ESSAYS IN BANKING AND REGULATION

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Tirupam Goel
May 2016

ESSAYS IN BANKING AND REGULATION

Tirupam Goel, Ph.D.
Cornell University, 2016

Abstract

The broad goal of this dissertation is to further our understanding of the relationship between real and financial sectors of an economy, to identify inefficiencies in financial sector intermediation, and to design financial regulation policies that can address these inefficiencies. The three chapters of this dissertation contribute to specific aspects of the above goal.

In the first chapter, I develop a general equilibrium macroeconomic model with a dynamic banking sector in order to characterize optimal size-dependent bank leverage regulation. Bank leverage choices are subject to the risk-return trade-off, and are inefficient due to financial frictions. I show that leverage regulation can generate welfare gains, and that optimal regulation is tighter relative to the benchmark and is bank-size dependent. In particular, optimal regulation is tighter for large banks relative to small banks, and it leads to the following welfare generating effects. First, as small banks take more leverage, they grow faster conditional on survival, leading to a *selection effect*. Second, small bank failures are less costly while entrants have higher relative efficiency, leading to a *cleansing effect*. Third, tighter regulation for large banks reduces their failure rate, which generates welfare since large banks are more efficient and costlier to replace, leading to a *stabilization effect*. The calibrated model rationalizes various steady state moments of the US banking industry, and points towards qualitatively similar but quantitatively tighter leverage regulation relative to the proposition in Basel III accords.

In the second chapter, I study the financial contagion problem when banks in order to hedge against idiosyncratic shocks, engage in two-dimensional as opposed to one-dimensional interactions with other banks. To this end, I develop a double-edge interbank network model where banks engage in debt contract and securitization transactions with other banks. I show that the standard intuition of financial contagion does not translate from the one-dimensional case to the two-dimensional case *i.e.* financial contagion can either weaken or worsen depending on the network and parameter configuration. In particular, I derive parametrization for the case where financial contagion worsens.

In the third chapter, we investigate whether countercyclical capital-ratio regulation (CCR) should be implemented strictly as a rule, or whether regulators should have discretion with respect to the timing and magnitude of changes in capital-ratio requirement. Using a simple model we prove the proposition that under information asymmetry, discretionary CCR leads to an increase in policy uncertainty relative to rule-based CCR. We prove a similar proposition for a general finite-horizon economy. Finally, we document that since discretionary CCR enables the regulator to respond to unexpected shocks, a benevolent regulator faces the welfare trade-off while choosing between rule-based and discretionary CCR.

BIOGRAPHICAL SKETCH

Tirupam was born in Calcutta (Kolkata), India, where he lived with his parents for the first eighteen years of his life. After completing his school education in 2004, Tirupam left Calcutta to attend the Indian Institute of Technology (IIT) Kanpur for an integrated bachelor's-master's degree in mathematics. At IIT he met some truly inspiring colleagues and teachers, interactions that came to play the most formative role in his early twenties — he also decided to pursue doctoral studies during this period. He graduated from IIT Kanpur in 2009 with excellent grades and a best master's thesis award, after which he worked in the banking industry for a couple of years. It is also during this time that he married Jyoti. In 2011, Tirupam embarked upon his doctoral study in economics at Cornell University, which has been the most enriching professional experience for him till date. He expects to graduate from Cornell in May of 2016 with this dissertation, and M.A. and Ph.D. degrees in economics. He feels very fortunate to have had the opportunity to study at two great institutions — IIT Kanpur and Cornell. In his free time, Tirupam likes to cook, paint with oil on canvas, or listen to music from the Indian subcontinent.

Dedicated to
my grandfather,
my parents,
and my wife

ACKNOWLEDGEMENTS

I would like to begin by expressing my most sincere gratitude towards my advisors Assaf Razin and Maxim Troshkin. Their advice and critique prompted me to think deep and fundamental, and achieve greater quality of my work. My learnings under their supervision will continue to positively influence me for a long time. No words can describe my gratitude towards my advisors.

I would like to express my immense gratitude towards Eswar Prasad and Karel Mertens, with whom I had several insightful discussions at various stages of my dissertation research. Their questions and suggestions have always nudged me in better directions and have prompted me to think about big-picture as well as micro-foundation issues. Their advice played a critical role in the development of my research aptitude.

I would like to sincerely and especially thank John Abowd for giving me the opportunity to work with him, and for giving me maximum flexibility as his research assistant. His support at various stages and in various ways has made life in graduate school much easier for me and my wife. I would like to thank Lars Vilhuber for the several useful comments and suggestions he gave on my work. I would like to thank my colleagues at the Labor Dynamics Institute (LDI), and also express thanks for the excellent working environment I had as part of the LDI team.

I am thankful for the beneficial discussions I had with Julieta Caunedo and Christopher Huckfeldt, and for the comments from Cornell Macro-lunch participants. Thanks to the Department of Economics for providing all the resources one needs to excel in graduate school, and to the Cornell Institute for Social and Economics Research (CISER) for their state-of-the-art computational resources without which, this dissertation could not have been finished in time, literally.

I want to extend my special thanks to Praveen Kulshreshtha, without whose support and encouragement, I would not have been able to pursue graduate studies at Cornell. I would like to thank Amit Mitra, Sharmishtha Mitra, Debasis Kundu, and Seema Nanda for their belief in me. And I would like to thank my alma-mater IIT Kanpur for the most inspiring and intellectually stimulating undergraduate education I could hope for.

Graduate school is a journey with ups and downs, and I was very fortunate to have some close friends during the same. Debi has been *the* mentor and friend since day one at Cornell. Mehreen and Sanket have been true friends and well wishers. Isha has been the friend and colleague with whom I really enjoyed talking macroeconomics. Thanks to my friends for all the revitalizing parties, game nights and philosophical discussions we enjoyed together.

Finally, I would like to express my utmost gratefulness towards my wife Jyoti for encouraging me to pursue my aspirations at the cost of hers, for patiently supporting me throughout these five years, and for always standing beside me in my pursuit with her love and care. It is because of her love and affection that graduate school has been a very smooth and beautiful experience for me.

This dissertation and several of my accomplishments in life would not have been possible without the blessings and sacrifices of my grandparents Shri Ram Singh Goel and Shrimati Sheela Rani Goel, and the blessings and support of my parents Shri Rajiv Kumar Goel and Shrimati Kusum Goel who toiled tirelessly during my childhood so that I can get the best education, who prayed during my undergraduate and graduate education for my success, and who encouraged me to pursue my aspirations at the cost of allowing their child to live away from them. No set of words can express my feelings and gratitude towards them.

Contents

Chapter 1: Banking Industry Dynamics and Size-Dependent Leverage

1 Introduction **1**

2 Leverage and Bank Dynamics **10**

3 Model **12**
- 3.1 Households . . . 13
- 3.2 Firms . . . 15
- 3.3 Capital Firm . . . 16
- 3.4 Banks . . . 17
- 3.5 Analytical properties . . . 25
- 3.6 Bank distribution . . . 26

4 Stationary General Equilibrium **27**
- 4.1 Existence and solution . . . 29

5 Data and Estimation **32**
- 5.1 Computations . . . 35

6 Counterfactual Experiment **37**
- 6.1 Optimal regulation . . . 40
- 6.2 Sensitivity Analysis . . . 45
- 6.3 Aggregate shocks . . . 47

7 Conclusion **48**

Chapter 2: Financial Contagion in Multiple-edge Interbank Networks

1 Introduction **51**

2 Contagion in multiple-edge networks **54**
- 2.1 Model . . . 55

3 Empirical evidence **60**

4 Model **61**
- 4.1 Shocks . . . 62
- 4.2 Interbank Networks . . . 65
- 4.3 Single-edge network: fixed asset return case . . . 65
- 4.4 Multiple-edge network: stochastic return case . . . 69

5 Conclusion **75**

Chapter 3: Countercyclical capital regulation:
Rules versus Discretion

1 Introduction **77**

2 Rules versus Discretion **80**
2.1 Time inconsistency . . . 80
2.2 Policy uncertainty . . . 83
2.3 Rules versus Discretion . . . 86

3 Model **87**
3.1 Main result in a simple setting . . . 87
3.2 General finite economy . . . 94

4 Conclusion **98**

A Appendix to Chapter 1 **99**
A.1 Proofs . . . 99
A.2 Data analysis . . . 104
A.3 Additional Figures . . . 106

