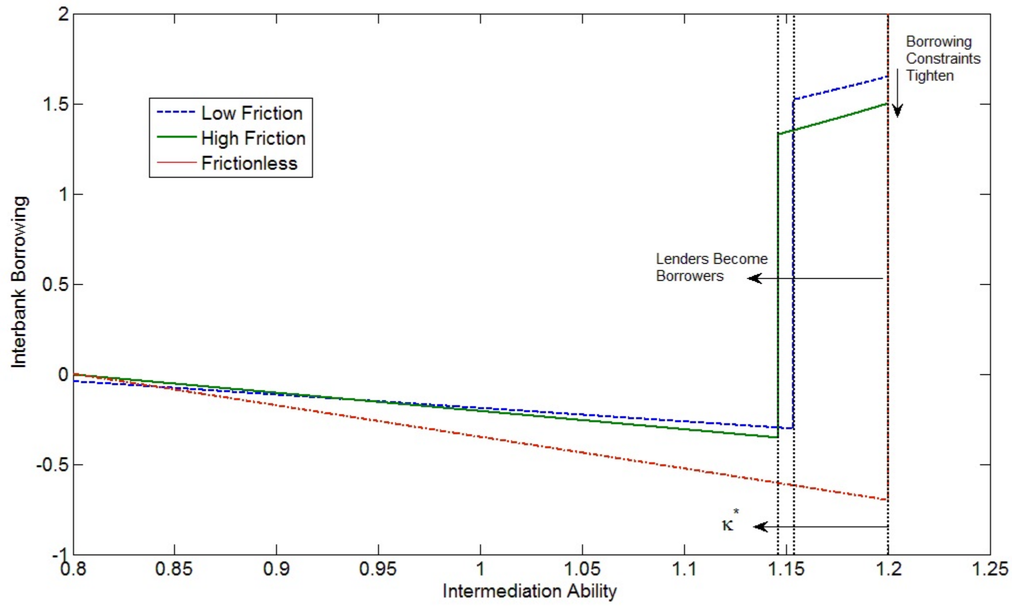
Introducing the friction limits the maximum the most able bank can borrow through the interbank lending market. By limiting the maximum the bank can borrow, funds that would go to the most able bank go to less able banks. Economy-wide investment and average returns banks obtain on investments decrease.

To visualize this, compare the steady state from the previous case with the steady state that obtains after an increase in the level of financial friction, θ. In figure 1.7, the steady state borrowing curve from the model above are plotted along with steady state borrowing from $\theta = 0.4$ (solid green), as in the previous section, to 0.5 (dashed blue).

Among interbank borrowers, a change in θ has both an "intensive" and an "extensive" effect on steady state behavior. The intensive effect is a change in leverage for all banks: the value of the leverage ratio $\phi_t(a)$ decreases as θ increases, 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. Moreover, the leverage curve gets less steep, so an increase in ability results in less additional borrowing for higher θ.

On the extensive side, as the friction increases and demand for interbank loans drops, the price of borrowing should decrease, at least relative to the value from lending

Figure 1.7: 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. In all three cases, ability is distributed uniformly in $[0.8, 1.2]$.

to firms. This decrease in price causes some banks that were interbank lenders to become borrowers and invest. The increase in demand for borrowing that results from this extensive movement will offset some of the initial fall.

Both of these effects result in a decrease in the volume of interbank loans that are made in the economy. This is what we should expect - both demand and supply of interbank loans falls.

The magnitude of a given change in the level of the friction decreases as the initial friction increases. For parameter values for which steady state equilibria exist, an increase in the level of the friction always results in a decrease in total interbank borrowing. But the size of this decrease itself decreases as the initial level of the friction increases. Since the leverage ratio increases at an increasing rate in the friction, a higher initial friction level essentially means that we are starting higher up on the leverage ratio curve, where the increases are larger.

Among interbank lenders, we see a decrease in the slope of the line that determines how much lenders lend; more able lenders still lend more than less able ones, but the size of the difference between the two decreases as the level of the friction increases. This is because the steady state level of capital decreases as the friction increases, and because the cost of refinancing is a linear function of the steady state level of capital, the slope decreases.

1.4 Homogeneous and Heterogeneous Banking Sector Comparison

In this section, I compare the response of this model to that of Gertler and Kiyotaki (2010), a model with a homogeneous banking sector. I calibrate both models to match characteristics of the 2007 economy, and show that downturns in the heterogeneous banks model are deeper.

1.4.1 Calibration

The model demands the choice of 9 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 (2010) and Christiano et al. (2005) and are given in Table 2.2 below.

The parameter φ, the inverse elasticity of labor supply, is chosen so that the Frisch elasticity is ten. 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.

Adjustment costs take the quadratic form:

$$f\left(\frac{I_t}{I_{t-1}}\right) = \frac{c_I}{2}(\frac{I_t}{I_{t-1}} - 1)^2 \tag{1.49}$$

The parameter θ and the distribution of ability types are the main quantities of interest. Together, they govern the average leverage held by all banks and the bank size distribution that is generated in the model. The equilibrium existence restriction (1.39) on θ limits how small this parameter can be chosen.

The distribution of ability types will always take the form of a bounded Pareto distribution, for two reasons: first, the bank size distribution over the period takes this approximate shape[19], and second, using the Pareto distribution is computationally convenient. This distribution has a pdf, given by

$$p(x) = \frac{\rho \underline{\kappa}^{\rho} \overline{\kappa}^{\rho}}{\overline{\kappa}^{\rho} - \underline{\kappa}^{\rho}} x^{-\rho-1} \tag{1.50}$$

where L and H are the lower and upper bounds of the distribution. In accordance with assumption 1, the distribution will always be chosen to have mean 1.

In what follows, the level of θ, the bounds of the Pareto distribution L and H, and the shape parameter ρ will be adjusted differently in each case.

[19] 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. For this exercise, I exclusively use the Pareto distribution for computational convenience.

In Gertler and Kiyotaki (2010), all banks are identical in ability. Demand for interbank loans is created by exogenously specifying that a constant fraction π^i of islands are the only ones allowed to purchase new capital every period. In my model, the banks with firms that purchase new capital are those with ability above the cutoff; therefore, the fraction of all islands that purchase new capital is an endogenous object. Therefore, I choose π^i in the simulation of the model so that in steady state, the fraction of banks that purchase new capital in my model (i.e. those with ability higher than the cutoff) matches the fraction of banks that purchase new capital in theirs.

I choose a calibration that matches the baseline calibration of their paper closely, with two exceptions. Their choice of the investing fraction is 25%. My model tends to generate smaller investing fractions for a Pareto ability distribution, so though I choose parameters for the ability distribution so that the induced investing fraction is as close to their baseline choice as possible, I am only able to generate a fraction in my model around 10%.

All impulse responses are generated using the full set of equilibrium equations in Appendix A with Dynare.

1.4.2 Crisis Response

The 2008 financial crisis, and banking crises more broadly, have their roots in more fundamental weaknesses in the wider economy. In this paper, I do not take a stand on the sources of these weaknesses - given some initial weakness in the economy, I am interested in any additional negative effects imperfect financial intermediation might create.

As such, I initiate a downturn in this economy with an exogenous shock to the value of capital in the economy. Since banks in the model derive their net worth (equity) partly

Table 1.3: Parameters for Heterogeneous and Homogenous Banks Comparison

Parameter		Value	Target
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
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)
Heterogeneous Banks Model			
Parameter		Value	Target
Friction	θ	0.3	Leverage Ratio - 2007
Min Ability	$\underline{\kappa}$	0.9	High Investing Bank Fraction
Max Ability	$\overline{\kappa}$	1.18	Mean 1 Assumption
Shape	ρ	5	Skewness - 2007
Homogeneous Banks Model			
Parameter		Value	Target
Fraction of Banks that Invest	π^i	0.1017	Heterogeneous Banks Investing Fraction

from the capital investments they've already made, this shock to capital translates to a shock to the net worth of all banks.

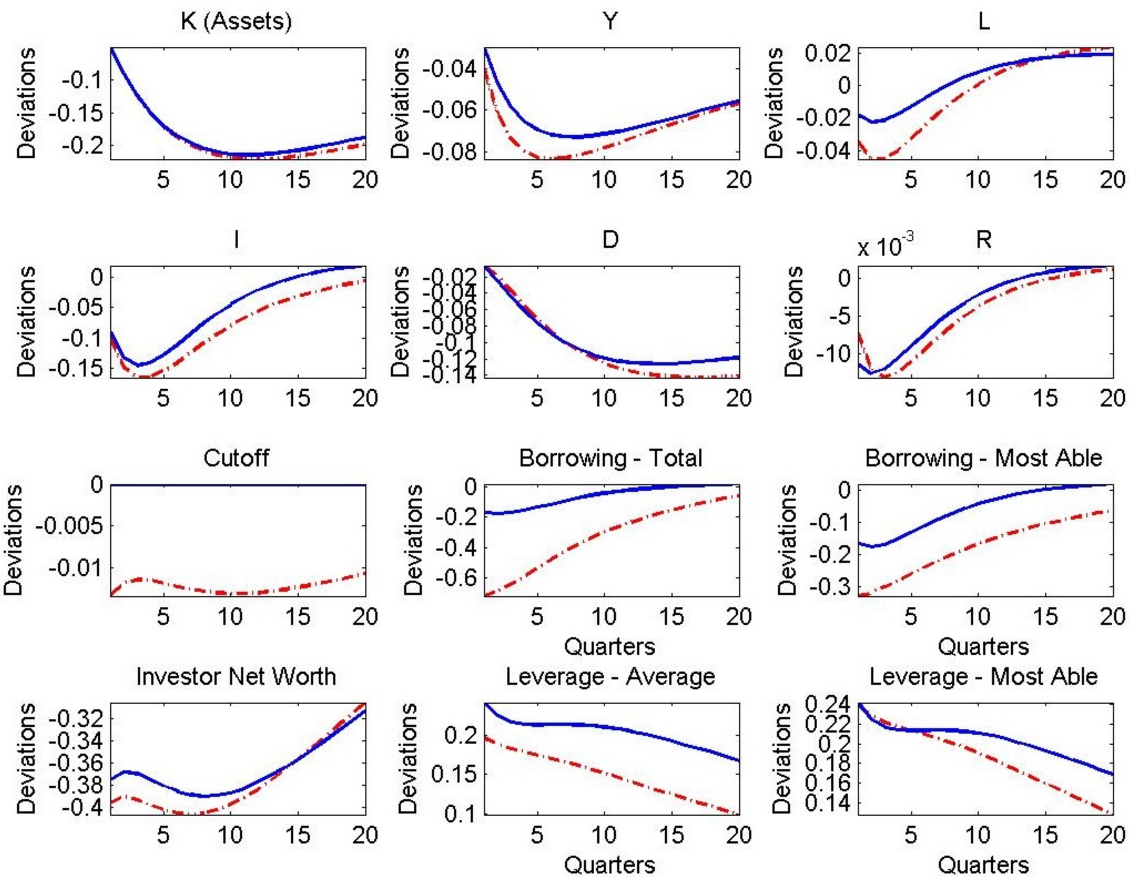
In the baseline experiment, I initiate a crisis in the model with a 5% shock to the capital quality ψ_t - the level is chosen to initiate a drop in output similar to that seen during the 2008 crisis. The shock has a persistence of 0.8, implying it returns to its steady state value after five years.

Figure 1.8 shows the impulse responses of both models, in terms of percentage deviations from steady state, to the initial shock. Generally, the shock makes capital more scarce, which causes banks to reduce their asset holdings, and because of the persistence of the shock, this reduced capital is reduced next period as well. This de-leveraging process ultimately produces the hump shape in capital in the figure, and drives the other responses.

In both models, the story proceeds roughly the same. 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 models, the fall is roughly the same. Output Y falls, and because the marginal productivity of labor falls upon impact, demand for labor falls. Wages fall as well, so the supply of labor and employment falls. Both Y and L fall in the figure, though to different minimum levels. The decrease in wages and employment forces households to save less today, leading to lower supply of deposits in both models

Second, on the bank side, because capital is held as assets by banks, the initial destruction of capital leads to a significant decrease in bank net worth - this is seen in the bottom left panel. This occurs because the initial decline in capital has second-round effects, reducing both the price of capital, which affects net worth through the value of the existing capital held by banks, and the net profits the bank receives from firms.

Figure 1.8: Impulse Responses: Heterogeneous and Homogeneous Banking Sectors



Note: Impulse responses to -5% financial shock for heterogeneous banking sector (dash) and homogeneous banking sector (solid). Shock persists for 20 quarters.

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. Because the economy faces adjustment costs, investment falls even further in the periods just after the initial shock, as banks have to de-lever more to cover their adjustment costs.

In addition, the shock to capital quality also destroys the net worth of lender banks. This, combined with the drop in deposits supplied by households, reduces the supply of interbank loans greatly. 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; we see this in the figure in the panel labeled average leverage. This happens 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.

As outlined above, dispersion can amplify the effects of financial frictions. An implication of this is that the downturns generated in this model can be deeper than those generated in homogeneous ability model. Banks in a heterogeneous banking sector rely more on a well-functioning interbank lending market, so when interbank lending is interrupted by a financial shock, the heterogeneous banking sector sees a deeper resulting downturn.

With heterogeneous banks, because more able banks are more leveraged, the average bank size is larger than the average bank size in the homogeneous case. This means that the steady state level of investment is also larger for the heterogeneous bank model. In addition, a larger portion of that 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.

Net worth decreases more in the heterogeneous banks case because the net profits from firms decrease more, which occurs because output Y falls more in the heterogeneous banks case. This itself has to do with the fact that investment falls farther - thus, the relationship between investment today and net worth tomorrow produces a feedback effect which amplifies the shock.

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.

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 able, and therefore less leveraged when they do borrow. Their entrance brings down the average leverage in the economy. No switching occurs in the homogeneous banks model, so in comparison, average leverage in the homogeneous banks model is more strongly countercyclical.

Ultimately, the farthest output deviates from its steady state value in the homogeneous economy is 15% smaller than the farthest it deviates in the heterogeneous economy. Investment sees more modest amplification of the initial shock, but note that it also takes longer to return to its steady state value.

As shown above, the banks that switch from lending to borrowing stabilize the interbank lending market, partially 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.

1.5 Quantitative Effects of Rising Concentration

In this section, I consider the impact of two changes to the size distribution that resulted from the increase in US banking sector concentration over the last three decades: bank sizes became more dispersed, and the number of large banks increased relative to the number of small banks. I calibrate the model to match these changes qualitatively, and show that they tend to exacerbate downturns as well.

As shown above, the US banking sector has become increasingly concentrated (in asset size) over the last two decades. In this section, I consider the following question: if a less concentrated bank size distribution faced the same shock to the quality of capital, would the resulting downturns be worse? If so, by how much?

As seen in the data, increasing concentration was accompanied by an increase in the dispersion and increase in mass of the right tail of the size distribution. I will match all three qualitative changes by varying the width and the shape parameter of the ability distribution from the calibration above. An increase in the width, or the difference $\overline{\kappa} - \underline{\kappa}$, will increase the dispersion as measured by the coefficient of variation in my model, as the mean of the distribution is always 1. Since the density of the Pareto distribution reaches a peak on its left side, and an increase in the shape parameter tends to increase the size of this peak, an increase in the shape parameter will decrease the mass of the right tail of the distribution.

Banks in the model either borrow as much as they can or lend as much as they can - their equilibrium value functions are linear. Because of this, in order to generate the inequality in bank sizes observed in the data, the fraction of banks that does any borrowing should be very small. Because there is also a restriction on how low a value the friction can take, the total new investment made by all banks in the banking sector will also be very small. Under the baseline calibration, a very high inequality economy becomes dominated by the actions of lending banks, an uninteresting case from the perspective of this paper.

One partial solution to this problem is to increase the depreciation rate of capital. I change the depreciation rate from the previous section, increasing it to 10%. By doing this, we increase the new investment each borrowing bank makes, making the actions of borrowing banks more important for aggregates. This is only a partial solution in the sense that the model still cannot match the inequality we observe in the data, but it can get much closer than with the baseline calibration. In order to simplify the discussion of responses, I also drop habits and adjustment costs in this section.

I choose another ability distribution that generates more inequality than in the baseline calibration, though still not as much as what we observe in the data. The choices for these parameters, and the corresponding dispersion and skewness measures, are given in the first two rows of table 1.4; other parameters are the same as in the baseline calibration above. The increased inequality comes at the cost of a reduction in the induced investing fraction, which is significantly lower than in the previous section.

The next rows list the steady state concentration, dispersion, and kurtosis of the induced size distribution for each of the two cases. The change in parameters generates the same qualitative changes in the size distribution as we see in the data; dispersion increases, and kurtosis increases, as we move from the representative low concentration to high concentration case. In addition, average leverage increases, also matching the time trend in leverage in the economy. As far as concentration, the model underestimates the change in concentration as measured by the top 1% share of assets held, but overestimates the change in concentration as measured by the top 10% share.

As in the previous case, a downturn is initiated with a 5% negative shock to the quality of capital, ψ, and has a persistence of 0.8, so that it completely dissipates after five years.

Table 1.4: Parameters and Induced Distribution for Rising Concentration Analysis

Year	ρ	θ	$\underline{\kappa}$	$\overline{\kappa}$	c_I	δ
Low Concentration	7	0.1	0.81	2.15	0	0.1
High Concentration	6	0.1	0.835	2.8	0	0.1

Year	Top 1%	Top 10%	GINI	Coeff of Variation	Avg Leverage
Low Concentration	0.05	0.36	0.57	1.1	6.6
High Concentration	0.15	0.58	0.72	2.1	14

The impulse responses for these cases are given in figure 1.9. 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 the low concentration economy; thus, if the high concentration case represents the economy just before the financial crisis, the economy would have experience a downturn that was only 60% as deep.

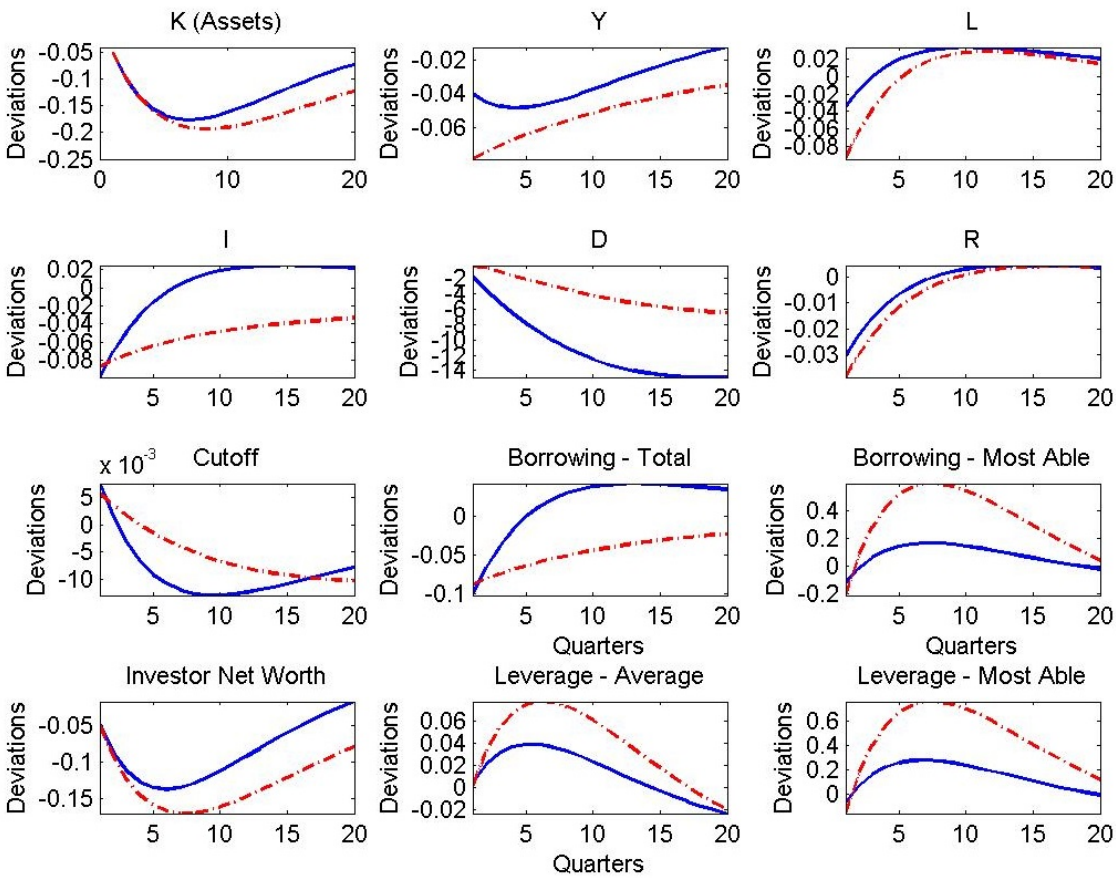
Second, output in the high concentration economy returns to steady state much more slowly than in the low concentration economy. Though not pictured, output in the high concentration economy takes twice as many quarters to return to its steady state level than in the low concentration economy.

The difference on impact of output in the two economies is due to the initial difference in the response of employment. As shown in the top right panel, employment falls much farther in the high concentration than 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 high concentration economy. As a result, employment falls farther initially, and output falls with it.

Because the high concentration economy is more dispersed, the average leverage of borrowing banks is higher. As in the case of amplification over a homogeneous banking sector, 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 also 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

Figure 1.9: Impulse Responses: High and Low Concentration



Note: Impulse responses to -5% financial shock for high concentration (dotted, red) and low concentration (solid, blue) banking sector. Responses are expressed in percentage deviations. Shock persists for 20 quarters.

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 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, dampening the negative effects of de-leveraging relative to the high concentration economy.

1.6 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.

Though this paper only considers positive questions, we can still discuss the role of policy responses[20] The model in this paper puts us in the unique position to consider targeted macro-prudential policies in a general macroeconomic framework. One such policy taken up by the US Treasury and the Fed was embodied in broad targeted asset purchases like the TARP, which in some cases equated to the government taking up part-ownership positions in some banks. In the context of this model, this would reduce the friction target banks face. 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 that that 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, which is not offset by the change in the friction for high levels of friction. In this case, offering a targeted program which doesn't affect the interbank market would improve welfare more than one which ignored this.

Despite the stylized assumptions in the model, the environment is rich, and I see several 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 arguments, and banks would smoothly transition from lending to borrowing as ability increased. With

[20] The policy responses discussed above are implemented in a companion paper, still in progress.

this, we could more easily generate high inequality.

Second, dropping assumption 2 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.

Chapter 2

Aggregate Implications of Capacity Constraints under Demand Shocks (with Florian Kuhn)

We investigate the impact of firm capacity constraints on aggregate production and productivity when the economy is driven by aggregate and idiosyncratic demand shocks. We are motivated by three observed regularities in US data: business cycles are asymmetric, in that large absolute changes in output are more likely to be negative than positive; capacity and capital utilization are procyclical, and increase the procyclicality of measured productivity; the dispersion of firm productivity increases in recessions.

We devise a model of demand shocks and endogenous capacity constraints that is qualitatively consistent with these observations. We then calibrate the model to aggregate utilization data using standard Bayesian techniques. Quantitatively, we find that the calibrated model also exhibits significant asymmetry in output, on the order of the regularities observed in GDP.

2.1 Introduction

Despite their (surprising) symmetry, two very asymmetric patterns have been observed in business cycles. First, big booms occur less often than big busts, that is, large absolute changes in output are more likely to be negative than positive. Second, firm productivity becomes more dispersed in recessions, and the productivity of the worst firms decreases relative to the mean (Kehrig (2011)).

We believe these patterns are closely related to the way firms change the intensity of their production processes over the business cycle. Capacity utilization, or the percentage of potential output that firms produce, seems to be procyclical, as does capital utilization, or the intensity with which firm use their machinery. It has also been noted that, even though average productivity is apparently procyclical, this is no longer the case when we take the changes in capacity utilization into account (Basu et al. (2006)).

In this paper, we propose a theory involving heterogeneous firms facing costs to utilizing previously installed capital. The costs endogenously constrain the capacity each firm utilizes, and these constraints change with economic conditions. We investigate the impact of these capacity constraints on aggregate production and productivity when the economy is driven by aggregate and idiosyncratic demand shocks. We devise a model of demand shocks and endogenous capacity constraints that is qualitatively consistent with these observations. We then use the mentioned stylized facts in combination with data on firms' input utilization rates in order to learn about the relevance of demand shocks for aggregate fluctuations. To assess this, we estimate the model using micro data and standard Bayesian techniques. We then compare to which extent it replicates the described stylized facts. We find that for our preferred calibration, the model exhibits asymmetry on the order of what we observe in the data.

Capacity utilization, and its relationship to business cycles, has been the subject of a long strand of research in macroeconomics. Hansen and Prescott (2005), motivated by the observation that many firms typically have idle capacity, construct a model in which a representative firm chooses capacity one period ahead, and finds that business cycles in such a model are asymmetric. Gilchrist and Williams (2005) investigate the use of "putty-clay" technology by firms, in which firms freely choose their capital a period in advance, and can only utilize that capital during production. They focus on the construction of a tractable aggregate production function, and find that business cycles are asymmetric as well. The model in this paper is closely related to this work.

This paper proceeds as follows. In the next section, we document our three stylized facts. Next, we propose our model, and discuss its important properties. In the next section, we estimate the key parameters of our model, simulate the model using these parameters, and discuss its quantitative performance. The last section concludes.

2.2 Three stylized facts relating to asymmetry of business cycles

Large deviations from trend in output are more likely to be negative The question whether business cycles are asymmetric is fairly old. As noted by McKay and Reis (2008), Mitchell (1927) characterized recessions as briefer and more violent than expansions. Starting with Neftci (1984) as well as DeLong and Summers, a large literature has investigated this question using more formal econometric techniques.

In general, results in the literature are mixed. Neftci (1984) finds evidence for asymmetry between increases and decreases of the US unemployment rate estimating an underlying Markov model. DeLong and Summers (1986) test skewness coefficients of output and employment data for the US and other industrialized countries. They do not find significant skewness in growth rates of output, nor significant differences in the length of expansions and contractions in output. Hamilton (1989) extends Neftci's work and estimates a two-state Markov process for output in which the recessionary and expansionary states have significantly different properties, a finding confirmed by later papers using similar methodology (Clements and Krolzig (2003), Hamilton (2005)). Sichel (1993) takes up DeLong and Summers (1986) nonparametric approach by looking at deepness of contractions and finds evidence of negative skew in levels of output and employment. Bai and Ng (2005) derive limiting distributions for coefficients of skewness and kurtosis under serial correlation and cannot reject the null of zero skew in US output growth rates. After nonparametrically estimated business cycle turning points, McKay and Reis (2008) compare average duration and growth during expansions and recessions, and do not find significant evidence that recessions are shorter or more violent.

These findings suggest to us that one should be specific in defining which aspect of potential asymmetry one is looking at, a point also made by McKay and Reis (2008). Our general reading of the literature is as follows: Evidence for skewness in output growth rates, which seem to be considered in most papers, is largely absent. There is some evidence for skewness in output levels. (Along both dimensions, employment seems to exhibit stronger asymmetry than output.) Additionally, for hidden-state Markov models, non-linear parameterizations fit the data significantly better than linear ones.

Our focus is on the claim that large cyclical swings in output are more likely to be negative than positive — in other words, production will spend more time far below trend than far above. This is related to negative skewness of output levels, for which there is some evidence (Sichel (1993)).

We now report some additional observations about the largest absolute deviations of the cyclical component of output. To this purpose, we take a measure of US production, detrend it, and then look at the largest observations of the detrended series in absolute value. Specifically, for some integer N we count in how many of the N periods of largest absolute deviations output was above trend vs below trend. We also compare the mean of the $N/2$ most positive deviations to the $N/2$ most negative deviations. Finally, we also compute the coefficient of skewness for the detrended series as a whole (so that this statistic is independent of N). We repeat this process for a number of different specifications, where a specification consists of a measure of output, a time span, a trend filter, and an integer N.

We report results in table 2.1. Our baseline cases use HP-filtered postwar data and often constitutes the weakest case in terms of differences between expansions and recessions since the HP filter tends to attribute parts of the cyclical movement into the trend at the edges of the sample. For almost all specifications we find that large deviations from trend tend to be negative.

The following discussion of the other two stylized facts is shorter, mainly because the evidence is not as mixed.

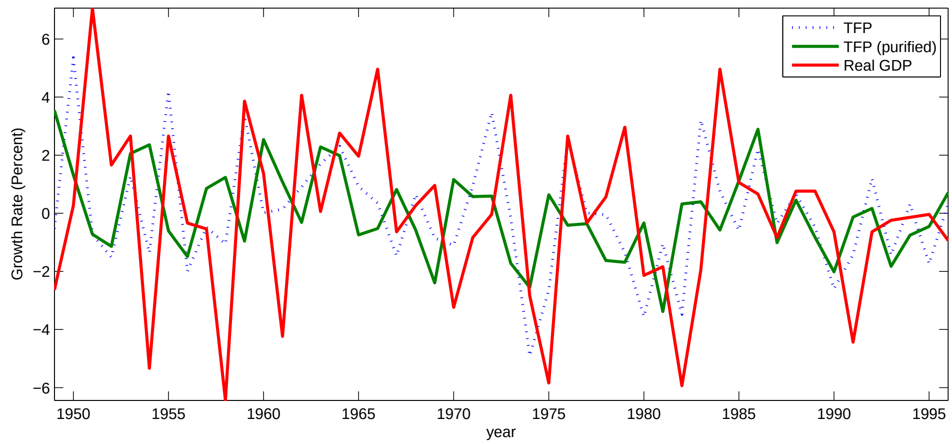
Simple TFP estimates are procyclical, but not so if corrected for utilization For this stylized fact we draw on Basu et al. (2006) and Basu et al. (2009) who discuss ways to improve the measurement of firm productivity. In particular, they construct a measure for aggregate technology that accounts for potential confounding influences of returns to scale, imperfect competition, aggregation across sectors and, especially relevant for us, utilization rates of factor inputs. Figure 2.1 displays annual output growth together with changes in a simple measure of TFP (the Solow Residual) as well as Basu, Fernald and Kimball's

Table 2.1: Strong Recessions are more Severe than Strong Expansions

Specification	# pos vs neg	mean pos vs neg	skewness
Quarterly GDP			
Baseline	16 vs 24	2.78% vs −3.68%	−0.60
$N = 20$	6 vs 14	3.12% vs −4.33%	−0.60
$N = 80$	40 vs 40	2.28% vs −2.87%	−0.60
until 2007	17 vs 23	2.28% vs −2.87%	−0.62
Linear filter	6 vs 34	7.99% vs −12.70%	−0.81
Rotemberg filter	6 vs 34	4.19% vs −5.68%	−0.33
Rotemberg filter, $N = 80$	29 vs 51	3.74% vs −5.14%	−0.33
Annual GDP			
Baseline	3 vs 7	3.20% vs −4.40%	−0.35
$N = 6$	0 vs 6	3.37% vs −4.83%	−0.35
$N = 20$	13 vs 7	2.99% vs −3.55%	−0.35
until 2007	4 vs 6	2.99% vs −3.55%	−0.35
from 1929	6 vs 4	16.69% vs −11.61%	+1.00
Linear filter	2 vs 8	7.29% vs −12.51%	−0.88
Linear filter from 1929	3 vs 7	20.50% vs −31.08%	−0.91
Rotemberg filter	1 vs 9	6.23% vs −13.50%	−0.87
Rotemberg filter from 1929	1 vs 9	16.15% vs −36.95%	−1.22
Monthly industrial production			
Baseline	50 vs 70	4.52% vs −5.90%	−0.65
$N = 40$	7 vs 33	5.48% vs −7.57%	−0.65
$N = 240$	124 vs 116	3.71% vs −4.45%	−0.65
until 2007	56 vs 64	4.39% vs −5.58%	−0.65
from 1919	54 vs 66	11.35% vs −13.59%	−0.55
Linear filter	33 vs 87	17.03% vs −22.69%	−0.52
Rotemberg filter	46 vs 74	7.47% vs −11.23%	−0.62

"# pos vs neg": Out of the N periods with largest absolute value, how many were positive and how many were negative. "mean pos vs neg": Mean of the $N/2$ largest periods vs mean of the $N/2$ smallest periods. "Skewness": Coefficient of skewness defined as $E\left[(x-\mu)^3/\sigma^3\right]$.
For all three series in the baseline, N correspons to a little less than 1/6 of observations, series were HP filtered and starting date is January 1949. "Quarterly GDP": $N = 40$, end date $2014:4$, HP(1600)-filtered. "Annual GDP": $N = 10$, end date 2013, HP(400)-filtered. "Monthly industrial production": $N = 120$, end date 2014/02, HP(10,000)-filtered. Alternative specifications differ from respective baseline only along listed dimensions.

Figure 2.1: GDP and TFP measures



Note: Real GDP series from FRED, along with TFP series and purified TFP series from Fernald (2012).

'purified' measure. The simple productivity measure is strongly procyclical: Correlation between output growth and simple TFP is 0.74. The improved technology measure does not exhibit this association with aggregate production; in fact purified TFP appears to be almost completely acyclical as its correlation with (contemporary) output growth is 0.02.

Since the mechanism we consider hinges strongly on the effect of adjustment in factor input utilization, we recalculate the mentioned coefficients of correlation using data provided by John Fernald[1] (see Fernald (2012)). This dataset provides among other things a TFP measure that *only* corrects for intensity of capital and labor utilization, allowing us to check if utilization is indeed relevant for the difference in cyclicality between the simple and the purified productivity measure (or if instead the difference stems mainly from the other 'purifying' steps taken by Basu, Fernald and Kimball). Additionally, it spans 15 more years at the end of the sample. Again, simple TFP is strongly procyclical with a correlation

[1]Data available at www.frbsf.org/economic-research/economists/jfernald/quarterly_tfp.xls

of 0.83 whereas utilization-corrected TFP is acyclical with a coefficient of -0.03.

Dispersion of firm productivity increases in recessions Our third fact is connected to a range of findings in the literature that relate recessions to increased cross-sectional dispersion among firms along several dimensions. Eisfeldt and Rampini (2006) show that capital productivity is more dispersed in recessions. Bloom (2009) and Bloom et al. (2012) show in seminal papers that shocks to the variance of firm productivity can cause drops in output; they include empirical evidence relating dispersion in sales growth, innovations to plant profitability, and sectoral output to times of low aggregate production.

Directly related to levels of firm productivity, Kehrig (2011) finds that the distribution of plant revenue productivity becomes wider in recessions; Bachmann and Bayer (2011) reach a similar result for innovations to the Solow residual in a dataset of German firms.

2.3 Model

We construct a simple monopolistic competition model in which a final good is produced with a variety of intermediate inputs. Each of the intermediate input producers must choose the price of their good and their capital capacity one period ahead of when they decide to produce. At the time of production, these firms are free to choose their labor input and *utilized* capital input. A convex cost of capital utilization ensures that firms choose interior values for their utilization rates.

This is a partial equilibrium model, in which wages, rental rates and, importantly, demand functions are independent of the outcomes in the intermediate goods sector.[2] Im-

[2]This setup could, for example, be rationalized as part of a general equilibrium in which there is a global final goods aggregator, a global household, and the intermediate goods sector represents a country's firms that jointly have mass 0 in final goods production.

portantly, this shuts down an interaction between binding constraints on intermediate goods firms and aggregate demand: An increasing share of constrained intermediate goods producers will potentially increase demand for other, non-constrained firms' goods and will also lower final goods production. Both these effects can have feedback effects on the share of constrained firms and thus deprive the Dixit-Stiglitz solution of its usual simplicity. For this reason we leave generalization to general equilibrium for future work.