B Appendix to Chapter 3 **110**
B.1 Proof . . . 110

Chapter 1

Banking Industry Dynamics and Size-Dependent Leverage Regulation

1 Introduction

A bank's leverage decision,[1] where leverage is defined as the ratio of bank assets to bank net worth, has implications for growth and solvency. On one hand, leverage increases the potential for higher return on net worth.[2] On the other hand, leverage increases variance in returns on net worth and the probability of bank failure, where failure is defined as the event of a bank's net worth falling below zero. In this sense, bank leverage decisions have implications for banking industry dynamics that include entry, exit and size distribution, and for such macroeconomic aggregates as output, investment, and distribution of consumption. In the presence of financial frictions in the banking industry, bank leverage decisions can be inefficient, and the heterogeneity of bank size and bank productivity matters for the size of this inefficiency. For example, the failure of a large, more productive bank, an event whose occurrence depends on its leverage, is socially more costly than the failure of a small, less productive bank. Therefore, there are potential gains from regulating leverage, and more so from having a leverage regulation that depends on the size of a bank.[3] The objective of this paper is to characterize the optimal *size-dependent* leverage regulation.[4]

[1]In this paper, "bank" is an umbrella term for financial intermediaries, and is meant to nest the concepts of commercial banks, savings and loan associations, investment banks, etc.

[2]High leverage, i.e., more debt financing relative to equity financing, also implies the tax advantage of debt over equity, but this feature is not pursued in this paper.

[3]I am assuming that bank productivity is not observed by the regulator, and therefore, regulation cannot be a function of bank productivity.

[4]Both net worth and assets are appropriate measures of bank size, and there is an equivalence mapping between these two approaches. I follow the former measure for mathematical convenience.

The question of bank leverage regulation has gained considerable importance since the 2008 financial crisis,[5] as high leverage played a critical role in rendering the scale of the crisis. In this regard, among other propositions, Basel:III [2011] proposes not just an overall tightening of regulation relative to Basel II, but also that systemically important financial institutions (SIFIs) must face tighter regulation relative to smaller banks. In implementing these two accords of Basel III, the regulator faces several trade-offs. The trade-off with respect to an overall tightening of leverage regulation has been the subject of numerous studies, for example Van den Heuvel [2008] and Corbae and D'erasmo [2014], and is also studied in this paper. However, the focus in this paper is on the trade-offs associated with imposing a size-dependent leverage regulation. For example, if the regulator imposes *relatively* tighter leverage regulation on smaller banks, it might hinder their growth and end up protecting larger banks. The latter consequence might be undesirable from an anti-trust perspective. Also, if larger banks are allowed relatively greater leverage, and if they end up assuming more leverage, larger banks become more prone to failure, which is socially costly. In the left panel of Figure 1, I document that on failing, larger banks[6] incur bigger costs for the economy.[7] On the other hand, if the regulator imposes *relatively* tighter leverage regulation for large banks, this might reduce the aggregate volume of financial intermediation by restricting bank lending, where large banks play the pivotal role. Also, large banks are more efficient,[8] as discussed in Hughes and Mester [2013] and several other empirical paper, and such a regulation might create disincentives for banks to grow large and efficient. Therefore, to determine the most effective size-dependent regulation, it is essential to have a framework in which one can study these trade-offs quantitatively.

To this end, I develop a model designed to examine the relationship between bank leverage, banking industry dynamics, and macroeconomic aggregates. I build on the real business cycle (RBC) model, and introduce a dynamic banking sector that consists of a continuum of heterogeneous banks. There is no aggregate, but only idiosyncratic uncertainty. The banking sector is modeled in the spirit of Hopenhayn [1992], but with debt financing and financial friction in the form

[5] Leverage regulation and capital regulation are related but distinct concepts. Following the convention used by the Bank of International Settlements (BIS) in its Basel accords, capital regulation computes capital charge based on risk weighted assets, whereas leverage regulation computes capital charge based on consolidated assets. Throughout this paper, I will refer to leverage regulation. However, since my model has a single risky asset held by banks, the notion of capital regulation and leverage regulation are coincident.

[6] Size of a bank is measured by its net-worth in this paper

[7] The net worth reported in the figure is as of the last regulatory filing by the bank before its failure. The estimated loss is the amount of FDIC funds disbursed to meet the shortfall in liabilities. In this sense, the estimated loss captures only the idiosyncratic effect of the failure of a bank. To capture the spillover effect of the failure of a bank through the fire sale channel, I need asset holding data on banks, which I do not have access to.

[8] For example, because they learn or adopt technology over time, in the process of growing large.

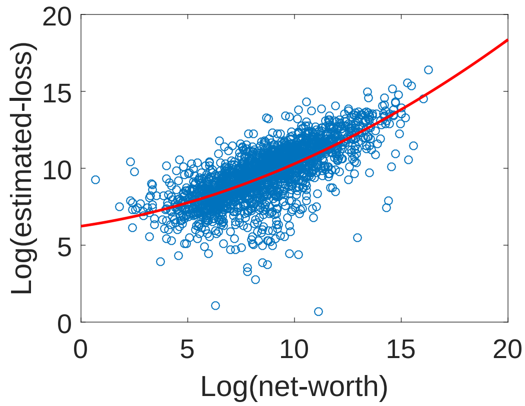


Figure 1: *(Left)* US savings and commercial banks: Log(net-worth) and log(estimated-loss) during 1985 - 2014; *(Right)* Log(net-worth) and ratio of total operating income to net-worth in 2000:Q4 (which is approximately the mid-point of the sample) *(source: FDIC Historical Statistics on Banking, and Call and Thrift Financial Reports)*

of credit constraints. Banks are essentially intermediaries that accept deposits from households and purchase shares in firms that operate the standard production technology. Banks have idiosyncratic efficiency in managing their equity in firms, and this idiosyncratic efficiency is modeled as a first-order Markov process. To motivate the key financial friction in the model, a limited commitment issue *à la* Gertler and Kiyotaki [2010] is assumed between banks and their depositors. Herein banks are assumed to be able to divert assets for personal consumption of their shareholders. To ensure that banks do not divert assets, they are subject to a borrowing constraint that limits the amount of deposits they can raise. The borrowing constraint ensures that the value of diverting assets is not larger than the value of continuing operations. Incumbent banks fail when net worth drops below zero, which can happen because of a large shock, or because of high leverage. Post failures, new banks enter as per a free entry process.

To quantify the distributional welfare consequences of a size-dependent policy, I consider a household sector that consists of workers and bankers.[9] The workers supply labor inelastically to earn wages. The bankers are heterogeneous and are modeled in the spirit of Aiyagari [1994], but with an income process which is endogenously determined. Bankers hold shares in the banking sector, and therefore receive dividend income from the bank whose shares they own. Since banks are subject to idiosyncratic efficiency shocks that govern dividend payments, bankers are subject

[9]In other words, a heterogeneous household setup is not required to solve the problem of the banking industry, or to understand the relationship between bank leverage and bank industry dynamics. A heterogeneous household setup is introduced to be able to talk about the welfare implications of changes in leverage policy.

to idiosyncratic dividend income shocks.

With regards to optimal regulation, the first main result of this paper is that, in steady state equilibrium, an overall tightening of the level of leverage regulation from 29.58 in the benchmark regime to 3.55 in the size-independent optimal policy regime generates utilitarian welfare gains of 2.2% in consumption-equivalent (CE) terms. This result points towards a more aggressive tightening of regulation relative to, for example, Nguyen [2014] and Begenau [2014], but reinforces the 25% capital ratio requirement proposed in Admati and Hellwig [2014],[10] and points in the same direction as total loss-absorbing capital (TLAC) requirements set by the Financial Stability Board. The second main result of this paper is that additional welfare gain of 8.1% in CE terms can be achieved, if leverage regulation is bank size specific.[11] In particular, the optimal size dependent regulation is characterized by relatively tighter regulation of 3.17 for large banks and a relatively liberal regulation of 4.12 for small banks. Qualitatively, the paper's results point in the same direction as the proposition in Basel III that large, systemically important financial institutions (SIFIs) should face tighter regulation. However, quantitatively, the paper's results suggest a more aggressive regulation relative to Basel III.