2.3.1 Final Goods Firms

There is a competitive sector of final goods producers which acts as a CES aggregator, that is, production of final goods Y_t requires purchasing a basket of intermediate inputs of differing varieties $\left\{y_{it}^d\right\}_i$ at prices $\{p_{it}\}_i$. Each final goods firm's production function is given by

$$Y_t = \left[\int (A_t b_{it})^{\frac{1}{\sigma}} {y_{it}}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}} \tag{2.1}$$

where σ is the (constant) elasticity of substitution between inputs. A_t and b_{it} are random (AR(1)) aggregate and variety-specific final good technology shocks, respectively. We will alternatively refer to these as "demand shocks" throughout the remainder of the paper. Under the assumption that b_{it} is lognormally distributed, which we will maintain throughout the paper, the aggregate shock A_t just shifts the mean of b_{it}.

As will be made clear below, because of capital utilization costs, intermediate goods firms will not be willing to supply more than a (variety specific) quantity y_{it}^C. Also, the price of final goods is normalized to 1. Thus, in each period t, final goods firms solve

$$\max_{(y_{it})_i} Y_t - \int_i p_{it} y_{it} di$$

Define the expenditure for varieties $I_t = \int_i (A_t b_{it})^{\frac{1}{\sigma}} p_{it} y_{it} di$, and price index $P_t^{1-\sigma} = \int_i (A_t b_{it})^{\frac{1-\sigma}{\sigma}} p_{it}^{1-\sigma} di$. Then

$$y_{it}^d = A_t b_{it} \frac{I_t}{P_t} \left(\frac{P_t}{p_{it}} \right)^{\sigma}$$

2.3.2 Intermediate Goods Firms

Intermediate goods producers transform capital and labor supplied by households into differentiated products that can be used as inputs by the final goods sector. Each producer monopolistically supplies a single variety.

We depart from the standard model in three ways. First, each intermediate goods firm chooses the price of its variety one period in advance. Second, also one period in advance, each firm invests some of its earnings (in units of final goods) as installed capital for production in the next period. Third, firms pay a cost this period for utilizing installed capital that is increasing in the fraction of installed capital used.

We will allow firms to utilize capital beyond installed capital this period, so that the utilization rate is bigger than 100%; firms are essentially able, albeit at great cost, to "overwork" the machines installed in the factory last period as much as the care to. For the cost functions and parameterizations we consider, the cost of utilizing capital keeps firms from overworking installed capital in any period.

Specifically, call installed capital by the producer of intermediate input i in period t k_{it}, and the actual capital used in period t $\tilde{k}_{it}$, so that $\tilde{k}_{it} \leq k_{it}$. Then the producer of intermediate input i has a production function given by

$$y_{it} = \tilde{k}_{it}^{\alpha} l_{it}^{1-\alpha} \tag{2.2}$$

Installed capital is rented at the end of period $t-1$ at the interest rate R_t, but producers do not repay households until the end of period t, as in the standard case. Capital depreciates by the rate δ after it is used, and firms maintain ownership of the undepreciated capital. Capital utilization comes at a cost given by the function $c\left(\tilde{k}, k\right)$. Thus, each firm i's profit in period t is given by

$$\pi_{it} = p_{it}y_{it} - w_t l_{it} - R_t k_{it} - c\left(\tilde{k}_{it}, k_{it}\right) + (1-\delta)k_{it} \tag{2.3}$$

We will assume that the capital utilization cost function is quadratic in the utilization rate:

$$c\left(\tilde{k}, k\right) = \chi \left(\frac{\tilde{k}}{k}\right)^2 k \tag{2.4}$$

Firm i maximizes the sum of per period profits, but since the only buyer of variety i is the final goods producer, the firm cannot supply any more than y^d_{it}. The firm's maximization problem is given by

$$\begin{gathered} \underset{(p_{it},k_{it})_{t=0}^{\infty}}{max} E_t \sum_{i=1}^{\infty} \Lambda_{t,t+i} \underset{(\tilde{k}_{it},l_{it})}{max}\, \pi_{it} \\ s.t.\; y_{it} \leq y^d_{it} \end{gathered}$$

We solve the firm's problem in three steps. First, find the firm's choice of $(\tilde{k}_{it}, l_{it})$ that minimizes its cost of supplying y_{it} given $(p_{it}, k_{it}, A_t, b_{it})$, where A_t and b_{it} are (random) aggregate and variety-specific demand shocks. Second, find the firm's choice of y_{it} that maximizes its profits given $(p_{it}, k_{it}, A_t, b_{it})$. Third, find the choices of (p_{it}, k_{it}) that maximize the firm's expected profits, where expectations are taken over the random shocks (A_t, b_{it}).

Firm Costs and Factor Demands

The firm chooses how much labor and capital to use in period t to minimize its period t costs.[3] The rental costs of installed capital are sunk at the time t. Thus, given $(p_{it}, k_{it}, w_t, R_t, b_{it}, A_t)$ and a desired level of output y, the firm chooses $(\tilde{k}_{it}, l_{it})$ to minimize

$$\min_{(\tilde{k}_{it}, l_{it})} w_t l_{it} + \chi \left(\frac{\tilde{k}_{it}^2}{k_{it}} \right)$$
$$s.t.\ \tilde{k}_{it}^{\alpha} l_{it}^{1-\alpha} \geq y$$

The solution to this problem is given by

$$\tilde{k}_{it} = y_{it}^{\frac{1}{2-\alpha}} \left(\frac{w_t}{2\chi} \frac{\alpha}{1-\alpha} \right)^{\frac{1-\alpha}{2-\alpha}} k_{it}^{\frac{1-\alpha}{2-\alpha}}$$

$$l_{it} = y^{\frac{1}{1-\alpha}} \tilde{k}_{it}^{-\frac{\alpha}{1-\alpha}}$$

With these factor demands, we can also find the firm's cost function:

$$C\left(y\right) = w_t^{\frac{2-2\alpha}{2-\alpha}} \chi^{\frac{\alpha}{2-\alpha}} k_{it}^{-\frac{\alpha}{2-\alpha}} \left(\frac{2-\alpha}{2-2\alpha} \right) \left(\frac{2-2\alpha}{\alpha} \right)^{\frac{\alpha}{2-\alpha}} y^{\frac{2}{2-\alpha}}$$

Firm Supply

Using the cost function obtained above, we can state the firm's profit maximization problem and obtain the optimal level of output y_{it}. As the only buyer of variety i is the final goods

[3]Minimizing period t costs minimizes the firm's total costs, since there adjusting these inputs have no effect on costs in other periods.

producer, the firm cannot supply any more than y_{it}^d. The firm chooses y_{it} to maximize[4]

$$\underset{y_{it}}{max}\ p_{it}y_{it} - C\left(y_{it}\right)$$
$$s.t.\ y_{it} \leq y_{it}^d$$

We can now write the solution to the above profit maximization problem. Given $(p_{it}, k_{it}, b_{it}, A_t)$,

$$y_{it} = \begin{cases} y_{it}^d & \text{if } b_{it} \leq \bar{b}_{it} \\ y_{it}^C & \text{else} \end{cases}$$

where

$$\begin{aligned} y_{it}^C =& \frac{\alpha}{2\chi(1-\alpha)}\left(1-\alpha\right)^{\frac{2-\alpha}{\alpha}} p_{it}^{\frac{2-\alpha}{\alpha}} w_t^{-\frac{2-2\alpha}{\alpha}} k_{it} \\ \bar{b}_{it} =& \left(\frac{P_t^{1-\sigma} p_{it}^{\sigma}}{A_t I_t}\right) y_{it}^C \end{aligned}$$

The cutoff $\bar{b}_{it}$ gives the lowest level of demand shock that ensures that the constraint $y_{it} \leq y_{it}^d$ does not bind. For all $A_t b_{it} > \bar{b}_{it}$, the marginal cost of supplying y_{it}^d exceeds the variety-specific price p_{it}, and therefore the firm would earn higher profits by only supplying y_{it}^C. This level of output acts as an endogenously determined limit on output (and therefore capital utilization); though this limit can differ across firms in the general model, in the solution we consider, it will be the same for all firms.

[4]The firm's profit function should also include the cost of repaying capital to the household, rented last period. Because this cost is sunk in period t, however, we ignore that cost in solving the period t output decision.

Firm Expected Profit

With the firm's supply function in hand, we can write down the state-by-state profit function, along with the expected profits the firm will earn given a choice of (p_{it}, k_{it}) and information about previous period random shocks $(A_{t-1}, (b_{i,t-1})_i)$.

After uncertainty is realized, profits (less capital rental costs) are given by

$$\pi_{it}^n(p_{it}, k_{it}, b_{it}, A_t) = \begin{cases} p_{it}y_{it}^d - C\left(y_{it}^d\right) & \text{if } b_{it} \leq \bar{b}_{it} \\ p_{it}y_{it}^C - C\left(y_{it}^C\right) & \text{else} \end{cases}$$

Thus, the firm chooses p_{it} and k_{it} to solve

$$\underset{p_{it}, k_{it}}{max}\ E_{t-1}\pi_{it}(p_{it}, k_{it})$$

where the profit agents expect to receive in period t, looking from the end of period $t-1$, is

$$E_{t-1}\pi_{it}(p_{it}, k_{it}) = \int \left(\int_0^{\bar{b}_{it}} \left[p_{it}y_{it}^d - C\left(y_{it}^d\right)\right] p_b(x)dx + \int_{\bar{b}_{it}}^{\infty} \left[p_{it}y_{it}^C - C\left(y_{it}^C\right)\right] p_b(x)dx \right) dF_A$$

2.3.3 Solution

We begin by putting structure on the random shocks b_{it} and A_t. The distribution of the random shock A_t follows an AR(1) process with normal error, while the distribution of the idiosyncratic shock b_{it} follows an AR(1) process with lognormal error:

$$b_{it} = \rho_b b_{i,t-1} + \epsilon_{bt}$$

$$ln(A_t) = \rho_A ln(A_{t-1}) + \epsilon_{At}$$

$$\epsilon_{bt} \underset{d}{\sim} ln\mathcal{N}(\mu_b, \sigma_b)$$

$$\epsilon_{At} \underset{d}{\sim} \mathcal{N}(\mu_A, \sigma_A)$$

In the estimation and simulation that follow, we make two assumptions on the structure of the demand shocks A_t and b_{it}:

- $\rho_b = 0$ (aggregate errors are normal, idiosyncratic errors are lognormal)
- $\mu_b = -\frac{\sigma_b^2}{2}$ (normalize mean of shock to 1)

We solve the full set of equilibrium equations under the above assumptions in the appendix. Importantly, the iid assumption implies that the one-period ahead prices and installed capital choices p_{it}, k_{it}, and $\overline{b}_{it}$ are identical across firms in every period. (We'll interchangeably call these common choices p_t and k_t.) In turn, this implies that the endogenous capacity constraint, y_{it}^C, and the corresponding demand shock cutoff $\overline{b}_{it}$, are identical for all firms.

Now consider the effect of an increase in the aggregate shock A_t. First, the cutoff $\overline{b_t}$ doesn't depend on the aggregate shock A_t. This implies that when we increase A_t, essentially increasing the mean demand shock faced by any firm, we necessarily make more firms capacity constrained.

We can now show that the model qualitatively matches the second two stylized facts described above, though we do not lay out the proofs in detail here. The second fact states that while measured TFP is procyclical, "purified" TFP (controlling for factor utilization) is less procyclical. The analog for both of these quantities is

$$\text{Measured TFP: } \frac{y_{it}}{k_{it}^{\alpha} l_{it}^{1-\alpha}} = \left(\frac{\tilde{k}_{it}}{k_{it}}\right)^{\alpha}$$
$$\text{Purified TFP: } \frac{y_{it}}{\tilde{k}_{it}^{\alpha} l_{it}^{1-\alpha}} = 1$$

The first quantity is just a function of the capital utilization rate; using the aggregate (mean) capital utilization rate from the appendix, it can be shown that the derivative of the above function with respect to A_t is positive, and therefore increasing in A_t. Since we use A_t to stand-in as the underlying source of business cycle fluctuations, we can say that this quantity is procyclical. The second quantity, on the other hand, is totally acyclical.

The third fact states that productivity dispersion increases in recessions. One way to show this is to show that the coefficient of variation of measured TFP is an decreasing function of A_t, where

$$CV_{TFP} = \frac{\sigma_{TFP}}{\mu_{TFP}} = \frac{\int \left(\frac{\tilde{k}_{it}}{k_{it}}\right)^{2\alpha} di}{\left(\int \left(\frac{\tilde{k}_{it}}{k_{it}}\right)^{\alpha} di\right)^{2}} - 1$$

Again, though we do not show it here, the derivative of the above function with respect to A_t is negative, and therefore decreasing in A_t.

2.4 Estimation

In the previous section, we showed that the model qualitatively implies a connection between capacity utilization and business cycle asymmetry. We now want to investigate to which extent the capacity and capital utilization that we observe in the data is quantitatively relevant, in that it implicity generates a significant portion of observed business cycle asymmetry. We do this by first estimating model parameters only with the capacity and capital utilization data, and then simulating the calibrated model. In this section, we outline the data we use for the calibration, then outline our calibration procedure and results.

2.4.1 Data

We use two aggregate time series on the manufacturing sector for the estimation: a measure of capacity utilization and capital utilization. All data are annual, and we restrict our attention to the range 1974-2004 for which the capital utilization data is available. We obtain aggregate capacity utilization across surveyed firms from FRED[5] from the Quarterly Survey of Plant Capacity Utilization conducted by US Census Bureau (2014). The data surveys plant managers across several sectors of the economy. Managers are asked to estimate the current market value of their production as a percentage of the value of production that would occur if plants were operating at "full capacity", i.e. using all machinery in place, and as much labor/fuel/etc. as necessary to operate it.

We obtain capital utilization data from Gorodnichenko and Shapiro (2011)[6], using their preferred measure of plant hours per week. This measure uses information on labor hours to estimate how intensively a plant's machines are being used.

For the output measure we use data on real GDP from FRED[7]. We detrend the annual production series using an HP(100) filter. Since plant hours per week and capacity

[5] Available at https://research.stlouisfed.org/fred2/series/MCUMFN
[6] Available at http://www-personal.umich.edu/ shapiro/data/SPC/
[7] Available at https://research.stlouisfed.org/fred2/series/IPMAN

utilization are naturally bounded from both sides, it is not clear that they should be detrended. We choose to HP-filter them with a high penalty parameter of $10,000$, in order to remove very long-run trends. Finally, for output and workweek of capital, we use log deviations from trend since the levels of these variables are not informative in the context of our model. However, we continue to express capacity utilization in levels, since the level of its long-run average of around 78% corresponds naturally to its equivalent in the model.

2.4.2 Estimation Procedure

We estimate four parameters: the cost parameter χ, the variance of the idiosyncratic demand process σ_b, as well as persistence and variance of the aggregate demand shock ρ_A and σ_A, respectively.

Given a set of parameters (χ, σ_b), and the mean wage w_t, we use the model to calculate mean capital utilization and capacity utilization. We then match this against the capital utilization and capacity utilization from the data.

For capacity utilization we refer to full capacity as y_{it}^C, the maximum output an individual firm is willing to supply when demand is very high. While a more natural counterpart to the survey question would be to compute output at some level of maximum capital utilization, it is not entirely clear what this would mean in the context of the model.

We obtain our estimates using a standard Bayesian approach; the procedure (see Fernández-Villaverde (2010)) is described in more detail in the appendix. The results are displayed in table 2.2. The first three parameters are chosen as standard in the literature. For the elasticity of substitution σ, the estimates in the literature range from around 4 to 10. For now we choose a value on the higher end of estimates (giving firms less monopolistic power), and note that it would be a good idea to include the parameter in future

Table 2.2: Parameters

Parameter		Value
Set outside estimation		
Effective Capital Share	α	0.33
Discount Factor	β	0.96
Depreciation Rate	δ	0.1
Elasticity of Substitution	σ	10
Autocorr. Demand Shock	ρ_b	0
Estimates		
Utilization Cost	χ	0.299
Std. Dev. Demand Shock	σ_b	4.67
Autocorr. Agg Shock	ρ_A	0.496
Std. Dev. Agg Shock	σ_A	0.041

estimations. The last parameter set outside the estimation is ρ_b, the persistence of the idiosyncratic demand shock. In the context of the current model, this is without loss of generality: since firm decisions are linear in expected demand, the distribution of $E[b_i]$ among firms does not matter (nor are we using data identifying this distribution). The next four parameters represent the median values of their repective marginal posterior distribution; we give further information about the posteriors in the appendix. While the variance of the idiosyncratic demand shock may seem high, it is worth keeping in mind that the mean of the distribution over b_i is kept normalized to 1 (the lognormal distribution has parameters $(-\sigma_b^2/2, \sigma_b)$). This means that for most firms the realization of their demand shock actually becomes smaller as σ_b increases.

2.4.3 Results

We now assess to which extent the model reproduces the three stylized facts under the calibrated parameters found in the previous section.

To do so, we use the the median estimates for χ and σ_b and compare the effects of symmetric aggregate demand shocks on output, aggregate (non-corrected) TFP, and

Table 2.3: Quantitative Performance

Business Cycle Asymmetry		
% Δ from SS level	20% change in A	-20% change in A
ΔKU	0.87	-0.92
ΔY	1.12	-1.19
	Increase	Decrease
Shock needed for 1% Change in Y	18	-17
Shock needed for 5% Change in Y	93	-88

	Skewness in Y
Model	-0.37
Data: Rotemberg Filter	-0.33
Productivity Dispersion	
Dispersion	Correlation with Y
Measured TFP Variance (model)	-0.98
TFPR Variance (data)	-0.5

Note: Performance of the model with respect to business cycle asymmetry and productivity dispersion. First, we list percent changes from steady state for production-weighted aggregate capital utilization (KU), capacity utilization (YU), and output (Y). Next, we list the shock to the aggregate shock A that would be needed to produce a 1% increase and decrease in Y, respectively, and then again for a 10% change. Next, we list the skewness of a simulated output series given the standard deviation for aggregate shock process A that generates the observed standard deviation of output. Last, we list correlation of productivity dispersion with output, both for a measured TFP in the model and estimates from Kehrig (2011).

dispersion in average costs of production.

The quantitative comparison results are given in table 2.3. We only report demand-weighted aggregates, to reflect final output as the product of the final goods aggregator. We first consider the impact of a change in the aggregate demand shock A, which corresponds to a change in the mean of the idiosyncratic demand shocks b_i. When A increases, more firms have higher demand shocks, but by the same token, more firms are producing at the endogenous output limit y_t^C.

We first shock the model with a shock equal to the calibrated standard deviation of $\sigma_{A_Y} = 0.2$. Generally, even for a large change in A, the model delivers small changes in all aggregate variables. The first three rows of the table show this: a 20% shock to aggregate demand relatively modest changes in output, capital utilization, and capacity utilization. We believe these magnitudes have something to do with the large estimated σ_b - since the distribution of demand shocks is already wide with $A = 1$, a small change in A doesn't affect the demand of many firms.

That said, when the model is calibrated to match the magnitude of observed changes in output, it does exhibit asymmetry in output. Given the same percentage magnitude aggregate shock, output decreases (as a percentage of its steady state value) by more than it increases (1.19% decrease vs 1.12% increase). This difference is small, but as the next rows show, it does have significant implications for the frequency of large busts versus large booms.

Under the calibration for the aggregate shock process A_t, we know the distribution of possible aggregate shocks. We can use the calibrated parameters to simulate the model for several demand shocks and calculate the implied skewness in aggregate output in the model. We see skewness of output in the model is on the order of what we see in the data; for example, the skewness in the filtered quarterly GDP data is a little less than the skewness implied by the model.

The model also seems to predict more skewness at higher levels of volatility. In the second two rows, we list the level of A_t necessary to induce a 1% increase and decrease in GDP, and then a 5% increase and decrease in GDP. At the 1% level, the necessary aggregate shocks are very close together in magnitude, only differing by a percentage point. At the 5% level, the difference between the negative shocks increases to 5 percentage points. We take this to imply that the utilization effects we capture in the model may be more pronounced for large changes in output.

Finally, the model also predicts significantly higher TFP dispersion in busts. Using the formula

$$\widetilde{TFP}_{it} = \frac{y_{it}}{k_{it}^{\alpha} l_{it}^{1-\alpha}} \tag{2.5}$$

We calculate the variance of $\widetilde{TFP}_{it}$ for many draws of the aggregate shock with $\sigma_A = 0.2$. We then calculate the correlation between the variance of TFP and output; we find it to almost perfectly negatively correlated. The corresponding estimate from Kehrig (2011) is much (unsurprisingly) significantly smaller, on the order of 0.5.

As shown above, the analog of "purified" TFP will always be acyclical in this model, while simple TFP will be procyclical, so we cannot meaningfully add to our model's performance on this measure in this quantitative exercise. We can speak to the model's general performance, however, by comparing the implied correlation between output growth and simple TFP.

2.5 Conclusion

We present a model of demand shocks which can qualitatively explain three business cycle observations: Deep recessions, procyclicality of the Solow residual while purified TFP is acyclical, and countercyclical dispersion in firms' Solow residuals and average costs. The main assumption is that firms set prices before the level of their demand is realized. We then demonstrate how to estimate the model using time-series data on the level of capacity utilization as well as the comovement of output, capacity utilization and capital utilization.

Further research is needed to derive the general equilibrium solution to the model as well as adapting the model to allow for a more informed estimation using more data.

Chapter 3

Technology Spillovers and Local Interactions

This paper explores the role of distance in selection of equilibrium in an economy with two steady-state Pareto-ranked equilibria, L and H. The economy has a finite number of sectors. In each one, a high productivity and a low productivity technology are available for use in production. The high productivity technology requires a fixed set up cost for production. Sectors are linked by localized production complementarities. I consider selection under a learning rule in which agents imitate their most successful neighbor. As distance between neighbors decreases, the possible profits from industrialization increase, and the likelihood that the learning rule process converges to a steady state matching the H equilibrium increases. The result suggests that, in the presence of localized technology spillovers, there may be important gains to economic growth from infrastructure development.

3.1 Introduction

Coordination failures have been identified as a source of friction in several areas of economics, and development is certainly no exception. Poverty traps, inefficiently high rent-seeking, and persistent, poor institutional regimes have all been identified as possibly serious obstacles to development in the last decade. If we take it for granted that a country is in a bad equilibrium, but a better one is possible, what can policy makers feasibly change to shift the country to the good equilibrium?

One answer to this last question has been transportation and communication infrastructure, or more generally, the cost per unit distance of transporting goods or knowledge. Intuitively, better roads mean goods can come from farther away with less risk of damage or delay; better phones mean access to more efficient techniques and expertise. Investment in transportation and communications networks are usually the province of public agents (i.e. national governments), and therefore represent a viable policy tool for escaping a bad equilibrium, assuming they do matter. At the same time, the impact of these networks on economic growth, let alone on equilibrium selection, is not very well understood. And despite these intuitively positive implications of "being connected to more people" for growth, most infrastructure investment of this kind has historically been made as a *response* to increasing growth, not the result of a sustained public policy effort to increase it (Banister and Berechman (2000)).

This paper investigates the role that "being connected to more people" might play in guiding society to one equilibrium over another when more than one exists. I want to answer the question: can poor transportation and communication account for the presence of coordination failures? I examine a simple model of industrial production with two steady state equilibria and find that as distance increases, the probability that any initial spatial configuration of industrial producers of a society will converge to a high income equilibrium decreases. If we take it for granted that economies are constantly shifting this spatial configuration, this result suggests that countries are more likely to find the "right path" and head toward a good equilibrium as networks improve (the effective distance between you and your neighbors decreases).

In order to understand the role that distance plays in equilibrium selection, I need two components that are not very well studied in the coordination failure literature: a means by which society selects an equilibrium (dynamics), and variation in the effects of agents' decisions over space (local interactions). To address the problem of equilibrium selection, I utilize a set of dynamics proposed in Eshel et al. (1998). I use a myopic, "imitation" decision rule, where agents choose next period decisions to match the decisions of their most profitable neighbors. Agents can choose between one of two techniques for production of a pre-determined good. In every period, each agent observes the profits earned by all other agents in her neighborhood, and calculates the average profit earned by the agents using each of the two techniques. Agents choose the technique next period that yielded the highest average profit for the other agents in the neighborhood this period. By using this rule, I implicitly assume that imitation is an important part of decision making in this setting. The steady states of the process induced by this rule will point to the equilibrium that will be selected under more general dynamics.

To address the issue of distance, I append a currently existing model containing multiple equilibria with a special, "local" structure. Specifically, I combine a model of industrial coordination failure, Murphy et al. (1989), with a spatially limited demand spillover. The idea is simple: agents can only work and buy goods in their neighborhood, an area that lies within a certain radius of their home. Demand for the goods in their neighborhood increases when more people in their neighborhood work in a factory, and so as more factories are built in a neighborhood, the incentives for a new factory to be built in the same neighborhood increase.

Coordination failures in industrial development have been the subject of several papers in the macroeconomics literature. Keynes popularized the notion as a central idea in his explanation for why prices sometimes adjust too slowly for markets to clear; more recent work (including Murphy et al. (1989), which motivates this paper) has formalized this idea in a variety of models. As noted in Cooper (1994), formalizing this notion requires an imperfectly competitive market, since otherwise flexible prices clear markets by the First Welfare Theorem. For example, in the economy proposed in Murphy et al. (1989), agents in a multi-sector economy choose between the use of two production techniques. The market imperfection here is an assumption of possible monopoly if one of the techniques is selected, essentially fixing prices for the goods produced.

Equilibrium selection in these coordination failure models is not as well understood. Cooper (1994) studies the dynamics of equilibrium selection in a model economy with multiple equilibria and a stochastic aggregate state variable that determines payoffs. He imposes a decision rule which has agents use the previous period value of the state variable as a prediction for the value of the variable this period. Equilibria in this economy, perhaps unsurprisingly, are persistent in the sense that once a particular equilibrium occurs, that equilibrium continues to occur until a large change in the state variable occurs. Durlauf (1993) examines a multisector model economy which allows multiple equilibria and a technology spillover that can be limited in scope with respect to other sectors. The restriction is a fully general version of the case in this paper in the sense that it allows arbitrary spillover dependence between sectors. The spillover takes the form of a productivity shock conditional on the technology choices in each sector. Among other results, the author showed that multiple long-run equilibria persist if each industry is sufficiently sensitive to the choices of other industries. This sensitivity could be interpreted to be the size of a neighborhood in my model, and therefore the persistence of multiple equilibria as sensitivity increases seems to contradict the findings of this model.

With respect to the theory of equilibrium selection in more general coordination games, Kandori et al. (1993) and Ellison (1993) study the selection among equilibria of a repeated 2 by 2 coordination game. In the first paper, the authors find that, given myopic learning and random experimentation, the absorbing set of the process includes only one equilibrium, reducing the multiple equilibria of the original game to just one. The second paper studies the rate of convergence to this equilibrium, and finds that large populations have a slower rate of convergence than small ones. In terms of equilibrium selection, the dynamics induced by myopia and mutation do a worse job of selection as the number of experimenters increases. In Galeotti et al. (2010), one result among others seems to jibe with the main result of this paper, that is, that if agents connected in a network face a strategic complementarity in their interaction, their level of participation in that interaction increases as the number of their neighbors increases.

Though less often used in the literature, imitation seems to be an important component of deciding whether to adopt a new technology. The psychic/social costs of new technology adoption are smaller when a neighbor has already adopted a new technology. A farmer is more likely to implement a new procedure that a neighbor currently uses, as some of the uncertainty about the benefits to the technique (and I would also argue the possibility of the use of this technique) is reduced (Evenson and Westphal (1995)). Moreover, in a model of local interactions, it makes sense that the set of technologies one can choose to adopt should come from a spatially restricted area; the imitation rule ensures that this is always the case. From a modeling perspective, the imitation rule buys us two positive qualities. First, it is simple, and tracking the evolution of even a large society is relatively easy with the aid of a computer. Second, as noted in Eshel et al. (1998), imitation allows cooperative outcomes in an otherwise non-cooperative setting without having to rely on repeated interactions between agents. In the model below, agents only live for one period, but we see the "cooperative", high income equilibrium arise in several cases.

My basic model starts with a simple adaptation of the "wage premium" model of Murphy et al. (1989) outlined in Basu (1997). Broadly, it consists of an N-person coordination game repeated an infinite number of times. Section 2 outlines the coordination game played every period and provides the preliminary result that positive profits are more easily achievable with larger neighborhoods, and section 3 outlines the structure of the repeated game and the solution concepts we are interested in. Section 4 sketches a possible empirical exercise and concludes.

3.2 Model - Stage Game

Consider an economy composed of N identical towns arranged on a circle, so that each town has exactly two other towns near it. Each town, denoted by i, is composed of a single industry and a large number of consumer-workers. Time is discrete, and each period is denoted by t. The firms and consumer-workers in each town i have limited information about the state of other towns. The industry of each town is devoted to the production of a single good, but goods vary across towns; in particular, the industry of each town produces one of K goods. Each town i is part of a "neighborhood" n_i, which contains a subset of all the towns of the world. Finally, I restrict N and assign good production to each town so that in any neighborhood, each of the K good varieties can be found. I state these assumptions more formally below.[1]

[1] The restriction on N affects the dynamics of the model, but it significantly simplifies the structure of demand. As I will show below, if all K goods are available in any neighborhood, and we imbue agents with the same preferences over all K goods, the demand of an agent for any good is identical across towns and goods. Without this assumption, some neighborhoods would not include some good varieties, which would cause some agents to be always worse off than others because their neighborhood does not include the same set of goods as another. I leave this interesting general case for future work.

Assumption 1. K is an odd positive integer.

Assumption 2. If N is the number of towns in the economy, then there is a positive integer h such that $N = hK$.

Definition 1. K is the number of total goods in the economy, with K odd, and N is the total number of towns in the economy. Then the neighborhood of town i, $n_i(K)$, is a set of towns that include town i and its $K-1$ nearest neighbors, so that

$$n_i(K) = \{i - \frac{K-1}{2}, ..., i, ..., i + \frac{K-1}{2}\}$$

Assumption 3. In a neighborhood of any town i, $n_i(K)$, there is exactly one town which produces each of the varieties $j \in (1, 2, ..., K)$.

To illustrate this structure, imagine the production of the burnt-orange fabric used in my Texas football t-shirt. The t-shirt was manufactured by the Co-op, a UT affiliated seller. The Co-op, as many other apparel manufacturers do, purchased the basic fabric from a textile mill. Textile production requires a knitter, curer, and dyer, and these functions need not occur under the same roof, as it often doesn't. Though the Co-op can choose to purchase the fabric anywhere, transport costs are smaller when these intermediate inputs are produced closer to the Co-op's manufacturing plants. Thus, the Co-op requires a knitter, curer, and dyer within a "neighborhood", that is, a radius outside of which transport costs are too large for supplying these goods to be profitable. A knitter in that neighborhood understands the complementarity between her service and the service provided by the curers and dyers.

3.2.1 Firms

Each industry is composed of a single large firm (alternatively, "industrial producers") and a competitive fringe of small, identical firms (alternatively, "cottage producers"). The large firm differs from the small firms only in the technology it uses for production. Both technologies require labor input at the beginning of the period to produce goods at the end of the period.

I place three important spatial restrictions on firms in the model. First, all goods and services are "local", that is, the labor demanded by any firm in town i can only be supplied by the consumer-workers in the neighborhood of town i. Second, the total population of consumer-workers in each town i is constant, and in particular equal to 1. (I denote this total population of consumer-workers in each town i at time t as $L_{i,t}$.)

Timing in the model is as follows: at the beginning of each period, the potential owners of the large firm in each town decide whether to produce. If they do, they pay the fixed set up cost F at the beginning of the period, while any cottage producers pay a 0 set up cost. Consumers in each town then buy goods from the towns in their neighborhood, including their own, and production meets demand exactly, so that consumers simultaneously act as workers, earning wage income in the process. At the end of the period, profits are realized, which can be negative if demand fails to cover fixed costs.

Denote the labor demanded by the large firm in industry i in time t as $l_{1,i,t}$, and the labor input by the small firms in industry i in time t as $l_{2,i,t}$. Denote the output of the large firm in industry i in time t as $y_{1,i,t}$, and output of the small firms as $y_{2,i,t}$. These quantities are produced through production functions $f_1 : \mathbb{R} \to \mathbb{R}$, $f_2 : \mathbb{R} \to \mathbb{R}$.

I impose significant restrictions on the forms of these production functions, following Murphy et al. (1989) and Basu (1997). The production functions are both linear, with production by the large firm exhibiting increasing returns to scale, and production by the small firms exhibiting constant returns to scale.

Assumption 4. For any town i at time t, denote output of the large firm by $y_{1,i,t}$ and the aggregate output of the small firms by $y_{2,i,t}$. Output as a function of labor input to both firm types $(l_{1,i,t}, l_{2,i,t})$ is given by the functions f_1, f_2, such that:

$$\begin{aligned} y_{1,i,t} &= f_1(l_{1,i,t}) = \alpha(l_{1,i,t} - F) \\ y_{2,i,t} &= f_2(l_{1,i,t}) = l_{2,i,t} \end{aligned}$$

where $\alpha > 1$.

These technologies and industry structures are intended to represent the adoption of a new, more productive technology in a competitive sector of the model economy. Before a new technology is introduced, there are a competitive number of firms using the old technology. Though the new technology is more productive, it requires a higher fixed cost to set up than the old technology; this difference in fixed costs could represent the training cost associated with operation of the new technology, or the costs of necessary physical infrastructure for the set up of the new technology. Since the new technology has lower marginal cost of production (firms using the new technology can increase their productivity by a factor of α), the first firms to introduce the technology can undercut prices charged by the old technology firms and still enjoy higher profits, essentially because they enjoy monopoly power.

The industry structure above, which is presented without modification from Basu (1997), carries an important consequence for prices in the economy. Since there are a competitive number of small firms, the price of any good i produced in that sector is equal to marginal cost, in this case 1. Large firms have monopoly power, but because both the large and small firms produce the same good, the large firm cannot charge any more than the cottage sector price, since otherwise the small firms could undercut it. On the other hand, the monopoly power of the large firm implies that if it charged any less than the cottage sector price, it could increase profits by increasing its price by a small amount. The same argument holds for each good i. Thus, the price charged for any good in the economy is exactly 1.

3.2.2 Households

Consumers prefer variety, in the sense that they want to consume a basket of each of the K types of goods. I define these preferences more formally in the definition and discussion below.

Definition 2. Consumers in any town i in any period t choose a basket of K goods to maximize the function $u_{i,t} : \mathbb{R}^K \to \mathbb{R}$ such that

$$u_i(x_1^i, x_2^i, ..., x_K^i) = \begin{cases} \prod_{j=1}^{K} x_j^i - v & \text{if employed in large firm} \\ \prod_{j=1}^{K} x_j^i & \text{otherwise} \end{cases} \tag{3.1}$$

where $v > 0$ is the utility cost of working in the industrial sector. This cost can capture many different ideas, but in the leading example, I use it to represent the costs to learning how to operate new machinery. We will see later that this cost also causes multiple equilibria to arise in this economy (as in Murphy et al. (1989)).