Intuitively, an overall tightening of regulation produces welfare gains primarily from lower taxes, lower entry and exit costs due to lower bank turnover rate, and higher dividend payments from a banks to households. The last effect is driven primarily by the fact that in the optimal regulation regime, there exists a larger mass of better capitalized banks that pay higher dividends in the aggregate. In the optimal regime, these gains dominate the welfare loss due to lower capital stock, output, and lower wages. The welfare gain from a size-dependent regulation follows from three channels. First, allowing small banks to take on more leverage allows them to grow faster, conditional on survival of the more efficient ones, leading to a *selection effect*. Second, although allowing small banks to take on more leverage results in a higher failure rate, these failures are less costly as the entrant banks easily replace failed banks in terms of size. Moreover, the mean efficiency of entrants is higher than the mean efficiency of failed banks, leading to a *cleansing effect*. Third, tighter regulation renders large banks less prone to failure, which improves welfare since large banks are more costly to replace as entrants need time to grow large, and in this process,

[10] A 25% capital ratio requirement is equivalent to a leverage regulation of 4 in a model with only one risky asset. With both risky and risk-free assets in the model, the optimal regulation would be less aggressive.

[11] These are large welfare gains relative to those found in other studies in literature. I argue that these gains are not implausible. For example, the cost of the 2008 financial crisis is of a similar order in permanent consumption terms, and when regulation can avoid the occurrence of such crises, the gains from regulation will be of this order. Technically, the model's predicted welfare gain will be smaller once transition dynamics are taken into consideration.

Year	Name	Highlights
1988	Basel I	(Tier1 + Tier2) Capital at least 8% of risk weighted assets; fixed risk weights
2004	Basel II	(i) revised definitions and computations, model-based risk weights; (ii) supervisory review; (iii) disclosure to strengthen market discipline
2011	Basel III	(i) tighter regulation; (ii) capital conservation buffer; (iii) countercyclical buffer; (iv) non-risk weighted capital regulation (leverage regulation); (v) liquidity requirements (vi) surcharge for SIFIs
2010	Dodd-Frank	(i) enhanced regulatory authority; (ii) identify and address distressed SIFIs; (iii) Volcker rule; (iv) expanding class of assets that are regulated *(among other accords of Dodd-Frank act)*

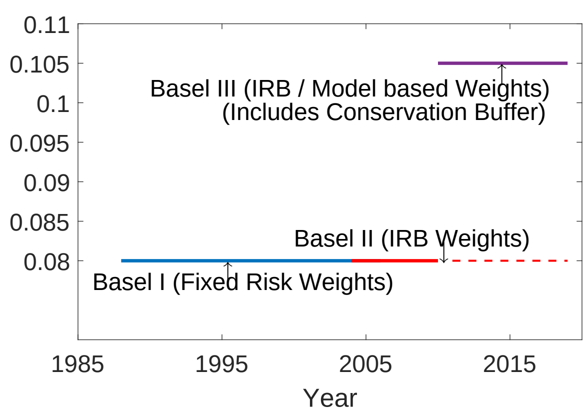
Table 1: Highlights of Bank Regulation in the US. Year is when the regulation was implemented, or when the implementation process begun.

incur period operating costs. In this sense, large banks, which are more efficient in equilibrium and intermediate more capital, are not replaced immediately, and therefore stabilizing them has welfare gains, the *stabilization effect* of size-dependent policy.

With regards to results the model produces on dynamics of the banking industry, I show that in the steady state, an equilibrium with entry, exit, and non-degenerate size distribution exists. In the stationary equilibrium, bigger banks are more highly levered, and there is positive correlation between bank size and bank efficiency. This follows intuitively from the fact that more efficient banks have better future prospects, hence they can therefore take on more leverage, and potentially grow faster. The paper also documents some results pertaining to how banking industry dynamics respond to changes in regulation. For example, a tightening of regulation results in the following effects (a weakening of regulation results in qualitatively opposite effects). First, the bank failure rate is lower, which increases the mean age of banks. This shifts the size distribution of banks to the right, and also results in a larger equilibrium mass of banks. The latter effect is instrumental in driving up aggregate dividend payments, even though each bank might not pay higher dividends. Second, since the mean leverage of banks is smaller, banks are better capitalized. In particular, the aggregate net worth of banks is larger. Third, tightening regulation results in a less concentrated industry as indicated by a lower Herfindahl Index.

Data and Stylized facts In order to characterize leverage regulation quantitatively, I use US commercial and savings banks Call and Thrift Financial Reports (CTR) data and Historical Statistics on Banking (HSOB) from Federal Deposit Insurance Corporation (FDIC). I draw household asset ownership and inequality data from Survey of Consumer Finances (SCF) and Organization for Economic Co-operation and Development (OECD) statistics, respectively. Aggregate macroeconomic data is obtained from the Federal Reserve Economic Data (FRED) database. I focus on the stationary equilibrium of the model, and estimate/calibrate model parameters to match select data moments.

I begin with some history of *capital regulation* of banks in the US, which has evolved considerably over time. The highlights are summarized in Table 1. Regulation has become generally tighter in its incidence and broader in scope over time (see, for example, Acharya [2011]). In particular, the cap on leverage has decreased, or equivalently, the minimum capital ratio requirement has increased (Please see the left panel of figure 2). Much of this tightening of regulation is motivated by the higher realized or estimated cost of a financial crisis.



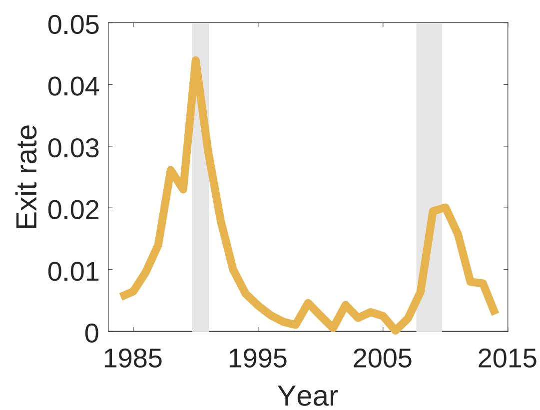


Figure 2: *(Left)* Historical Basel Total (tier1 + tier2) capital requirements *(Right)* US commercial and savings banks: Yearly failure rate (Source: FDIC Historical Statistics on Banking)

I now present some stylized facts observed in the data that I use to discipline my model. In the right panel of figure 2, I graph the yearly exit (failure) rate of US savings and commercial banks, which is one of the targets of my estimation procedure. The shaded regions approximately represent the two business cycle contractions that also featured financial sector crises: (i) the savings and loan crisis of the 1990s and (ii) the financial crisis of 2008. I take note of the following observations. First, exit rate peaks during the savings and loans crisis and during the financial crisis, and is stable otherwise, i.e., during 1995-2005. Even structurally, the 1995-2005 period is less affected by regulatory changes in the banking industry because most of the bank branching deregulation

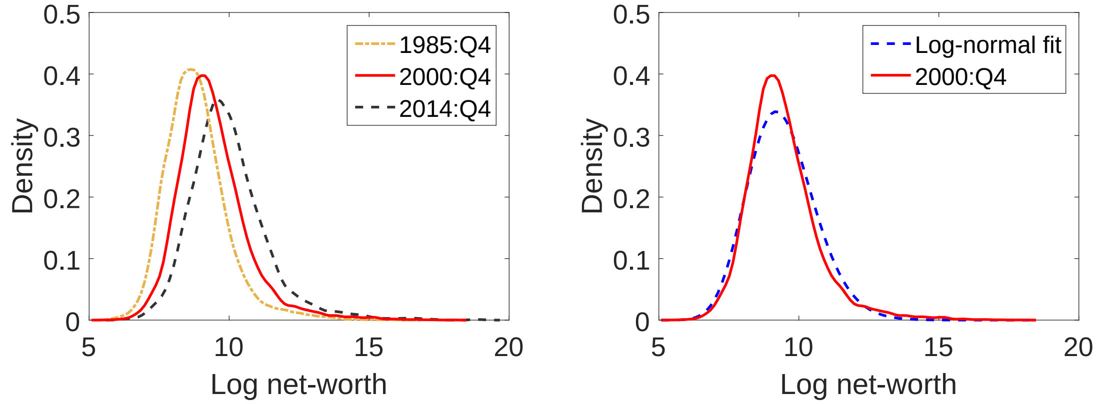


Figure 3: *(Left)* Kernel density estimate of log net-worth distribution; *(Right)* Log-normal fit to 2000Q4 data (source: FDIC Call and Thrift Financial Reports)

happened prior to 1995. In particular, although the Riegle-Neal Interstate Banking and Branching Efficiency Act was implemented in June 1997, US states were most active in removing geographic limits to banking prior to 1995 (Jayaratne and Strahan [1997]). Therefore, since 1995-2005 is a relatively stable period for banking industry dynamics, I use this period for calibration/estimation.