3.3 Solution Properties

Because there are a large number of small firms in any town, the wage in the cottage sector must equal the marginal cost to production in that sector, or 1. The wage paid in the industrial sector, on the other hand, is set by the large firm just high enough to draw a worker out of cottage labor. Since factory work involves the utility cost v, the wage paid to industrial workers must equal $1 + v$. Notice that beyond this fixed utility cost of industrial employment, there is no disutility from working in this environment, so no matter what technology the towns use, the entire labor force of any town will be employed somewhere in the neighborhood.

The firm owners in town i also behave as consumers. But they introduce an unresolved complication to the model: since firm owners have to pay the fixed cost of setting up a factory before they observe demand, they may incur negative profits in any period. Negative profit shows up in the model as reduced demand by the consumers of town i, since (as I mention below) the income used for purchasing goods is equal to the sum of firm profits and labor wages. I interpret this reduced demand as the worker ownership of firms, so that firm losses are covered by the wages of the workers.

With the above wages and preferences, we can calculate the demand and profit functions for consumers in town i for each good $j \in 1, 2, ..., K$. Given wealth w, and noting that the price for the goods produced in any town is always 1, the demand of consumers in town i for each good j is given by

$$x^i_j = \frac{w}{Kp_j} = \frac{w}{K} \tag{3.2}$$

The cottage producers always earn zero profit. On the other hand, if the large firm produces quantity $y_{i,t} > 0$, it must first pay the fixed set up labor cost F, which costs $F(1 + v)$ at the prevailing wage rate. (Note that, given the timing of the model developed above, the large firm must pay this fixed cost at the beginning of the period, before it can observe demand.) Producing $y_{i,t}$ only requires $l_{i,t} = \frac{y_i}{\alpha}$ beyond the set up cost. Thus, the firms of town i of type $j \in 1, 2$ choose a production level $y_{j,i,t}$ in each period t to maximize the function $\pi_{i,t} : \mathbb{R} \to \mathbb{R}$

Proposition 1. The profits earned by firms in town i in period t is $\pi_{i,t} = \pi_{1,i,t} + \pi_{2,i,t}$, where $\pi_{1,i,t} : \mathbb{R} \to \mathbb{R}$ and $\pi_{2,i,t} : \mathbb{R} \to \mathbb{R}$ are functions given by

$$\pi_{1,i,t} = by_{1,i,t} - F(1+v) \text{if } y_{1,i,t} > 0 \quad (3.3)$$

$$\pi_{2,i,t} = 0 \quad (3.4)$$

where $b = 1 - \frac{1+v}{\alpha}$, and $y_{1,i,t}$ is the output of the large firm.

Notice that positive profits are impossible if $\alpha < 1 + v$; I therefore only consider the case where $\alpha > 1 + v$.

Cottage workers earn less than industrial workers, and cottage producers profit less than industrial producers. Since preferences are identical across towns, we know demand increases linearly in wealth, by a factor of $\frac{1}{K}$. Thus, since the consumers of a town in which cottage producers are active have less wealth than the consumers of a town where industrial producers are active, the demand from the consumers of the cottage town for a neighbor's goods will be smaller than the demand from an industrial town. Thus, industrialization of any town confers a positive spillover on its neighbors; the citizens of the newly industrialized town raise the demand for their neighbors' goods and thereby make it more profitable for industrial producers in other towns to enter the market.

We are now in a position to solve for the demand for the good in town i at time t. Since the goods market should clear in every period, in every town, it must be that $y_{i,t}$, the income earned by the consumers of town i at time t, must equal the demand from workers and firm owners in the neighborhood, that is, $y_{i,t} = \sum_{j \in n_i} \pi_{j,t} + w_{j,t} l_{j,t}$, where $w_{j,t}$ is wages earned. (This basically implies that we assume firms produce to meet demand.) We can break down these components based on whether each town j is industrialized at time t or not. Before we proceed, I want to introduce an indicator that equals one if the large firm is active in town i at time t, i.e. i is "industrialized" at time t.

$$d_{i,t} = \begin{cases} 1 & \text{if } y_{1,i,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

Given the quantity $y_{j,t}$ for all $j \in n_i$, the wealth of the firm owners of town i's neighborhood is 0 if the firm is a cottage producer, and $by_{j,t} - F(1+v)$ if the firm is an industrial producer. The wealth of laborers is $y_{j,t}$ if employed in cottage production, and $(1+v)y_{j,t}$ if employed in industrial production. Finally, only the fraction $\frac{1}{K}$ of each individual's wealth is used to purchase goods of type i. Noting this, we expand the original goods market clearing condition to

$$Ky_{i,t} = \sum_{j \in n_i} d_j(by_{j,t} - F(1+v)) + \sum_{j \in n_i} (1-d_j)y_{j,t} + (1+v)(K - \sum_{j \in n_i} (1-d_j)y_{j,t}) \tag{3.6}$$

where the last term follows from the fact that the total labor used in the neighborhood cannot exceed $\sum_{j \in n_i} L_{j,t} = K$ by assumption. This implies that labor employed in industrial production must equal the total labor available minus that used in cottage production.

The above simplifies to

$$y_{i,t} = \frac{\sum_{j \in n_i, j \neq i} ((b+v)d_j - v)y_j + (1+v)(K - F \sum_{j \in n_i} d_j)}{K - (b+v)d_i + v} \tag{3.7}$$

We can note two things immediately from the above equation. First, the demand for the goods of any town implicitly depends on the demand of every other town on the circle, since demand for town i's good depends on demand for town $i+1$'s good, which in turn depends on the demand for town $i+2$'s good, and so on. Second, demand in every town depends on the value of K, the size of one's neighborhood.

We can obtain a system of N equations using the above relationship for demand in town i, but it is convenient to introduce some matrix notation to summarize the above relationships. Call the vector of indicators for industrial production $\mathbf{d}_t$, so that if $y_{1,i,t} > 0$, the i-th element of $\mathbf{d}_t$ is 1. Set $\mathbf{D}_t = diag(\mathbf{d}_t)$.

We can specify the neighborhood structure of the world by an N by N matrix $\mathbf{S}$. Each entry $\mathbf{S}_{i,j}$ corresponds to the relationship between town i and town j; if j is in i's neighborhood n_i, $\mathbf{S}_{i,j} = 1$, otherwise, $\mathbf{S}_{i,j} = 0$. Note that this matrix is independent of t, since spatial relationships do not change over time in this model.

Using the above notation, we can summarize the N equation system by rearranging the terms of the previous equation. Introduce two more matrices, $\mathbf{G}_t$ and $\mathbf{h}_t$.

$$\mathbf{G}_t = K\mathbf{I}_N - (b+v)\mathbf{S}\mathbf{D}_t + v\mathbf{S} \tag{3.8}$$

$$\mathbf{h}_t = K(1+v)\mathbf{1} - F(1+v)\mathbf{S}\mathbf{d}_t \tag{3.9}$$

Provided that the matrix $\mathbf{G}_t$ has an inverse, the demand vector for all towns in period t, $\mathbf{y}_t$, is given by

$$\mathbf{y}_t = (\mathbf{G}_t)^{-1}\mathbf{h}_t \tag{3.10}$$

This provides a compact form for writing the demand equations, and more importantly allows us to study demand completely in terms of the independent variables, $d_{i,t}$.

The system of demand equations has a solution when $\mathbf{G}_t$ has an inverse, and $\mathbf{G}_t$ has an inverse when it has rank N. $\mathbf{G}_t$ has 3 components: an N by N identity matrix, the N by N matrix $-(b+v)\mathbf{S}\mathbf{D}_t$, and the N by N matrix $v\mathbf{S}$. Establishing the invertibility of $\mathbf{G}_t$

is just a matter of showing that any acceptable choice of parameters b and v cannot yield a matrix with rank less than N.

Proposition 2. For any $K \in \mathbb{N}$, $\mathbf{D}_t$ defined as above, and $\mathbf{S}$, if $\alpha > 0$, $v < \alpha - 1$, and $b = 1 - \frac{1+v}{\alpha}$, $\mathbf{G}_t$ has an inverse.

Proof. $K\mathbf{I}_N$ has rank N. The entries of $\mathbf{SD}_t$ are just the entries $\mathbf{S}_{i,j}d_{j,t}$, so every element of $\mathbf{S}$ is greater than or equal to $\mathbf{SD}_t$, and since $b \neq v$, the elements of $v\mathbf{S} - (b+v)\mathbf{SD}_t$ are either equal to $-b < 0$ or $v > 0$. The i-th column of $\mathbf{G}_t$ then has either K, $K - b$, or $K + v$ in the i-th entry, and either 0, $-b$, or v in the j-th entry, where $j \neq i$. Since $b < K$ for all $K \geq 1$, this column can be expressed as $\mathbf{e}_i + x_i$, where x_i has $-b$, v, and 0 as entries. But then the N element vectors form a basis for the matrix $\mathbf{G}_t$, and $\mathbf{G}_t$ has rank N. □

3.3.1 Existence of Multiple Equilibria

We can view the above environment as an N-person simultaneous move coordination game, played by the potential industrial producers of each town i in each period t. Based on information they have about demand in other towns and the incentives of other potential producers, they decide whether to pay the fixed set up cost and enter the market. Payoffs are weakly higher for all N players when they coordinate. As in the 2 by 2 coordination game, there are (at least) two Nash equilibria under appropriate conditions on the fixed set up cost F, in each of which all players play the same action.

Two equilibria arise for an arbitrarily large circle under a simple condition on the fixed cost, F. In the first, all towns use cottage production, while in the second, all towns are industrialized. I label these equilibria L and H, respectively. For the H equilibrium to exist, it should be the case that if all towns have industrialized, there should be no incentive for an industrial firm owner to deviate to cottage production in any town i. This amounts to the requiring that profits from industrial production exceed the profits from cottage production. If all towns have industrialized, all labor is employed in industrial production, and income in every town is identical. Moreover, since the entire population of each town is employed in industrial production, goods market clearing implies that income from demand is equal to the amount produced if all workers in each town were used as industrial laborers:

$$y_i = \alpha(L_i - F) = \alpha(1 - F) \tag{3.11}$$

A deviation to industrial production cannot be advantageous if profits from industrial production (given industrial production by all others) strictly exceed 0. Thus, a sufficient (but not necessary) condition for the H equilibrium to exist is that industrial profits strictly exceed cottage profits:

$$\begin{aligned} by_i - F(1+v) &= b\alpha(1-F) - F(1+v) > 0 \\ &\Rightarrow F < b \end{aligned}$$

If F is too low, the L equilibrium will not exist, since entry costs will not present a barrier even to towns with small demand for its good. That is, if all towns are engaged in cottage production, it should not be the case that a potential industrial producer in any town has an incentive to set up a factory and produce. Since profits from cottage labor are 0, this means that the profits from industrial production, given demand from a world of cottage neighbors, cannot exceed 0. If all towns are engaged in cottage production, the entire labor force of each town is employed in cottage labor, and goods market clearing implies that

$$y_i = L_i = 1$$

For the L equilibrium to exist, it is sufficient to show that given demand when all towns are cottage producers, any potential industrial producer would make negative profits[2] :

$$\begin{aligned} by_i - F(1+v) &= b - F(1+v) < 0 \\ &\Rightarrow F > \frac{b}{1+v} \end{aligned}$$

If we restrict F to be in between these two bounds, i.e. $F \in (\frac{b}{1+v}, b)$, not only do both the L and H equilibria exist, they are the only two equilibria that exist. To see this, consider an arbitrary set of strategies that don't constitute the L or H equilibrium, assume it is an equilibrium, and show a contradiction. If $\mathbf{d}_t$ is a set of strategies that don't constitute the L or H equilibrium, then $\mathbf{d}_t \neq \mathbf{1}$ and $\mathbf{d}_t \neq \mathbf{0}$. This implies that at least one industrialized town, i has an unindustrialized neighbor, say $i+1$. This implies that demand for town i's good is less than the $\alpha(1-F)$ in the H equilibrium and greater than the 1 demand in the L equilibrium; denote this demand level $\bar{y}$. We have two cases: $F > \frac{b\bar{y}}{1+v}$ or $F \leq \frac{b\bar{y}}{1+v}$. In the first case, the industrialized producer i will make negative profits this period, and would have done strictly better by not entering at all. In the second case, since the cottage producer $i+1$ shares the same neighbors as i, and the demand level $\bar{y} = \frac{1}{3} + \frac{\bar{y}}{3} + \frac{y_{i-1}}{3}$, then

[2]Since industrializing raises worker wages in one's own town, the potential industrial producer should also consider this increase in demand when calculating the value of deviating from the L equilibrium. However, per unit output, an industrial producer will pay an extra v in wages over the cottage wage labor bill. But the producer only gets back the fraction $\frac{1}{K}$ of this amount, since workers spend the extra money on the other goods in the neighborhood. Thus, the producer's potential profits are strictly lower than if we were to ignore the effect as above, and if F satisfies the above condition, it will ensure the existence of the L equilibrium.

the cottage producer would obtain $y_{i+1} > \frac{y_{i+1}}{3} + \frac{\bar{y}}{3} + \frac{y_{i+2}}{3} \geq \frac{y_{i+1}}{3} + \frac{\bar{y}}{3} + \frac{1}{3}$. Thus, the cottage producer would have done better by switching to industrial production and neither case can be an equilibrium.

The next condition summarizes the discussion above.

Proposition 3. In any period t, if the fixed set up cost for industrial production F satisfies

$$F \in (\frac{b}{1+v},\ b) \tag{3.12}$$

then only two equilibria exist in this period: one in which $y_{1,i,t} = 0$ for all towns i, another in which $y_{1,i,t} = \alpha(L_i - F)$, denoted L and H, respectively.

One can show that the H equilibrium Pareto dominates the L equilibrium. To see this, note that in the L equilibrium, agents fulfill their demand $\frac{1}{K}$ for every good in their neighborhood, while in the H equilibrium agents fulfill their demand $\frac{\alpha(1-F)}{K}$ for every good in their neighborhood. Under the above condition, $F < b$, so $\alpha(1-F) > \alpha(1-b) = \alpha(1 - 1 + \frac{1+v}{\alpha}) = 1 + v$. Since $v > 0$ by assumption, $\alpha(1-F) > 1$. Thus, every agent in every town enjoys higher utility in the H equilibrium than the L equilibrium, and the H equilibrium Pareto dominates the L equilibrium.

3.3.2 Neighborhood Size and Profits

The above environment will become part of a repeated game in which each stage is the N-person coordination game played by the industrial producers. I am interested in the dependence of the dynamics of the model on the neighborhood size, K. One result that follows from the structure of the stage game itself is that as neighborhood size increases, industrial producers can have more unindustrialized neighbors and still achieve positive profits. To see this, start with the condition that industrial owners have positive profits, and rewrite it using the matrix notation developed above:

$$
\begin{aligned}
\pi_t &> \mathbf{0} \Leftrightarrow \\
b\mathbf{y}_t - F(1+v)\mathbf{1} &> \mathbf{0} \Leftrightarrow \\
\mathbf{y}_t &> \frac{F(1+v)}{b}\mathbf{1} \Leftrightarrow \\
(\mathbf{G}_t)^{-1}\mathbf{h}_t &> \frac{F(1+v)}{b}\mathbf{1} \Leftrightarrow \\
\mathbf{h}_t &> \mathbf{G}_t\frac{F(1+v)}{b}\mathbf{1} \Leftrightarrow \\
K(1+v)\mathbf{1} - F(1+v)\mathbf{S}\mathbf{d}_t &> (K\mathbf{I}_N - (b+v)\mathbf{S}\mathbf{D}_t + v\mathbf{S})\frac{F(1+v)}{b}\mathbf{1} \Leftrightarrow \\
\mathbf{S}\mathbf{d}_t &> K(\frac{F(1+v)-b}{vF})\mathbf{1} = cK\mathbf{1}
\end{aligned}
$$

The i-th entry of the N element vector on the left hand side is just the sum $\sum_{j \in n_i} d_{j,t}$, that is, the number of industrialized towns in the neighborhood of town i. The above condition thus tells us that if the number of an industrial producer's industrialized neighbors this period is more than some constant $c = \frac{F(1+v)-b}{vF}$ multiplied by the size of the producer's neighborhood, demand for her good is high enough this period to ensure positive profits. Moreover, by proposition 3, $c \in (0,1)$. Because this number is less than 1, only a fraction of town i's neighbors need to be industrialized for an industrial producer to enjoy positive profits in town i. For K small, this fraction is greater than the largest integer smaller than K, which implies that a single cottage producer in town i's neighborhood will cause any industrial producers to incur negative profits. As K increases, however, the number of cottage producers that the neighborhood can "tolerate" before causing negative profits for industrialists increases with it. More formally,

Proposition 4. If K is the size of the neighborhood of town i, $c = \frac{F(1+v)-b}{vF}$, and $d_{i,t}$ indicates whether town i is industrialized in period t, then if $d_{i,t} = 1$ and

$$\sum_{j \in n_i} d_{j,t} > cK \tag{3.13}$$

then $\pi_i > 0$.

Moreover, if $n_C = \sum_{j \in n_i} (1 - d_{j,t})$ is the proportion of cottage producing towns in n_i in period t, then as K increases, the value of n_C that satisfies $K - n_C > cK$ increases.

To see the second statement, assume $K - n_C > cK$ for some K. Since $K > 0$, rewrite this equation as $1 - \frac{n_C}{K} > c$. But then if K increases by 1, n_C can increase by 1 and still satisfy the inequality.

This fact is a simple consequence of algebra, but it has important implications when we consider the model dynamics below. Basically, when neighborhoods have more overlap, the world can begin with more cottage producers and still end up in the H equilibrium.

3.4 Imitation and Dynamics

In order to study equilibrium selection, I impose a simple, per period decision rule on the model described above. The dynamics induced by this decision rule are easy to compute, but not as easy to describe more generally, since the actions of each agent depend on the actions of all other agents. In this section, I first specify my restrictions on the larger equilibrium selection problem and how the larger problem relates to this one. Second, I derive a necessary condition of the imitation rule that only uses the number of neighbors. Third, I show that as neighborhood size increases, the process induced by the imitation rule necessarily leads to the steady state in which all neighbors are industrialized.