In Figure 3, I document the distribution of inflation-adjusted log net worth at three points of time during the last three decades. For any point in time during this period, the size distribution of bank net-worth in the United States can be fairly well approximated, based on the Kolmogorov-Smirnov Goodness of fit test, by a log-normal distribution. The fit for bank size distribution in 2000:Q4, for example, is shown in the right panel. Two remarks on the tail of the bank size distribution are as follows: (a) log-normal is not adequate to capture the tail, and a power law distribution can better describe the tail (b) the model in this paper does not attempt to explain the existence of a long / heavy tail of the bank distribution - it rather focuses on specific percentiles of the size distribution of banks. Over time, the distribution is shifting to the right. However, the inflation-adjusted distributions at any two points in time during 1995-2005 are not statistically significantly different based on the two-sample Kolmogorov-Smirnov (KS) test. In this sense, the size distribution has remained fairly stable during the calibration period.

Related Literature The first strand of literature to which this paper is related, aims to understand how inefficiencies in the banking sector affect the real economy, and to identify which tools regulators should use to address these inefficiencies. The key contribution of this paper helps us understand how, in the presence of financial frictions, leverage choices affect output, investment, and distribution of consumption through the banking industry dynamics channel. Few other papers have this feature. For example, Corbae and D'erasmo [2014] develop a quantitative model of an imperfectly competitive banking industry. Bank dynamics (entry and exit) are driven by shocks to borrowers production technology, and the authors show how competition between banks affects interest rates and borrowers' decisions regarding amount of loans and levels of risk. Covas and Driscoll [2014] study capital and liquidity regulation in a dynamic banking sector, and its effect on loan supply, output, and interest rates. Nguyen [2014] studies the welfare implications of capital requirements in the presence of moral hazard due to deposit insurance. Christiano and Ikeda [2013] study the welfare gains possible due to leverage restrictions in a macroeconomic model with unobservable banker effort. Begenau [2014] shows that higher capital requirements might actually increase bank lending when households value safe and liquid assets, providing another rationale for tighter regulation. Other papers that discuss time-invariant or counter-cyclical leverage regulation that is not dependent on bank sizes include Benes and Kumhof [2011], Zhu [2007] and Christensen et al. [2011].

In this paper, I consider a competitive banking industry in which credit friction arises because of a limited commitment issue between banks and its depositors. By assuming a stylized competitive banking industry instead of a more general imperfectly competitive industry, I am able to achieve tractability in studying endogenously determined size distribution. The fact that banking industry dynamics in general, and size distribution of banks in particular, respond to regulation in my model is the feature which allows me to study size-dependent regulation.

This paper is also related to the literature on macroeconomic models with financial sectors. For example, Gertler and Kiyotaki [2010] and Gertler and Karadi [2011] study unconventional monetary policy and other policy experiments in a macro model in which financial frictions arise due to a limited commitment issue. Adrian and Boyarchenko [2012] use a value-at-risk constraint to generate the endogenously determined borrowing constraints which play the central role in amplifying shocks. In this paper, the credit constraint arises because of a limited commitment issue between banks and their depositors. The main contribution here is to model a heterogeneous financial sector with endogenous entry and exit. In this sense, the paper offers an alternative channel, namely industry dynamics, through which to study the effect of the financial sector on macroeconomic aggregates.

This pursuit is related to another strand in the literature that studies the dynamics of non-financial firms. For example, Cooley and Quadrini [2001], Clementi and Hopenhayn [2006], Albuquerque and Hopenhayn [2004] study the dynamics of firms that borrow subject to borrowing constraints, but do not study the relationship between leverage and the risk of return on net worth or the risk of failure. The implication of borrowing on industry dynamics has been studied theoretically in Miao [2005], and empirically in, for example, Lang et al. [1996], but only in the context of non-financial firms. These studies are not directly applicable to financial firms (equivalently referred to as *banks* in this paper) as the capital structure choice of banks is characterized by higher leverage compared with non-financial firms (Kalemli-Ozcan et al. [2012]). In fact, as DeAngelo and Stulz [2013] show and Hanson et al. [2010] argue, high leverage is optimal for banks. This is because of the nature of the business of financial firms, wherein obtaining cheap funding is key to surviving the competition. Because of the fact that financial firms are more highly levered in general, the interaction between leverage and financial industry dynamics is also more pronounced, more so in the presence of financial frictions. In Section 2, I provide some intuitions for these mechanisms using simple examples. I incorporate these intuitions in my model to understand how financial leverage shapes financial industry dynamics in the presence of financial frictions. Some papers that follow a closely related route are Boyd and De Nicolo [2005], Corbae and D'erasmo [2013], and Corbae and D'erasmo [2014]. The former discusses endogenous bank failures in a static setting, where competition among banks reduces loan rates, incentivising borrowers to take on less risky projects and fail less often - which means fewer losses and failures for banks. The latter use models of imperfect competition to study a range of bank regulations, including capital requirements.

In studying the effect of industry dynamics on the real economy, in a non-financial context, Clementi and Palazzo [2015] investigate how firm entry and exit amplify aggregate fluctuations. The mechanism in their paper is that after a positive aggregate productivity shock, the number of entrants increases. Of these entrants, those who survive, generate a wider and longer expansion. This insight is closely related to the intuition in my paper, according to which allowing smaller banks to take on more leverage leads to a selection effect such that surviving entrants are able to grow faster on an average.

The remainder of the paper is organized as follows. In Section 2, I present examples to provide intuitions for model mechanisms. I present the model in Section 3, equilibrium in Section 4, discuss the data, estimation, and computations in Section 5, and counterfactual experiments in Section 6. Section 7 concludes.

2 Leverage and Bank Dynamics

In this section I provide intuitions supporting my view of how leverage affects bank dynamics. Leverage is defined as the ratio of total assets to net worth. Failure is defined as the event of a bank's net worth dropping below zero for the first time. In the examples below, a bank's stylized balance sheet of a bank is characterized by net worth and leverage tuple (n, l). As per this notation, assets equal ln and liabilities equal $(l-1)n$. I will assume that asset price is $Q = 1$, so that the number of units of an asset is same as the value of the asset. I focus on dynamics on an individual bank balance sheet in the event of negative shocks to its assets. The intuitions generated by these dynamics form the basis for the full model presented in Section 3.

Leverage and bank failures To illustrate the direct effect of leverage on failure, consider two banks of the same size as measured by net worth, holding the same assets. Bank A has balance sheet position (n, l_A) and bank B has balance sheet position (n, l_B). Assume, without loss of generality, $l_B > l_A$. Also assume that the liabilities of a bank do not change due to asset side shocks. Then the percentage drops ϵ_A, ϵ_B in asset prices required for bank A and bank B to fail independently are given as:

$$\epsilon_A = 1/l_A, \quad \epsilon_B = 1/l_B \implies \epsilon_B < \epsilon_A$$

Remark 1: *Given any probability distribution of asset price shocks, the probability of failure for a more highly levered bank is higher.*

Leverage and de-leveraging Consider the same two banks A and B such that $n_A = n_B$ and $l_B > l_A$. Assume that due to some exogenous heterogeneity, l_A and l_B are the optimal levels of leverage for banks A and B, respectively. Now consider a negative shock ϵ to asset prices, one that is not large enough for any of the banks to fail. Then, the post-shock leverage of the banks is:

$$l'_A = \frac{(1-\epsilon)l_A}{1-\epsilon l_A}, \quad l'_B = \frac{(1-\epsilon)l_B}{1-\epsilon l_B}$$

$$\implies l'_B > l'_A$$

$$\implies l'_B/l_B > l'_A/l_A$$

Remark 2: *A bank that is more highly levered compared with another bank will remain so after a negative asset price shock. The percentage increase in leverage is greater for a bank that is more* *highly levered in the first place.*