3.4.1 Imitation as Selection Process

Imagine again the environment above as a game, but this time consider repeating the game so that the two equilibria described in the last section exist in every stage of this game. At the beginning of each period, potential industrial firm owners in town i have to choose whether to set up a factory for industrial production or do nothing. If these agents set up a factory, they recieve profits from its success or failure. As noted above, let the variable $d_{i,t}$ equal 1 if the firm owner chooses the industrial technique and 0 otherwise. A strategy for each agent i should specify the value of the variable $d_{i,t}$ for every period, for every possible history of decisions $h_t = (\mathbf{d}_{t-1}, \mathbf{d}_{t-2}, ...)$.

The imitation rule says that in the next period, you should choose the technology that the most profitable firm owners in your neighborhood chose this period. To be more specific, assume for a moment that $d_{j,t} = 1$ for at least 1 town $j \in n_i$ and $d_{k,t}0$ for at least 1 town $k \in n_i$. Then $d_{i,t+1} = 1$ if

$$\frac{\sum_{j \in n_i} d_{j,t}\pi_{j,t}}{\sum_{j \in n_i} d_{j,t}} > \frac{\sum_{j \in n_i}(1 - d_{j,t})\pi_{j,t}}{\sum_{j \in n_i}(1 - d_{j,t})} \tag{3.14}$$

Since the profits to cottage production are always 0, the right hand side of the equation is always 0. But then since $\sum_{j \in n_i} d_{j,t}$ is always positive, the above condition is satisfied whenever the numerator $\sum_{j \in n_i} d_{j,t}\pi_{j,t} > 0$. Now since this rule is meant to model imitation, it should also be the case that the absence of one type in i's neighborhood this period should imply that i cannot take on that type next period. That is, if $\sum_{j \in n_i} d_{j,t} = 0$, $d_{i,t+1} = 0$, and if $\sum_{j \in n_i} d_{j,t} = K$, $d_{i,t+1} = 1$. Summarizing all of these features, I state a formal version of the decision rule below.

Definition 3. Given the vector of production choices $\mathbf{d}_t$ for every town in period t, the imitation rule is a production choice function $g(\mathbf{d}_t) : \mathbb{R}^N \to \mathbb{R}^N$ for every town in period $t+1$ where each component $d_{i,t+1} \in \mathbb{R}$ is given by

$$d_{i,t+1} = \mathbf{1}(\sum_{j \in n_i} d_{j,t}\pi_{j,t} > 0) + \mathbf{1}(\sum_{j \in n_i} d_{j,t} = K) \tag{3.15}$$

where n_i is the neighborhood of town i, and K is the size of every neighborhood.

Without the above rule, repetition of the stage game could lead to cooperative outcomes. With the above rule, I force agents to act in a way that does not allow this behavior; though this comes at the cost of generality in the statements I can make about agents' behavior, I can study the effects of neighborhood size on the dynamics induced by the above rule without any additional assumptions. When we consider the term "high income equilibrium" in the context of dynamic games, we normally mean a state in which every agent in the economy is playing a strategy which is a best response to every other agent's action, given that all other agents are playing best responses to every other agent's action as well. Moreover, a strategy specifies an action in every period in which the game is played. In two senses, then, our conventional definition of equilibrium is trivially satisfied under the imitation rule: first, agents are not best responding, they are responding the only way they know how, namely, by imitating. Second, because agents are all playing an identical strategy every period, the concept of strategic stability that is built into the definition of equilibrium in a dynamic game no longer has meaning. Strategic stability is ensured by assumption, in the sense that everyone plays the same strategy.

For these reasons, I do not consider the effect of distance on the selection of equilibria in this section, but rather the effect of distance on the selection of the steady states of the imitation rule process. These are very different concepts, and it is not clear that the latter concept tells us anything about the former, which is essentially the question posed at the beginning of the paper. I argue, however, following Cooper (1994), that the steady states of the imitation rule process are informative about equilibrium selection under the assumption that agents are one-period lived, and that these equilibria represent an important special case.

When agents only live for one period, cooperative outcomes cannot arise from repeated interactions. Despite the significant restriction, this special case is still important. First, as in Cooper (1994), this restriction on behavior prevents the repeated interactions outcome from becoming a confounding factor. Second, it is not clear that agents in this type of environment should ever engage in repeated interactions; the stage game is equivalent to a coordination game played by the potential monopolists each period. If we are trying to describe the plight of a potential-new-technology-adopting-garment-manufacturer, for example, the decision maker today will probably not be the same person as the decision maker tomorrow, especially if, upon entering today, the manufacturer faces inadequate demand and incurs negative profits.

Under the "one period life" assumption the dynamics become much simpler. There are exactly two equilibria in the stage game, and if there are T total periods in the repeated game, there are 2^T equilibria.[3] If we only consider steady state equilibria, there are exactly 2 of them, one where potential monopolists don't enter in every period, another where potential monopolists enter in every period and earn positive profits.

The steady state equilibria of this system constitute a version of the multiple equilibria that motivated the study. With this in mind, I reformulate the original problem as

[3]To see this, note that the agents in period t could play either of the two equilibria in every period. Since the decisions of agents in any other period are unaffected by the equilibrium played in period t, no matter what the choice of period t agents, either of the two equilibria can be played in every other period.

follows: given that one of two steady state equilibria are possible under the assumption that agents only live for one period, is the high income equilibrium more likely to emerge as neighborhood size increases? The imitation rule, and the process induced by this rule, are then meant to help us answer this question. If imitation is an important component of how individuals adopt new technology, then if the imitation process converges to a steady state that matches one of the equilibria of the system, it seems reasonable to think that the steady state of the process would arise more often as an equilibrium in a more general setting.

3.4.2 Nearest Neighbors Criterion

The matrix notation developed in section 2.3 implies a simpler form of the imitation rule which only relies on the number of industrialized neighbors in every neighborhood. Using this form, I arrive at a simple result that follows from the profits condition of section 2.4. Recall that the imitation rule is a condition on a sum of profits earned by different industrial producers: the potential industrial producer in town i should decide to enter production next period if the sum of profits earned by all industrial producers in the neighborhood this period is strictly positive. I will show that the condition on the sum of profits has the same form as the condition for profits in each town. First, consider the sum of the incomes of towns i and $i + 1$, using the expanded forms given by the original equation defining income in section 2.2:

$$(K-(b+v)d_{i,t}+v)y_{i,t}+(K-(b+v)d_{i+1,t}+v)y_{i+1,t}=$$
$$\sum_{j\in n_i,j\neq i}((b+v)d_{j,t}-v)y_{j,t}+(1+v)(K-F\sum_{j\in n_i}d_{j,t})+$$
$$\sum_{j\in n_{i+1},j\neq i}((b+v)d_{j,t}-v)y_{j,t}+(1+v)(K-F\sum_{j\in n_{i+1}}d_{j,t})$$

Call $\tilde{x}_{i,t}=x_{i,t}+x_{i+1,t}$, where $\tilde{x}_{N,t}=x_{N,t}+x_{1,t}$. Then:

$$(K-(b+v)\tilde{d}_{i,t}+v)\tilde{y}_{i,t}+(b+v)(d_{i+1,t}y_{i,t}+d_{i,t}y_{i+1,t})=$$
$$\sum_{j\in n_i,j\neq i}((b+v)\tilde{d}_{j,t}-v)\tilde{y}_{j,t}+$$
$$(1+v)(2K-F\sum_{j\in n_i}\tilde{d}_{j,t})-\sum_{j\in n_i,j\neq i}(b+v)d_{j+1,t}y_{j,t} \quad (3.16)$$

We can express the above condition in terms of the matrix notation introduced in section 2.2. Define a new N by 1 vector, $\tilde{\mathbf{h}}_t$, where

$$\tilde{\mathbf{h}}_t=2K(1+v)\mathbf{1}-F(1+v)\mathbf{S}\tilde{\mathbf{d}}_t \quad (3.17)$$

Then we can reformulate the equation for the sum $y_{i,t}+y_{i+1,t}$ as

$$\mathbf{G}_t\tilde{\mathbf{y}}_t=\tilde{\mathbf{h}}_t+(b+v)(D_{t,1}+D_{t,-1})\mathbf{y}_t \quad (3.18)$$

where the matrices $D_{t,1}$ and $D_{t,-1}$ are shifted versions of the matrix $D_t=diag(\mathbf{d}_t)$, so that the i-th diagonal element of D_t^1 is $d_{i+1,t}$, and the i-th diagonal element of D_t^{-1} is $d_{i-1,t}$. The matrix $\mathbf{G}_t$ is identical to the matrix in section 2.2.

If $\pi_{i,t} + \pi_{i+1,t} = \tilde{\pi}_{i,t} > 0$, then $y_{i,t} + y_{i+1,t} = \tilde{y}_{i,t} > \frac{2F(1+v)}{b}$. When we place this relation into the equation above, we obtain

$$\mathbf{G}_t(\frac{2F(1+v)}{b}) < \tilde{\mathbf{h}_t} + (b+v)(D_{t,1} + D_{t,-1})\mathbf{y}_t$$

Since $\mathbf{y}_t > \mathbf{0}$, $d_{i,t} \geq 0$ for all towns i, periods t, and $b, v > 0$ by assumption, we know the second term on the right hand side of the equation is positive. But then we know that the condition holds if

$$\tilde{\mathbf{h}_t} > \mathbf{G}_t(\frac{2F(1+v)}{b})$$

Finally, expand the matrices and rearrange as in section 2.4 to obtain the final necessary condition:

$$\mathbf{S}\tilde{\mathbf{d}_t} > 2K(\frac{F(1+v) - b}{vF})\mathbf{1} = 2cK\mathbf{1} \tag{3.19}$$

The above condition states that, given any two neighboring, industrialized towns i and $i+1$, we can just add the number of industrialized neighbors in the neighborhood of each town to ensure that the sum of the profits earned by producers in each town are strictly positive. Also, we can easily generalize the above condition to larger sets of industrialized neighbors. Given another town $i+2$ and defining $\tilde{\tilde{x}}_{i,t} = \tilde{x}_{i,t} + x_{i+2,t}$, the only component of equation 3.16 that would change is the coefficient on K. More formally,

Proposition 5. Say town i is industrialized, and it has industrialized neighbors $k_j \in n_i$, with neighborhoods n_{k_j}. Then if

$$\sum_{j=1}^{m} \sum_{l \in n_{k_j}} \mathbf{S}_{k_j,l} d_{l,t} > mcK\mathbf{1} \tag{3.20}$$

where $c = \frac{F(1+v)-b}{vF}$, then $\sum_{j=1}^{m} \pi_{k_j} > 0$.

3.4.3 Effects of Increasing Neighborhood Size

I show that the above condition is easier to satisfy for larger neighborhood sizes. This fact follows almost immediately from the above condition. First, from proposition 4, we know that the maximum number of cottage producers that any neighborhoood can "tolerate" before causing industrial producers to incur negative profits, n_C, increases as neighborhood size increases. Second, we know from the above condition that by taking the sums of industrialized neighbors for different neighborhoods, we obtain a minimum condition for the sum of industrialized neighbor profits to be positive. Now consider an arbitrary initial state, $\mathbf{d}_0$. The process can converge to one of two steady states[4], one in which all towns are cottage producers, the other where all towns are industrial producers. The only way the process will converge to the industrial steady state is if cottage towns imitate other profitable industrial towns in their neighborhood. The only way the other towns are profitable is if they meet proposition 5. But as I showed in the proof of proposition 4, this minimum condition is more easily met when neighborhood size increases for any neighborhood, and proposition 5 is just a sum of conditions in proposition 4, the final result follows.

Proposition 6. The probability p_0 that an initial state $\mathbf{d}_0$ will lead the imitation rule process to converge to the H steady state increases as K increases.

Proof. By proposition 4, the maximum number of cottage towns in any neighborhood to ensure positive profits, n_C, increases as K increases. By proposition 5, this in turn implies

[4] I am unable to show this formally, but the process can converge to other steady states. I argue, however, that these states are not of primary interest because they are "unstable", that is, randomly changing the production technique for a small number of towns causes the process to wander away from this state.

that, for any cottage producer in the initial state, the cottage producer will see industrial neighbors with positive average profits. For an arbitrary initial state, then, the likelihood that a cottage producer will imitate an industrial producer next period will increase as neighborhood size increases, and the same will be true for all subsequent periods. Thus, a higher proportion of arbitrary initial states will converge to the H steady state, and if we measure this proportion by p_0, p_0 increases with K. □

Again, by my argument in the beginning of this section, the fact that the process is more likely to converge to the H steady state as neighborhood size increases implies that, if we take it for granted that imitation is an important part of equilibrium selection in this environment, the true dynamics (whatever they may be) are more likely to converge to the H steady state equilibrium as neighborhood size increases.

3.5 Conclusions

I see value in the above work as laying the foundation for several avenues of future research. In this section, I outline my plans for this work and thoughts on how to test some of the components of the model and its likely implications.

3.5.1 Future Work

This project presents a number of interesting avenues for future research. First, fully characterizing the absorbing sets and basins of attraction for the imitation rule process, as in Eshel et al. (1998), would help deepen intuition about the model. This could be done computationally by "working backwards": given a final state of the process, all the previous period states that could possibly lead to the final state can be calculated, since we've specified an explicit form for the decision rule that would lead it there. Using the matrix notation developed above, we could specify a difference equation for the process, classify the system, and investigate whether other theory might be applicable. It seems that because the structure of the model is fairly simple, this system should resemble an already well-studied one in the literature.

Second, an important omission from this draft of the paper is a model of the dynamics when returns to industrial production are uncertain. Specifically, it would be interesting to include a symmetric productivity shock ϵ in the production function of the industrial producer, so that productivity becomes $\alpha + \epsilon$. There is one interesting effect I want to look for here: how the variance of ϵ affects the likelihood of the low income equilibrium. Presumably, agents with good shocks could support the industrialization of cottage producers that wouldn't occur otherwise, but then again the opposite is also true for bad shocks.

Third, I want to understand the effect of random experimentation on the model. If agents were to spontaneously "experiment" with the new technology, switching even when decision rules may say otherwise, it would be interesting to know how much of this experimentation is necessary to push society to the high income equilibrium when the low income equilibrium arises without it. From a policy perspective, we could view this kind of activity as the result of innovation policy or technology transfer programs for some industries, and its relationship to the multiple equilibria as a policy alternative for escaping an underdevelopment trap.

Fourth, it is important to understand under what conditions the imitation rule will be played by agents with a wider strategy space. As mentioned above, the imitation rule represents a single strategy played by all agents in every period in the game. If agents believed that other player's actions were informative about conditions they themselves could not observe, then imitation may be a best response.

Last, I want to explore the model in the context of other network structures. As we get more links between neighbors, neighborhoods are getting larger, but there are also more connections between neighborhoods. In the case of uncertain returns, this seems to imply that positive or negative shocks could be transmitted more widely than in the circle case.

I am confident that with the framework already in place, exploring these avenues should be relatively straightforward, representing simple applications of the above model to a wider variety of environments. There is an important literature on network games that was largely left out of this paper, and it seems to have enormous potential for helping to answer many of the questions I've set out.

3.5.2 Empirical Exercises

The original Murphy et al. (1989) model was just a convenient theoretical description for an effect proposed by another economist. Though there is evidence to support the model as it stands, it has not been tested in a systematic way. The same is true for the diffusion dynamics proposed in this paper. Therefore, an important exercise would be some sort of test of the original assumptions that underlie the model. Specifically, there are two questions that I'd like to answer. First, does the likelihood of new technology adoption change as the average distance from other technology adopters increases? This could be done using the data described in Doss (2003), a micro-level data set collected on agricultural technology adoption in Eastern Africa. Second, do the dynamics of technology adoption follow an imitation or best response type of rule? We may be able to distinguish between these two classes of rules by noting that if agents are best responding, they will not be making systematic errors in judgment that they might if they followed an imitation rule. It is difficult to make this more precise without developing the dynamics under best response rules.

Coordination failures are difficult to identify in practice. Many of the exogenous variables that might be used to determine outcomes in conventional growth studies would actually be endogenous in a multiple equilibria model, since multiple equilibrium relationships exist between variables. Bloom et al. (2003) test for the existence of multiple equilibria in national income using latitude and rainfall data. Thus, though it is likely that transportation and communication network indicators are likely to be endogenous in this type of model, it may be possible to exploit geographical variation in "neighborhood size" (e.g. how hilly a region is, presuming that impedes travel) to test whether one equilibrium is systematically selected over another.

Though the model above makes some sense at the micro level, it is difficult to understand how these effects will aggregate, especially when the parameters of the model are probably different across industries. After all the issues raised in this section are addressed, the next step would be to connect this model to Durlauf (1993). As I mentioned in the introduction, that paper seems to contradict the neighborhood size result found here, and an important and interesting task for the farther future may be to resolve this tension.