The first implication is intuitive. The second implication is more interesting for the following reason. Assume that, after a negative price shock, both banks re-optimize their balance sheets in order to get back to their respective pre-shock leverage levels. This re-optimization entails de-leveraging, i.e., reducing leverage. De-leveraging can be achieved either by raising fresh equity, or by selling assets to pay off liabilities. Since raising fresh equity is more difficult than selling assets in a business cycle downturn, I will assume that de-leveraging takes place through the sale of assets. Suppose bank i sells x_i units of assets at price $(1-\epsilon)$ to bring its leverage back to the (optimal) pre-shock level. For bank i, post-shock assets are $(1-\epsilon)nl_i$, and post-shock net worth is $(1-\epsilon l_i)n$. Then we have the following implication:

$$\frac{(1-\epsilon)nl_i - (1-\epsilon)x_i}{(1-\epsilon l_i)n} = l_i, \quad i \in \{A, B\}$$

$$\implies x_i = \frac{\epsilon n l_i (l_i - 1)}{1-\epsilon}$$

$$\implies x_B > x_A$$

This means that a more highly levered bank would sell more assets for a given shock. However, the more insightful implication follows from the comparison of percentage asset sale, defined as the ratio of units of assets sold to a bank's pre-shock assets. This implication, which follows trivially from the fact that $l_B > l_A$, is a key idea of this paper.

$$\tilde{x}_i = \frac{x_i}{nl_i} = \frac{\epsilon(l_i - 1)}{1-\epsilon}$$

$$\implies \tilde{x}_B > \tilde{x}_A$$

Remark 3: *Given a negative asset price shock, a more highly levered bank would sell a larger fraction of its assets while re-optimizing its balance sheet.*

This is where the fire sale externality issue bites because after a negative asset price shock, the volume of asset sale matters as each additional unit of assets sold in the market depresses asset prices further, reinforcing the initial shock. This mechanism constitutes an externality because banks do not internalize the effect of asset sales on the solvency of other banks.

Bank size matters In the above examples, I considered banks of the same size. In this example, I show that if two banks are equally levered but are of different sizes, the larger bank will have a larger effect on asset prices. Consider two banks with net worth and leverage positions (n_A, l) and (n_B, l) with $n_B > n_A$, i.e., bank B is larger than bank A, and both banks are equally levered. As before, I will assume that l is the optimal operating leverage ratio for these banks. Suppose there is an asset price shock that depresses asset prices by a fraction ϵ. The updated balance sheets are as follows:

$$A : \left((1-\epsilon l)n_A, \frac{(1-\epsilon)l}{1-\epsilon l} \right); \qquad B : \left((1-\epsilon l)n_B, \frac{(1-\epsilon)l}{1-\epsilon l} \right)$$

The leverage of both banks goes up by the same amount. As previously, in order to re-optimize their respective balance sheets, bank i would need to sell assets, say x_i units:

$$x_i = \frac{\epsilon n_i l(l-1)}{1-\epsilon} \quad \forall i \in \{A, B\} \implies x_B > x_A$$

Remark 4: *Given a negative asset price shock, among equally levered banks, the larger banks will sell more assets while re-optimizing their balance sheets.*

This mechanism in itself does not lead to inefficiency. However, in the presence of the fire sale issue, this mechanism does lead to an inefficiency. The size of a bank begins to matter simply because of the greater effect of a large bank on asset prices, and hence on other banks' balance sheets.

3 Model

Overview Time is discrete and the horizon is infinite. There is no aggregate uncertainty. The economy consists of the following agents: a unit mass of atomistic households, a representative capital goods producing firm, and a continuum of islands, each with a non-financial firm and a financial firm. The mass of households consists of a fraction λ of bankers and a fraction $(1-\lambda)$ of workers.[12] The workers supply labor and own capital firms, while the bankers own the banks. Capital goods producing firms (henceforth *capital firms*) operate a technology subject to investment adjustment costs.

The non-financial firms (henceforth *firms*) produce output using a constant returns to scale

[12] This classification of households as workers and bankers is necessary to quantify distributional welfare effects of change in leverage regulation. It is not required for the analysis of bank leverage or its relationship with banking industry dynamics.

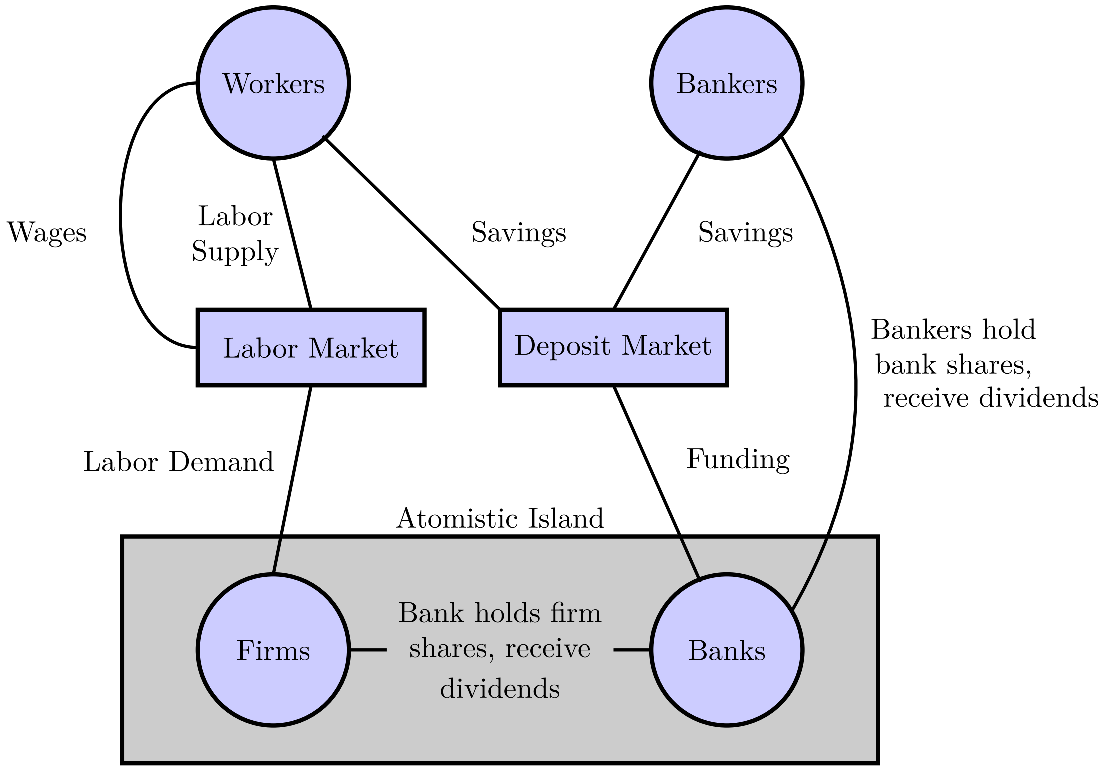


Figure 4: Physical setup of the economy

(CRS) technology that takes labor and capital as input. The capital stock on a given island is owned by the firm. A firm's share on a given island is held by the financial firm belonging to that island. Firms issue new shares to financial firms to purchase additional capital from capital firms. A firm's capital stock is subject to island-specific idiosyncratic *capital quality shocks*. Since financial firms own shares in the firms, these shocks essentially affect the balance sheet of financial firms, and will be interpreted as financial firm *efficiency* in this paper. The financial firm is essentially a financial intermediary that accepts deposits from households and uses them to purchase shares in firms. For the remainder of this paper, financial firms will be referred to as banks. The overview of the economy is presented in Figure 4, following which I discuss the model in greater detail.

3.1 Households

The household sector consists of a fraction λ of bankers and a fraction $1 - \lambda$ of workers. The total mass of households is one. For both types of households, I assume a constant relative risk aversion (CRRA) period utility function. d denotes deposits placed in the bank and R is the gross interest rate on deposits. To fund the deposit insurance system that ensures risk-free deposits, the government imposes income tax τ on the bankers. T denotes lump sum transfers from the government to the workers. I assume that the only way households can save or invest in firms is via the deposit market. This can be motivated in numerous ways, one being that households do not have the expertise to monitor borrowers, which banks have. The problems of workers and bankers are given below, where the subscript b indicates banker and the subscript w indicates worker.