Appendices

Appendix A

Bank Size, Leverage, and Financial Downturns

A.1 Equilibrium Conditions

In this section, I describe the full system of equations that describes equilibrium. Before proceeding, make a few simplifications to the model. First, 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:

$$\begin{aligned}\int_{\kappa\in B} b(\kappa)p(\kappa)d\kappa = \int_{\kappa\in B} \bar{b}(\kappa)p(\kappa)d\kappa = \int_{\kappa\in B} \left(\phi_t(\kappa)n_t(\kappa) - d_t(\kappa)\right) d\kappa \\ = -\int_{\kappa\in L} b(\kappa)p(\kappa)d\kappa = \int_{\kappa\in L} \left(n_t(\kappa) + d_t - Q_t(\kappa)k_t(\kappa)\right) d\kappa \quad \text{(A.1)}\end{aligned}$$

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

$$D_t = \int_{\kappa^*}^{\overline{\kappa}} p(\kappa)\phi_t(\kappa)((Z_t + (1-\delta)Q_t^i)K_t(\sigma+\xi) - \sigma R_{t-1}D_{t-1})d\kappa$$
$$- \int_{\underline{\kappa}}^{\kappa^*} p(\kappa)((Z_t + (1-\delta)Q_t(a))K_t(\sigma+\xi) - \sigma R_{t-1}D_{t-1})d\kappa$$
$$+ \int_{\underline{\kappa}}^{\kappa^*} p(\kappa)\frac{Q_t(a)}{\kappa(a)}(1-\delta)K_t d\kappa \quad \text{(A.2)}$$

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:

$$s_t(\kappa) = i_t(\kappa) + k_t(\kappa) = \frac{\kappa}{Q_t(\kappa)}(n_t(\kappa) + d_t(\kappa) + b_t(\kappa))$$
$$= \frac{\kappa}{Q_t^i}(1+\phi_t(\kappa))n_t(\kappa) \;\; \forall \kappa \in [\kappa^*, \overline{\kappa}] \quad \text{(A.3)}$$

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

$$Q_t^i \left(I_t + (1-\delta)K_t \int_{\kappa^*}^{\overline{\kappa}} p(\kappa) \right) = \int_{\kappa^*}^{\overline{\kappa}} \kappa(1+\phi_t(\kappa))n_t(\kappa)d\kappa \quad \text{(A.4)}$$

A.1.1 Full System

HH/Firms

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha} \tag{A.5}$$

$$K_t = \psi_t(I_{t-1} + (1-\delta)K_{t-1}) \tag{A.6}$$

$$Y_t = C_t + (1 + f(\frac{I_t}{I_{t-1}}))I_t \tag{A.7}$$

$$1 = E_t \Lambda_{t,t+1} R_t \tag{A.8}$$

$$u_{Ct} = (C_t - \gamma C_{t-1})^{-1} - \beta\gamma(C_{t+1} - \gamma C_t)^{-1} \tag{A.9}$$

$$\Lambda_{t,t+1} = \beta \frac{u_{Ct+1}}{u_{Ct}} \tag{A.10}$$

$$Z_t = \alpha A_t (\frac{L_t}{K_t})^{1-\alpha} \tag{A.11}$$

$$Q_t^i = 1 + f(\frac{I_t}{I_{t-1}}) + (\frac{I_t}{I_{t-1}})f'(\frac{I_t}{I_{t-1}}) - E_t \Lambda_{t,t+1}(\frac{I_{t+1}}{I_t})^2 f'(\frac{I_{t+1}}{I_t}) \tag{A.12}$$

$$\chi L_t^{\varphi} = (1-\alpha)\frac{Y_t}{L_t} E_t u_{Ct} \tag{A.13}$$

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

$$A_t = \rho_A A_{t-1} + \epsilon_{At} \tag{A.14}$$

$$\psi_t = \rho_\psi \psi_{t-1} + \epsilon_{\psi t} \tag{A.15}$$

Bank Optimization

$$Q_t(\kappa) = \frac{\kappa}{\kappa_t^c} Q_t^i \;\; \forall \kappa \in L \tag{A.16}$$

$$\nu_t = \nu_{bt} \tag{A.17}$$

$$\nu_{bt} = \frac{\kappa_t^c}{Q_t^i} \nu_{st} \tag{A.18}$$

$$\lambda_t(\kappa) = \frac{(\kappa - \kappa_t^c)\nu_{st}}{\kappa \theta Q_t^i - (\kappa - \kappa_t^c)\nu_{st}} \;\; \forall \kappa \in B \tag{A.19}$$

$$\lambda_t(\kappa) = 0 \;\; \forall \kappa \in L \tag{A.20}$$

$$\Omega_t(\kappa) = 1 - \sigma + \sigma \nu_{bt}(1 + \lambda_t(\kappa)) \tag{A.21}$$

$$\phi_t(\kappa) = \frac{\kappa(\nu_{st} - \theta Q_t^i)}{\nu_{bt} Q_t^i - \kappa(\nu_{st} - \theta Q_t^i)} \tag{A.22}$$

$$\nu_{bt} = E_t R_t \Lambda_{t,t+1}(1 - \sigma + \sigma \nu_{bt+1}) + E_t R_t \Lambda_{t,t+1} \sigma \nu_{bt+1} G_0 \tag{A.23}$$

$$\begin{aligned} \nu_{st} = E_t \Lambda_{t,t+1} \Psi_{t+1} [& (1 - \sigma + \sigma \nu_{bt+1} + \sigma \nu_{bt+1} G_0) Z_{t+1} \\ & + (1 - \sigma + \sigma \nu_{bt+1})(1 - \delta) Q_{t+1}^i \Big(\int_{\kappa_{t+1}^c}^{\overline{\kappa}} p(\kappa) d\kappa \Big) + \sigma \nu_{bt+1} G_0 (1 - \delta) Q_{t+1}^i \\ & + (1 - \sigma + \sigma \nu_{bt+1}) \Big(\int_{\underline{\kappa}}^{\kappa_{t+1}^c} p(\kappa) \kappa d\kappa \Big) (1 - \delta) Q_{t+1}^n] \end{aligned} \tag{A.24}$$

$$G_0 = \int_{\kappa_{t+1}^c}^{\overline{\kappa}} p(\kappa) \lambda_{t+1}(\kappa) \, d\kappa \tag{A.25}$$

Securities Market

$$Q_t^i (I_t + (1 - \delta) K_t \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa) d\kappa) = ((Z_t + (1 - \delta) Q_t^i)(\sigma + \xi) K_t - \sigma R_{t-1} D_{t-1}) G_1 \tag{A.26}$$

$$G_1 = \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa)\kappa(1+\phi_t(\kappa))\, d\kappa \tag{A.27}$$

Deposit Market

$$\begin{aligned} D_t = (Z_t + (1-\delta)Q_t^i)(\sigma+\xi)K_t - \sigma R_{t-1}D_{t-1})G_2 \\ - (Z_t - \sigma R_{t-1}D_{t-1})(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)d\kappa) \\ - (1-\delta)Q_t^n(\sigma+\xi)K_t(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)\kappa d\kappa) + Q_t^n(1-\delta)K_t(\int_{\underline{\kappa}}^{\kappa_t^c} p(\kappa)d\kappa) \end{aligned} \tag{A.28}$$

$$G_2 = \int_{\kappa_t^c}^{\overline{\kappa}} p(\kappa)\phi_t(\kappa)\, d\kappa \tag{A.29}$$

A.1.2 Steady State

HH

$$I = \delta K \tag{A.30}$$

$$C = [A(\frac{L}{K})^{1-\alpha} - \delta]K \tag{A.31}$$

$$\chi L^{\varphi} = (1-\alpha)A(\frac{L}{K})^{-\alpha}\frac{1-\beta\gamma}{1-\gamma}\frac{1}{C} \tag{A.32}$$

$$L = (\frac{Z}{\alpha A})^{\frac{1}{1-\alpha}}K \tag{A.33}$$

$$R = \frac{1}{\beta} \tag{A.34}$$

$$\Lambda = \beta \tag{A.35}$$

$$Q^i = 1 \tag{A.36}$$

Bank Optimization

$$\lambda(\kappa) = \frac{(\kappa - \kappa^c)\nu_s}{\kappa\theta - (\kappa - \kappa^c)\nu_s} \ \ \forall \kappa \in B \tag{A.37}$$

$$\lambda(a) = 0 \ \ \forall a \in L \tag{A.38}$$

$$\phi(\kappa) = \frac{\kappa(\nu_s - \theta)}{\nu_b - \kappa(\nu_s - \theta)} \tag{A.39}$$

$$Q^n = \frac{1}{\kappa^c} \tag{A.40}$$

$$\frac{1}{\kappa^c} = \nu_s(1 - \frac{\sigma}{1-\sigma}\overline{G_0}) \tag{A.41}$$

$$\nu_b = \kappa^c \nu_s \tag{A.42}$$

$$Z = \frac{\nu_s}{\beta\nu_b} - \frac{1-\delta}{\nu_b}\left((1-\sigma+\sigma\nu_b)(\int_{\kappa^c}^{\overline{\kappa}} p(\kappa)d\kappa) + \sigma\nu_b\overline{G_0}\right) - \frac{(1-\delta)Q^n}{\nu_b}\left((1-\sigma+\sigma\nu_b)(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)\kappa d\kappa)\right) \tag{A.43}$$

Securities Market

$$\frac{\sigma}{\beta}\frac{D}{K} = (Z+1-\delta)(\sigma+\xi) - \frac{\delta + (1-\delta)(\int_{\kappa^c}^{\overline{\kappa}} p(\kappa)d\kappa)}{\overline{G_1}} \tag{A.44}$$

Deposit Market

$$\frac{N^i}{K} = (Z+1-\delta)(\sigma+\xi) - \frac{\sigma}{\beta}\frac{D}{K} \tag{A.45}$$

$$\frac{D}{K} = \frac{N^i}{K}\overline{G_2} + Q^n(1-\delta)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)d\kappa\right) - \left(Z(\sigma+\xi) - \frac{\sigma}{\beta}\frac{D}{K}\right)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)d\kappa\right) - (1-\delta)Q^n(\sigma+\xi)\left(\int_{\underline{\kappa}}^{\kappa^c} p(\kappa)\kappa d\kappa\right) \tag{A.46}$$

$$G_0 = \int_{\kappa_{t+1}^c}^{\overline{\kappa}} \frac{p(\kappa)(\kappa - \kappa_{t+1}^c)\nu_{st+1}}{\kappa\theta Q_{t+1}^i - (\kappa - \kappa_{t+1}^c)\nu_{st+1}} d\kappa \tag{A.47}$$

$$G_1 = \int_{\kappa_t^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa\kappa_t^c\nu_{st}}{\kappa_t^c\nu_{st} - \kappa\nu_{st} + \kappa\theta Q_t^i} d\kappa \tag{A.48}$$

$$G_2 = \int_{\kappa_t^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa(\nu_{st} - \theta Q_t^i)}{\kappa_t^c\nu_{st} - \kappa\nu_{st} + \kappa\theta Q_t^i} d\kappa \tag{A.49}$$

$$\overline{G_0} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)(\kappa - \kappa^c)\nu_s}{\kappa\theta - (\kappa - \kappa^c)\nu_s} d\kappa \tag{A.50}$$

$$\overline{G_1} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa\kappa^c\nu_s}{\kappa^c\nu_s - \kappa\nu_s + \kappa\theta} d\kappa \tag{A.51}$$

$$\overline{G_2} = \int_{\kappa^c}^{\overline{\kappa}} \frac{p(\kappa)\kappa(\nu_s - \theta)}{\kappa^c\nu_s - \kappa\nu_s + \kappa\theta} d\kappa \tag{A.52}$$

To solve the model, I first guess a productivity cutoff κ^c. Then the value from interbank lending $\nu_b = \kappa^c\nu_s$, and the equation for interbank lending above becomes an equation in one variable, which we can solve for parameters. We can then use the equation for ν_s to get Z, and use the investing islands securities market equation to get $\frac{D}{K}$. With this in hand, we can test the deposit market clearing condition, by calculating both the right and left hand sides of equation (21).

Appendix B

Aggregate Implications of Capacity Constraints

B.1 Solution

In this section, we list the full set of equilibrium equations under the assumptions that idiosyncratic demand shocks b_{it} are iid, and the shock process has mean 1, so that

$$b_{it} = \epsilon_{bt}$$
$$\epsilon_{bt} \underset{d}{\sim} ln\mathcal{N}(-\frac{\sigma_b^2}{2}, \sigma_b)$$

The first assumption implies that the one-period ahead prices and installed capital choices p_{it} and k_{it} are identical across firms in every period. We'll interchangeably call these common choices p_t and k_t. In turn, this implies that several quantities in the model also become identical across firms, e.g. $\overline{b}_{it} = \overline{b_t}\ \forall i$. We present only the simplified equations below.

Since this is a partial equilibrium model, both wages w_t and rental rates for capital R_t are treated as given.

Final Goods Firms

$$Y_t = \left(\left[\int_0^{\overline{b_t}} (A_t b_{it})^{\frac{1}{\sigma}} y_{it}^{d\,\frac{\sigma-1}{\sigma}} di \right] \right)^{\frac{\sigma}{\sigma-1}} \tag{B.1}$$

$$Y_t = I_t \tag{B.2}$$

$$P_t = (\int_0^{\infty} (A_t b_{it})^{\frac{1-\sigma}{\sigma}} p_{it}^{1-\sigma} di)^{\frac{1}{1-\sigma}} = A_t^{\frac{1}{\sigma}} p_t \tag{B.3}$$

$$I_t = \int_0^{\infty} (A_t b_{it})^{\frac{1}{\sigma}} p_{it} y_{it}^d di = p_t \int_0^{\infty} (A_t b_{it})^{\frac{1}{\sigma}} y_{it}^d di \tag{B.4}$$

$$y_{it}^d = \frac{I_t}{P_t} \left(\frac{P_t}{p_{it}} \right)^{\sigma} A_t b_{it} = \frac{I_t}{P_t} A_t^2 b_{it} \tag{B.5}$$

Intermediate Goods Firms

$$y_{it}^C = y_t^C = C_1 p_t^{\frac{2-\alpha}{\alpha}} w_t^{-\frac{2-2\alpha}{\alpha}} k_t \tag{B.6}$$

$$\bar{b}_{it} = \overline{b_t} = \frac{P_t}{A_t I_t} y_t^C \tag{B.7}$$

$$\tilde{k}_{it} = C_0^{\frac{1-\alpha}{2-\alpha}} k_{t-1}^{\frac{1-\alpha}{2-\alpha}} w_t^{\frac{1-\alpha}{2-\alpha}} y_{it}^{\frac{1}{2-\alpha}} \tag{B.8}$$

$$l_{it} = y_{it}^{\frac{1}{1-\alpha}} \tilde{k}_{it}^{-\frac{\alpha}{1-\alpha}} \tag{B.9}$$

$$C(y) = C_2 \frac{w_t^{\frac{2-2\alpha}{2-\alpha}}}{k_t^{\frac{\alpha}{2-\alpha}}} y^{\frac{2}{2-\alpha}} \tag{B.10}$$

$$C_0 = \frac{\alpha}{2\chi(1-\alpha)} \tag{B.11}$$

$$C_1 = \frac{\alpha}{2\chi(1-\alpha)} (1-\alpha)^{\frac{2-\alpha}{\alpha}} \tag{B.12}$$

$$C_2 = \left(\frac{2-\alpha}{2-2\alpha} \right) \left(\frac{\alpha}{2\chi(1-\alpha)} \right)^{\frac{-\alpha}{2-\alpha}} \tag{B.13}$$

$$C_3 = \frac{2C_1^{\frac{\alpha}{2-\alpha}} - 2C_2}{\alpha C_1^{\frac{\alpha}{2-\alpha}}} \tag{B.14}$$

$$h_{1t}(p_{it}, k_{it}) = \int \left(\int_0^{\overline{b_t}} \left[p_{it} y_{it}^d - C\left(y_{it}^d \right) \right] p_b(x) dx \right) dF_A \tag{B.15}$$

$$h_{2t}(p_{it}, k_{it}) = \int \left(\int_{\overline{b_t}}^{\infty} \left[p_{it} y_{it}^C - C\left(y_{it}^C \right) \right] p_b(x) dx \right) dF_A \tag{B.16}$$

$$h_t = E_{t-1} \pi_{it}(p_{it}, k_{it}) = h_{1t} + h_{2t} \tag{B.17}$$

The FOC for p_{it} is given by

$$\frac{\partial h_t}{\partial p_{it}} = \frac{\partial h_{1t}}{\partial p_{it}} + \frac{\partial h_{2t}}{\partial p_{it}} = 0$$

And for k_t

$$\frac{\partial h_t}{\partial k_{it}} = \frac{\partial h_{1t}}{\partial k_{it}} + \frac{\partial h_{2t}}{\partial k_{it}} = R_t - (1-\delta)$$

Exogenous Shock Processes

$$b_{it} \underset{d}{\sim} ln\mathcal{N}(-\frac{\sigma_b^2}{2}, \sigma_b) \tag{B.18}$$

$$ln(A_t) = \rho_A ln(A_{t-1}) + \epsilon_{At} \tag{B.19}$$

$$\epsilon_{At} \underset{d}{\sim} \mathcal{N}(0, \sigma_A) \tag{B.20}$$

B.2 Estimation Details

B.2.1 Moment Conditions

The aggregate (mean) capital utilization rate can be found with

$$\int \frac{\tilde{k}_{it}}{k_t} di = \int \frac{(C_0 w_t)^{\frac{1-\alpha}{2-\alpha}}}{k_t^{\frac{1-\alpha}{2-\alpha}}} y_{it}^{\frac{1}{2-\alpha}}\ di \tag{B.21}$$

$$= \frac{(C_0 w_t)^{\frac{1-\alpha}{2-\alpha}}}{k_t^{\frac{1-\alpha}{2-\alpha}}} \left(\frac{A_t I_t}{P_t}\right)^{\frac{1}{2-\alpha}} \left(\int_0^{\frac{b_t}{A_t}} x^{\frac{1}{2-\alpha}} p_b(x) dx\right) + C_0^{\frac{1-\alpha}{2-\alpha}} C_1^{\frac{1}{2-\alpha}} w_t^{-\frac{1-\alpha}{\alpha}} p_t^{\frac{1}{\alpha}} \left(\int_{\frac{b_t}{A_t}}^{\infty} p_b(x) dx\right) \tag{B.22}$$

We can express the aggregate (mean) capacity utilization rate by

$$\int \frac{y_{it}}{y_{it}^C} di = \left(\frac{A_t^2 I_t}{A_t^{\frac{1}{\sigma}} p_t^{\frac{2}{\alpha}} k_t}\right) w_t^{\frac{2-2\alpha}{\alpha}} \left(\int_0^{\frac{b_t}{A_t}} x p_b(x) dx\right) + 1 \tag{B.23}$$

In the estimation, we assume that wages w_t are given. Then as long as p_{it} and k_{it} are only functions of the parameters (χ, σ_b) and the aggregate expenditure $I_t = Y_t$. The aggregate expenditure is itself only a function of A_t. Then the above capital utilization and capacity utilization rates are also only functions of (χ, σ_b) and, through its effect on I_t, also a function of A_t. We can write p_{it} and k_{it} as functions of (χ, σ_b) using the FOCs from every firm's (common) one period ahead decisions.

To get the FOCs for the firm's one period ahead decisions, first take the partial derivatives of the components of the expected profit function:

$$\frac{\partial h_{1t}}{\partial p_{it}} = \left(\int_0^{\overline{b_t}} \left(y_{it}^d + (p_{it} - C'(y_{it}^d)) \left(\frac{\partial y_{it}^d}{\partial p_{it}} \right) \right) p_b(x) dx \right) + \left[p_{it} \overline{y_{it}^d} - C\left(\overline{y_{it}^d} \right) \right] p_b(\overline{b_t}) \frac{\partial \overline{b_t}}{\partial p_{it}}$$
$$\frac{\partial h_{2t}}{\partial p_{it}} = \left(y_{it}^C + (p_{it} - C'(y_{it}^C)) \left(\frac{\partial y_{it}^C}{\partial p_{it}} \right) \right) \left(\int_{\overline{b}_{it}}^{\infty} p_b(x) dx \right) - \left[p_{it} y_{it}^C - C\left(y_{it}^C \right) \right] p_b(\overline{b}_{it}) \frac{\partial \overline{b_t}}{\partial p_{it}}$$

$$\frac{\partial h_{1t}}{\partial k_{it}} = \left(\frac{\alpha}{(2-\alpha) k_{it}} \int_0^{\overline{b_t}} C(y_{it}^d) p_b(x) dx \right) + \left[p_t \overline{y_{it}^d} - C\left(\overline{y_{it}^d} \right) \right] p_b(\overline{b_t}) \frac{\partial \overline{b_t}}{\partial k_t}$$
$$\frac{\partial h_{2t}}{\partial k_{it}} = \left((p_t - C'(y_t^C)) \frac{\partial y_t^C}{\partial k_t} + \frac{\alpha}{(2-\alpha) k_{it}} C(y_t^C) \right) \left(\int_{\overline{b_t}}^{\infty} p_b(x) dx \right) - \left[p_t y_t^C - C\left(y_t^C \right) \right] p_b(\overline{b_t}) \frac{\partial \overline{b_t}}{\partial k_t}$$

where $\overline{y_{it}^d} = y_{it}^d(\overline{b_t})$.

Expanding further,

$$\frac{\partial h_t}{\partial p_{it}} = (1-\sigma) \frac{I_t}{P_t} + \sigma (y_t^C)^{\frac{-\alpha}{2-\alpha}} \left(\frac{I_t A_t^2}{P_t} \right)^{\frac{2}{2-\alpha}} \left(\int_0^{\frac{\overline{b_t}}{A_t}} x^{\frac{2}{2-\alpha}} p_b(x) dx \right) + y_t^C \left(\int_{\frac{\overline{b_t}}{A_t}}^{\infty} p_b(x) dx \right) \tag{B.24}$$

$$\frac{\partial h_t}{\partial k_{it}} = \frac{\alpha}{2-\alpha} C_2 \left(\frac{I_t A_t^2}{P_t} \frac{w_t^{1-\alpha}}{k_{it}} \right)^{\frac{2}{2-\alpha}} \left(\int_0^{\frac{\overline{b_t}}{A_t}} x^{\frac{2}{2-\alpha}} p_b(x) dx \right) + \left(\frac{\alpha y_t^C}{p_{t-1} k_{t-1}} \right) \left(\int_{\frac{\overline{b_t}}{A_t}}^{\infty} p_b(x) dx \right) \tag{B.25}$$

We can now use equations (B.22) and (B.23) together with (B.24) and (B.25) to obtain moment conditions.

B.2.2 Estimation Procedure

Call the two model moments $f_{1t}(\chi, \sigma_b|A_t) = \int \frac{\tilde{k}_{it}}{k_t} di$, $f_{2t}(\chi, \sigma_b|A_t) = \int \frac{y_{it}}{y_{it}^C} di$, and their observed analogs m_{1t} and m_{2t}. Through their dependence on output, these functions also depend on the aggregate shocks A_t. Roughly, then, the estimation procedure proceeds in two steps: first, given a set of possible parameters $(\hat{\chi}, \hat{\sigma_b})$, calculate the model moments for the entire distribution of possible shocks A_t. Second, calculate the expected difference between the model and corresponding data moments, given the prior distribution of shocks A_t; if the difference is large, adjust the prior distribution, and repeat until the difference is minimized. Third, in an outer step, repeat this calculation for all periods t. Fourth, calculate a likelihood of observing the entire sequence of data moments, given your minimized shock distribution in each period. Last, repeat the entire process for other possible parameter sets $(\hat{\chi}, \hat{\sigma_b})$ until the likelihood is minimized.

In solving the inner problem of adjusting the probability distribution of shocks A_t, we make use of the fact that the A_t has a parametric prior distribution with mean 0, and so reduce the problem of finding the best distribution to one of finding the best σ_A. In order to force the algorithm to search over many parameter sets, we assume that the observations m_{1t} and m_{2t} are observed with measurement error. We employ a Bayesian particle filter to solve the inner problem and generate a likelihood function. We then minimize the likelihood function using the Metropolis-Hastings algorithm; both of these are described in more detail in Fernandez-Villaverde (2009).

B.2.3 Results

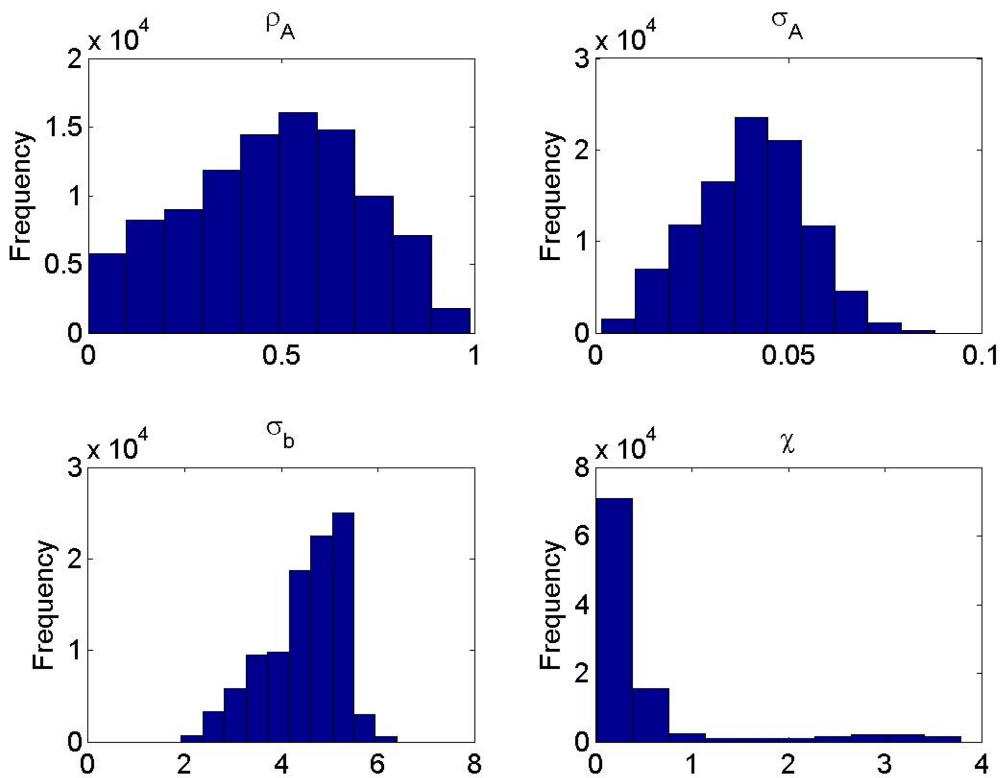
We assume the following prior distributions on each of the estimated parameters:

Table B.1: Assumed Priors

Parameter		Prior
Utilization Cost	χ	$ln\mathcal{N}(0,1)$
Std. Dev. Demand Shock	σ_b	$ln\mathcal{N}(-2,1)$
Autocorr. Agg Shock	ρ_A	$Uniform(0,1)$
Std. Dev. Agg Shock	σ_A	$ln\mathcal{N}(-2,1)$

We list our parameter estimates in the main text. Here, we plot the approximate posterior distribution for each of the four parameters of interest.

Figure B.1: Posterior Distributions: Parameter Estimates



Bibliography

Admati, A. R. and M. Hellwig (2013). *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It.* Princeton University Press.

Altunbas, Y., L. Evans, and P. Molyneux (2001). Bank Ownership and Efficiency. *Journal of Money, Credit and Banking*, 926–954.

Bachmann, R. and C. Bayer (2011). Investment dispersion and the business cycle. *NBER Working Paper* (w16861).

Bai, J. and S. Ng (2005). Tests for Skewness, Kurtosis, and Normality for Time Series Data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Banister, D. and J. Berechman (2000). *Transport Investment and Economic Development.* Spon Press.

Basu, K. (1997). *Analytical Development Economics.* MIT Press, Cambridge, MA.

Basu, S., J. Fernald, J. Fisher, and M. Kimball (2009). Sector-Specific Technical Change.

Basu, S., J. G. Fernald, and M. S. Kimball (2006). Are Technology Improvements Contractionary? *American Economic Review 96*(5), 1418–1448.

Beck, T., A. Demirgüç-Kunt, and R. Levine (2007). Bank Concentration and Fragility. *The Risks of Financial Institutions*, 193.

Bloom, D., D. Canning, and J. Sevilla (2003). Geography and Poverty Traps. *Journal of Economic Growth 8*(4), 355–378.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta Eksten, and S. Terry (2012). Really Uncertain Business Cycles. *NBER Working Paper* (w18245).

Boissay, F., F. Collard, and F. Smets (2013). Booms and Systemic Banking Crises. *ECB Working Paper, No. 1514*.

Boyd, J. H. and G. De Nicolo (2005). The Theory of Bank Risk Taking and Competition Revisited. *Journal of Finance 60*(3), 1329–1343.

Bremus, F., C. Buch, K. Russ, and M. Schnitzer (2013). Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity. (w19093).

Brunnermeier, M. K. (2009). Deciphering the Liquidity and Credit Crunch 2007-2008. *Journal of Economic Perspectives 23*(1), 77–100.

Carlstrom, C. T. and T. S. Fuerst (1997). Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. *American Economic Review*, 893–910.

Christiano, L., R. Motto, and M. Rostagno (2010). Financial Factors in Economic Fluctuations.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*(1), 1–45.

Christiano, L. J. and D. Ikeda (2011). Government Policy, Credit Markets and Economic Activity. *NBER Working Paper* (w17142).

Clements, M. and H. Krolzig (2003). Business Cycle Asymmetries. *Business & Economic Statistics 21*(1), 196–211.

Cocco, J. F., F. J. Gomes, and N. C. Martins (2009). Lending Relationships in the Interbank Market. *Journal of Financial Intermediation 18*(1), 24–48.

Cooper, R. (1994). Equilibrium Selection in Imperfectly Competitive Economies with Multiple Equilibria. *Economic Journal 104*(426), 1106–1122.

Corbae, D. and P. D'Erasmo (2010). A Quantitative Model of Banking Industry Dynamics. Technical report, Society for Economic Dynamics.

De Nicolo, G., P. Bartholomew, J. Zaman, and M. Zephirin (2004). Bank Consolidation, Internationalization, and Conglomeration: Trends and Implications for Financial Risk. *Financial Markets, Institutions & Instruments 13*(4), 173–217.

DeLong, J. and L. Summers (1986). Are Business Cycles Symmetric? *NBER Working Paper Series 1444*.

Doss, C. (2003). Understanding Farm Level Technology Adoption: Lessons Learned from CIMMYT Micro Surveys in Eastern Africa. *Mexico DF, CIMMYT 20*.

Durlauf, S. (1993). Nonergodic Economic Growth. *Review of Economic Studies 60*(2), 349–366.

Eisfeldt, A. L. and A. a. Rampini (2006). Capital Reallocation and Liquidity. *Journal of Monetary Economics 53*(3), 369–399.

Ellison, G. (1993). Learning, Local Interaction, and Coordination. *Econometrica 61*(5), 1047–1071.

Eshel, I., L. Samuelson, and A. Shaked (1998). Altruists, Egoists, and Hooligans in a Local Interaction Model. *American Economic Review 88*(1), 157–179.

Evenson, R. and L. Westphal (1995). Technological Change and Technology Strategy. *Handbook of Development Economics 3*, 2209–2299.

Fecht, F., K. G. Nyborg, and J. Rocholl (2011). The Price of Liquidity: the Effects of Market Conditions and Bank Characteristics. *Journal of Financial Economics 102*(2), 344–362.

Federal Reserve Bank of Chicago (2013). Commercial Bank Data. Available online at `www.chicagofed.org/webpages/banking/financial_institution_reports/commercial_bank_data.cfm`.

Federal Reserve Bank of New York (2009). Quarterly Trends for Consolidated US Banking Organizations. Technical report, Federal Reserve Bank of New York.

Federal Reserve Bank of St. Louis (2013). Assets and Liabilities of Commercial Banks in the United States. Available online at `http://research.stlouisfed.org/fred2/release?rid=22`.

Fernald, J. (2012). A Quarterly, Utilization-Adjusted Series on Total Factor Productivity. *Manuscript, Federal Reserve Bank of San Francisco*.

Fernández-Villaverde, J. (2010). The econometrics of DSGE models. *SERIEs 1*(1-2), 3–49.

Furfine, C. H. (1999). The Microstructure of the Federal Funds Market. *Financial Markets, Institutions & Instruments 8*(5), 24–44.

Galeotti, A., S. Goyal, M. Jackson, F. Vega-Redondo, and L. Yariv (2010). Network Games. *Review of Economic Studies 77*(1), 218–244.

Gertler, M. and P. Karadi (2011). A Model of Unconventional Monetary Policy. *Journal of Monetary Economics 58*(1), 17–34.

Gertler, M. and N. Kiyotaki (2010). Financial intermediation and credit policy in business cycle analysis. *Handbook of Monetary Economics 3*(3), 547–599.

Gilchrist, S. and J. C. Williams (2005). Investment, capacity, and uncertainty: a putty–clay approach. *Review of Economic Dynamics 8*(1), 1–27.

Gorodnichenko, Y. and M. Shapiro (2011). Using the Survey of Plant Capacity to Measure Capital Utilization. *US Census Bureau CES*, 11–19.

Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. (2005). What's Real about the Business Cycle? Technical Report w11161, National Bureau of Economic Research.

Hansen, G. D. and E. C. Prescott (2005). Capacity constraints, asymmetries, and the business cycle. *Review of Economic Dynamics 8*(4), 850–865.

He, Z. and A. Krishnamurthy (2012). A Macroeconomic Framework for Quantifying Systemic Risk. *Fama-Miller Working Paper*, 12–37.

Ivashina, V. and D. Scharfstein (2010). Bank Lending during the Financial Crisis of 2008. *Journal of Financial Economics 97*(3), 319–338.

Janicki, H. and E. Prescott (2006). Changes in the Size Distribution of US banks: 1960-2005. *Federal Reserve Bank of Richmond Economic Quarterly 92*(4), 291–316.

Jermann, U. and V. Quadrini (2012). Macroeconomic Effects of Financial Shocks. *The American Economic Review 102*(1), 238–271.

Kandori, M., G. Mailath, and R. Rob (1993). Learning, Mutation, and Long Run Equilibria in Games. *Econometrica: Journal of the Econometric Society 61*(1), 29–56.

Kehrig, M. (2011). The Cyclicality of Productivity Dispersion. *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*.

McKay, A. and R. Reis (2008). The Brevity and Violence of Contractions and Expansions. *Journal of Monetary Economics 55*(4), 738–751.

Mitchell, W. C. (1927). The Processes Involved in Business Cycles. In *Business Cycles: the Problem and its Setting*, pp. 1–60. NBER.

Murphy, K., A. Shleifer, and R. Vishny (1989). Industrialization and the Big Push. *Journal of Political Economy 97*(5), 1003–1026.

Neftci, S. N. (1984). Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy*, 307–328.

Puri, M., J. Rocholl, and S. Steffen (2011). Global Retail Lending in the Aftermath of the US Financial Crisis: Distinguishing between Supply and Demand Effects. *Journal of Financial Economics 100*(3), 556–578.

Rezitis, A. N. (2006). Productivity Growth in the Greek Banking Industry: a Non-parametric Approach. *Journal of Applied Economics 9*(1), 119–138.

Sichel, D. E. (1993). Business Cycle Asymmetry: a Deeper Look. *Economic Inquiry 31*(2), 224–236.

US Census Bureau (2014). Quarterly survey of plant capacity utilization. Available online at bhs.econ.census.gov/bhs/pcu/index.html.

Index

Abstract, vii
Acknowledgments, v
Appendices, 115
Appendix
  Bank Size and Downturns, 116
  Capacity Constraints, 123

Bank Size, Leverage, and Financial Downturns, 1
Bibliography, 136

Dedication, iv

Introduction, 61

Technology Spillovers and Local Interactions, 84

Vita

Chacko was born in Columbia, MD on 12 May 1983. He loved (and continues to love) school: he received a Bachelor of Science degree in Electrical Engineering and a Bachelor of Arts in Government from the University of Texas at Austin in 2005, and a Master of Arts in International Relations at the University of Chicago in 2006. It was at Chicago that he realized he was destined for a career in economics. He was accepted to the economics PhD program at the University of Texas at Austin in 2008.

Permanent address: 2225 Speedway
Stop C3100
Austin, Texas 78712

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.