Workers Each atomistic worker supplies a unit of labor inelastically, and saves in the deposit market. Workers maximize the present discounted value of lifetime utility, while choosing levels of consumption and deposits. Workers solve the following problem subject to the budget constraint. All the workers are identical, and solve the same problem. W denotes wages. Π is the dividends paid by capital firms, which are assumed to be owned by the workers.[13]

$$\max_{c_{wt},d_{wt}} \quad E_t \sum_{\tau=0}^{\infty} \beta^{\tau} \left(\frac{c_{wt+\tau}^{1-\gamma}}{1-\gamma} \right)$$

$$c_{wt} + d_{wt} = R_{t-1} d_{wt-1} + W_t + \Pi_t + T_t$$

Bankers Each atomistic banker holds shares in the banking sector, and receives stochastic dividends e_t from banks. They save through the deposit market, and are subject to an *ad hoc* borrowing constraint $\underline{d}$. Bankers jointly provide startup funding to new banks, and also incur the entry cost associated with new banks entering the banking industry. Therefore $M_t(n^e + c^e)$ is the total startup expense incurred by bankers, where M_t is the mass of entrants, n^e is the startup funding, and c^e is the entry cost. When a bank fails, it is replaced by a new bank, and the shareholders of the failed bank become the shareholders of the new bank. Bankers solve the following problem subject to the budget constraint:

$$\max_{c_{bt},d_{bt}\geq \underline{d}} \quad E_t \sum_{\tau=0}^{\infty} \beta^{\tau} \left(\frac{c_{bt+\tau}^{1-\gamma}}{1-\gamma} \right)$$

$$c_{bt} + d_{bt} = d_{bt-1} + (1-\tau)\big((R_{t-1}-1)d_{bt-1} + e_{bt}\big) - M_t(n^e + c^e)$$

Bankers are subject to idiosyncratic income shocks, which renders their state space two dimensional. First dimension of the state space is current dividend income, which is stochastic and follows a Markov process that is induced by the transition dynamics of the bank that is paying the

[13]This is without loss of generality. Capital firms could also have been owned by bankers with no qualitative change in the results of this paper.

dividend. The second dimension of the state space is the amount of savings (deposits) a banker has. In the long-run equilibrium of the economy with no aggregate uncertainty, this problem can be cast recursively, where the state space is given by (e, d). $U(e, d)$ is the lifetime utility of a banker with current dividend income e and deposits d.

$$U(e,d) = \max_{c,d' \geq \underline{d}} \quad u(c) + \beta \sum_{e'} U(e', d')\pi(e, e')$$

$$c + d' = d + (1-\tau)\big((R-1)d + e\big) - M(n^e + c^e)$$

The consumption and savings policy functions obtained above induce a distribution of bankers over the state space, which I will denote by $\mu_b(e, d)$.

3.2 Firms

Each island has a firm, and the firms are identical across islands. They operate a Cobb-Douglas production technology. The firm on island j produces final goods y_t^j by hiring labor l_t^j in the national market at wage rate W_t, and by using the capital stock k_t^j that it owns. The firms solves the following problem, where A_t is some aggregate productivity parameter:

$$\max_{l_t^j} \quad A_t(k_t^j)^\alpha (l_t^j)^{1-\alpha} - W_t l_t^j$$

Firm shares are held by banks, and each share represents claim to one unit of capital. The firm issues new shares to the bank on its own island to purchase additional capital. I assume that there are no financial frictions in the contracting between firms and banks, and that firms pledge all returns to their shareholding banks. Consequently, the price of capital goods is equal to the price of shares in the firm. Dividends per unit of firm capital, which is also equal to dividends per share paid to the bank on island j, is given below:

$$Z_t = \frac{y_t^j - W_t l_t^j}{k_t^j}$$

Note that dividend rate Z_t is equalized across the island because labor is hired in the national market at a national wage rate W_t. Next, capital stock on a given island is subject to idiosyncratic capital quality shocks, and evolves as follows:

$$k_{t+1}^j = (i_t^j + (1-\delta)k_t^j)\psi_{t+1}^j$$

where i_{t-1}^j is investment in period $t-1$ and ψ_t^j is the *capital quality* shock. The shock effectively translates into the return on financial assets (shares) a bank holds. In this sense, capital quality shock can be interpreted as the efficiency of a bank in ascertaining the quality of the capital whose claim it holds. In other words, ψ_t^j is a bank's efficiency in managing its equity in the firm. The two interpretations, that ψ is a capital quality shock and that ψ is bank efficiency, are equivalent because banks are owners of the firm, and shares of the firm are one-to-one backed by the physical capital stock of the firm. In particular, if s_t^j is the bank shareholding in the firms on island j at the *end* of period t, then it is related to the beginning of a period physical capital stock k_t^j as follows:

$$s_t^j = (i_t^j + (1-\delta)k_t^j)$$

where i_t^j is investment and δ is the physical depreciation rate. The above equation implies that shareholding at the *end* of a period is equal to the depreciated capital stock (depreciated because current-period production takes place in the middle of the period) plus new investment undertaken in the current period. The above equation also implies the following simple relationship between capital stock and bank shareholding in firms (bank assets):

$$k_{t+1}^j = \psi_{t+1}^j s_t^j$$

In this sense, s_t^j is a determinant of the expected capital stock at the beginning of period $t+1$, which is $E_t[\psi_{t+1}^j|\psi_t^j]s_t^j$.

So far I have referred to firms by their island indices. However, since (a) firms are identical in their production technology, (b) they operate a CRS Cobb-Douglas production technology, and (c) hire labor in a national market, I can refer to the economy-wide *representative* firm. In other words, I can represent aggregate output Y_t as a function of aggregate capital K_t, and not necessarily as a function of the distribution of capital stocks across islands. This also means that I can refer to the representative firm as one that issues equity in the national market and pays a national dividend rate Z_t. This interpretation makes computations easier, and is employed in the rest of this paper to render the analysis more tractable.

3.3 Capital Firm

The capital firm produces capital goods from consumption goods, but is subject to convex investment adjustment costs. It maximizes the present discounted value of profits accruing to the

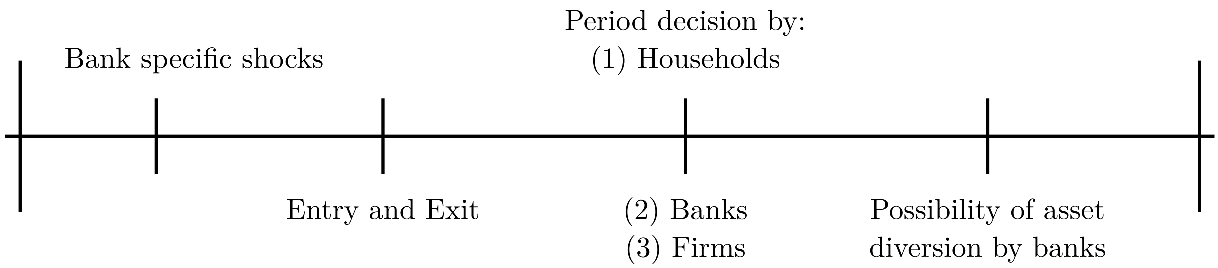


Figure 5: Timeline

household:

$$\max_{I_t} \quad E_t \sum_{\tau=t}^{\infty} \Lambda_{t,\tau} \Big(Q_t I_\tau - [1 + f(I_\tau / I_{\tau-1})] I_\tau \Big)$$

where $f(x)$ is a convex cost function and $\Lambda_{t,\tau}$ is the stochastic discount factor between time t and τ. Q_t is the price of capital goods.

3.4 Banks

There is a continuum of banks. The banks are owned by the bankers. The objective of each atomistic bank j at time t is to maximize the present discounted value of its dividends e_τ^j it pays to its shareholders:

$$E_t \sum_{\tau=t}^{\infty} \beta^\tau e_{t+\tau}^j$$

The discount factor employed by banks in their objective is assumed to be β instead of the stochastic discount factor of the bankers that hold shares in bank j. This is a simplifying assumption to avoid having to keep track of the distribution of bank shareholding, and the corresponding shareholder's inter-temporal marginal rate of substitution. This assumption reflects the fact that the bank uses an *average* discount factor in the economy instead of using the specific discount factor of its shareholders, which boils down to an assumption about the presence of friction in the association between bank management and bank ownership.

At the beginning of any period t, the bank has equity s_{t-1}^j in the firm (bank assets), and deposits d_{t-1}^j from households (bank liabilities). Since the bank owns the firms, and firm capital is subject to idiosyncratic shocks, at the beginning of every period, bank assets are subject to idiosyncratic shocks. As discussed earlier, this shock ψ_t^j is interpreted as a banks's idiosyncratic efficiency in managing its ownership in the firm. This shock captures the notion that a bank that is more *efficient* today makes better asset investment decisions, and hence has a better expected future

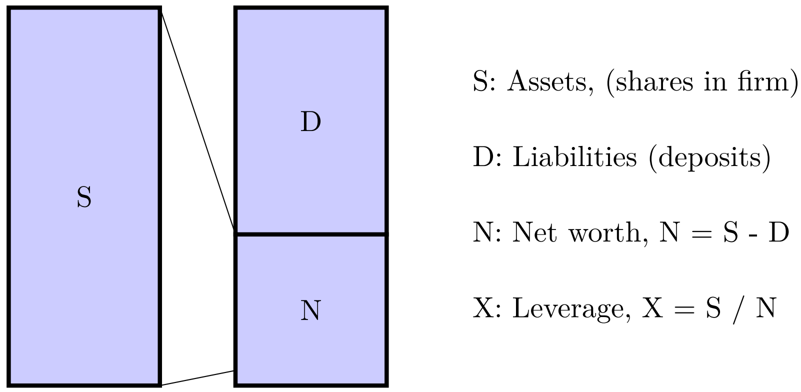


Figure 6: Stylized Bank Balance Sheet

return on its assets. I assume that ψ_t^j follows a first-order Markov process with transition function $F(\psi_t^j, \psi_{t+1}^j)$, which is the conditional distribution of bank efficiency in the next period given bank efficiency in the current period. Later, I will impose more structure on F such that it captures the notion of first-order stochastic dominance (FOSD) in current efficiency: banks with high efficiency today enjoy *better* distribution of efficiencies tomorrow.

Bank failures After banks acknowledge efficiency shocks at the beginning of the period, a bank is considered to have failed if its *beginning-of-period liabilities* are at least as large as its *beginning-of-period assets:*[14]

$$\psi_{t+1}^j Q_t s_t^j - R_t d_t^j \leq 0$$

The assets of the failed bank are liquidated and used to service as much of the total amount of liabilities (depositors) as possible. The remaining liabilities are paid off by the deposit insurance system which is funded by taxes on bankers. Once the exiting banks exit, a mass of new banks pay a fixed entry cost c^e and enter the market with start-up capital n^e. Upon entry, the new banks draw random idiosyncratic efficiencies from a stationary distribution $G(.)$, and begin operations as any other incumbent would.

[14] Note that beginning-of-period assets are valued at the asset prices of the previous period. The interpretation assumes that such a capital quality shock is realized during the transition from one period to another, and that depositors will make a run on the bank if the bank is insolvent as per the valuation of the bank at the very beginning of the period. Valuation at the beginning of the period is based on previous-period asset prices because current-period asset prices are determined later in the period, only after depositors have ensured that insolvent banks have exited. This intra-period liquidity issue can be motivated by bank creditors calling on the bank (or not willing to roll over the debt) before paying off assets, an issue that was key during the 2008 financial crisis.

Bank balance sheet Following bank exit and entry, firms produce output while demanding labor from households. Firms pay dividends Z_t per unit of equity to banks. The banks decide assets s_t^j and liabilities d_t^j for the current period, given dividends Z_t paid by firms, and taking market prices Q_t and R_{t-1} as given. Banks also pay an operating cost $c(N_t^j)$ as a function of their *end-of-period net worth* N_t^j, which for the rest of this paper is the measure of bank size. Banks are subject to the following balance sheet and flow of fund constraints:

$$N_t^j + d_t^j = Q_t s_t^j + e_t^j + c(N_t^j)$$

$$N_t^j = (Z_t + (1-\delta)Q_t)\psi_t^j s_{t-1}^j - R_{t-1} d_{t-1}^j$$

The first equation is the balance sheet identity, which states that the sum of a bank's net worth and liabilities are used in three ways: (a) to purchase assets (b) to pay dividends and (c) to pay operating costs. The second equation is the law of motion for N_t^j. The first term consists of dividend gain component Z_t and capital gain component $(1-\delta)Q_t$, adjusted by efficiency shock ψ_t^j, per unit of asset holding s_{t-1}^j in the previous period. The second term consists of the cost of liabilities.

Operating Costs That banks are subject to a size-dependent convex operating costs is critical for showing that (a) a solution exists for the bank's problem, and that (b) an endogenously determined size distribution exists. I motivate operating costs as increasingly higher resource costs that a bank has to incur in order to operate at larger scales. Expenses that form part of this cost include lobbying, advertising and brand-building efforts, in addition to regular infrastructure and salary expenses. It is in the former sense that in my model, operating costs are only partially observed in bank balance sheet or cash flow statements. An alternative motivation of convex operating costs comes from the firm dynamics literature, where the 'span of control' argument is used to represent the increasingly limited ability of a firm (manager) to operate at larger scales. For financial firms (which I call banks in my model), this is intuitive because larger banks operate in more diversified and complicated financial markets.

However we interpret convex operating costs, they imply a decreasing returns-to-scale (DRS) technology for banks, which is key to the existence of a bank size distribution.[15] Otherwise, with

[15] Operating costs are one way to introduce DRS to a bank's technology. I considered dividend adjustment costs, exogenous random destruction of banks, or operating costs that are convex in a bank's assets as alternatives to introduce the necessary curvature to the bank's problem. However, because of analytic convenience, and because I

CRS technology the size distribution would remain indeterminate. A DRS technology is also useful in proving the existence of a unique solution to the bank's problem.[16] I assume the following operating cost function $c(n) = cn^{\zeta}$, where c is the level and ζ is the exponent parameter of the cost function, which are to be estimated from the data.

Credit Friction In order to capture the key inefficiency of the banking sector, I introduce the following limited commitment issue between banks and their creditors. In particular, I assume that banks can divert a fraction of their assets *a'la* Gertler and Kiyotaki [2010]. Some other approaches to introducing financial inefficiency in a macroeconomic model are unobservable firm effort, as in Christiano and Ikeda [2013], and the moral hazard issue due to deposit insurance, as in Nguyen [2014], Begenau [2014]. My choice in this regard is guided by the focus of this paper on the relationship between leverage and inefficiency. The possibility of asset diversion induces a credit constraint on banks:

$$V_t^j \geq \theta Q_t s_t^j$$

which means that the value to a banker of diverting assets must not be greater than the end of period t charter (continuation) value of the bank. This constraint is a manifestation of the fact that creditors, depositors in this case, are aware of the asset diversion possibility, and therefore lend only up to the point at which banks do not have the incentive to divert assets for personal consumption by shareholders.

The endogenously determined borrowing constraint generates an inefficiency due to the fire sale externality issue - a financial friction that was at the heart of the recent financial crisis. To see how this externality functions, assume that a bank j is hit with a large idiosyncratic shock, such that a large fraction of its assets, and consequently its net worth, are eroded. Erosion of net worth can lead to de-leveraging through more than one channel. First, as shown in the examples in Section 2, asset side shocks can increase a bank's leverage, and it might want to reduce leverage to a pre-shock *optimal* level by selling assets. Second, since V_t^j is concave in net worth,[17] if the erosion of net worth is large enough, a bank might hit the credit constraint, and be forced to further de-lever. Finally, an asset price shock can force de-leveraging in a regime with leverage regulation as the

did not want to complicate the trade-off with respect to bank leverage decisions, I chose the size-dependent operating cost approach.

[16] A DRS technology is not necessary but sufficient to show that the endogenous state variable in the dynamic programming setup of the bank problem is bounded, paving way to show existence of a unique solution to the bank's problem.

[17] This will be proved later.

bank must maintain leverage below the regulatory threshold. Along either channel, selling assets increases the supply of capital in the market, thereby depressing the return on capital. Lower return on capital adversely affects the charter value of other banks, thereby tightening their credit constraints, forcing them to de-lever. Overall, the credit constraint, and ensuing fire sales constitute an externality since a bank, while de-leveraging by selling assets, does not internalize the effects of its own asset sales on the balance sheets of other banks.

Leverage I now introduce the notion of leverage since leverage decisions are a key focus in this paper. First, I define *effective net worth* n_t^j as net worth minus operating costs: $n_t^j = N_t^j - c(N_t^j)$. Next, I define dividend ratio h_t^j as the fraction of effective net worth paid out as dividends: $h_t^j = e_t^j / n_t^j$. Finally, I define leverage x_t^j as the ratio of assets to residual net worth, i.e., net worth after paying dividends from effective net worth:

$$Q_t s_t^j := (1 - h_t^j) n_t^j x_t^j$$

$$\text{Balance sheet constraint} \implies d_t^j = (1 - h_t^j)(x_t^j - 1) n_t^j$$

The flow-of-funds constraint, written in terms of effective net worth, is as follows, where $g(\omega) = \omega - c(\omega)$, is the function that gives effective net worth as a function of raw net worth:

$$n_{t+1}^j = g\Big((Z_{t+1} + (1-\delta) Q_{t+1}) \psi_{t+1}^j s_t^j - R_t d_t^j \Big)$$

$$\implies n_{t+1}^j = g\Big(\underbrace{\big(\frac{Z_{t+1} + (1-\delta) Q_{t+1}}{Q_t}\big)}_{\substack{R_{kt+1}\text{: Return on} \\ \text{assets}}} \psi_{t+1}^j (1 - h_t^j) n_t^j x_t^j - R_t (1 - h_t^j)(x_t^j - 1) n_t^j \Big)$$

$$\implies n_{t+1}^j = g\Big((R_{kt+1} \psi_{t+1}^j - R_t) x_t^j + R_t)(1 - h_t^j) n_t^j \Big)$$

Then, the cutoff of ψ_t that determines whether a bank fails in the next period, is a function of current-period leverage, and is given as follows:

$$\frac{R_t d_t^j}{Q_t s_t^j} = \frac{R_t (x_t^j - 1)}{x_t^j} =: \psi^*(x_t^j, R_t)$$

Also, the borrowing constraint on banks is written in terms of leverage as:

$$V_t^j \geq \theta(1-h_t^j)n_t^j x_t^j$$

Leverage Regulation Finally, I introduce leverage regulation to address the inefficiency in the model that arises due to the fire sale issue, and the fact that banks do not internalize the costs of failure they impose on the economy. This is formalized in the optimal regulation discussion. The regulation reduces bank failure probability in a steady state, but it also restricts total volume of financial intermediation, and it is this trade-off that I study in this paper. Leverage regulation is defined as a constraint on the maximum leverage a bank might assume. The dependence of regulation on net worth is a key feature of this paper, and will be discussed in greater detail later.

$$x_t^j \leq \chi(n_t^j)$$

Recursive formulation I cast the bank's problem recursively. The state vector for any bank j at the end of period t is $(\psi_t^j, s_{t-1}^j, d_{t-1}^j; O_t)$, where O_t is the aggregate state vector. In the general model, O_t comprises all market prices, all aggregate shocks, and the distribution of banks. In the baseline model with no aggregate uncertainty, O_t comprises of market prices only. In either version, since the bank takes market prices Q_t, R_{t-1} and dividends Z_t as given, it follows from the flow of funds constraint that n_t^j is a sufficient statistic for s_{t-1}^j and d_{t-1}^j. Consequently, a smaller and equivalent state vector for any bank j is $(\psi_t^j, n_t^j; O_t)$. The choice variables for any bank j are x_t^j and h_t^j. In writing the recursive problem of an arbitrary bank j, I omit the time subscript and bank specific superscript except when necessary. I use a -1 subscript to denote the previous period and a prime superscript to denote the next period. The recursive problem of an incumbent bank is given as follows:

$$V(\psi, n; O) = \max_{x\in[1,\chi(n)],h\in[0,1]} \left(hn + E\Lambda' \int_{\frac{R(x-1)}{x}}^{\overline{\psi}} V(\psi', n'; O')F(\psi, d\psi') \right)$$

$$n' = g\Big(((R_k'\psi' - R)x + R)(1-h)n \Big)$$

$$V(\psi, n; O) \geq \theta(1-h)nx$$

The problem indicates that current period value of a bank is the sum of its current period dividend payment and its discounted continuation value. The integral captures the fact that if next period realization of shock ψ' is less than $R(x-1)/x$, then the bank's *beginning* of period net worth is non-positive and the bank fails, which means zero continuation value. Net worth evolves as per a technology $g(.)$ banks posses, which is concave because $g(\omega) = \omega - c(\omega)$ and c(.) is convex. The incentive constraint is based on the current period value function, and imposes an endogenously determined constraint on the choice set. The choice set is bounded in the following sense. $[0, 1]$ is the bounded choice set for h because limited liability[18] $\implies h \geq 0$ and non-negative net worth $\implies h \leq 1$. Leverage, x, is bounded below by 1 since deposits are non-negative, and it is bounded above by $\mathcal{X}(n)$, which is the leverage regulation that restricts banks from taking on too much leverage. For the benchmark model that follows and its calibration, I consider a degenerate regulation $\mathcal{X}(n) = \hat{\mathcal{X}}$, as is the case in the data for the chosen calibration period. I will now discuss some properties of the bank's problem.

Proposition 1. *Given current bank efficiency ψ, bank failure probability is strictly increasing in leverage x, where the bank solvency probability for a bank characterized by (ψ, n) is given as*

$$p(\psi, n) = \int_{R(x-1)/x}^{\overline{\psi}} F(\psi, \partial\psi')$$

Proof. The interval of integration is the bank's endogenously determined region of solvency. The lower limit of this integral is the failure cutoff $\psi^*(x, R) = R(x-1)/x$ i.e. the bank fails if it realizes $\psi' < \psi^*(x)$. Then the proposition follows directly from the fact that $\partial\psi^*(x)/\partial(x) = R/x^2 > 0$. Note, however, that higher leverage does not necessarily correspond to a higher probability of failure, which leads me to the next proposition. ■

Proposition 2. *Assume that $F(\psi, \psi')$ is first-order stochastically increasing in ψ. Then, given a bank with current leverage x, the probability of failure is weakly decreasing in current efficiency ψ.*

Proof. This proposition follows directly from the definition of FOSD. If $\psi_1 > \psi_2$, then $F(\psi_1, \psi') \leq F(\psi_2, \psi') \; \forall \; \psi'$ and hence

$$p(\psi_1, n) = \int_{R(x-1)/x}^{\overline{\psi}} F(\psi_1, \partial\psi') \leq \int_{R(x-1)/x}^{\overline{\psi}} F(\psi_2, \partial\psi') = p(\psi_2, n)$$

Intuitively, when a bank's efficiency increases *ceteris paribus*, the distribution of next period efficiency is shifted to the right, and hence reducing the probability of a bad *(where bad depends on the*

[18] No borrowing from equity holders

leverage) shock. In light of these propositions, the trade-off a bank faces with respect to leverage and dividend decisions can be summarized below. ■

1. The trade-off with respect to leverage is as follows. Increasing leverage increases the variance of *total* premium $(R_k'\psi' - R)x$ a bank earns on its effective post dividend net worth. Depending on the distribution of ψ', and as long as expected premium is positive, increasing leverage raises continuation value. However, increasing leverage also reduces the probability of solvency (increases the probability of failure).

2. The trade-off with respect to dividends is as follows. Increasing dividend ratio h increases the current value for bank equity holders, but reduces n'. As I prove in the appendix, V is increasing in net worth, which means that increasing h decreases continuation value by reducing retained earnings.

Entrant bank problem There are infinitely many potential entrants and there is free entry. Entrants pay a cost c^e before they can enter the market. After entry, they draw their idiosyncratic efficiency from an invariant distribution G. All entrants start operations with a fixed start-up capital n^e provided to them by households.[19] Since, after entry, an entrant behaves exactly like an incumbent, the expected value of a potential entrant can be computed using the value function of incumbents as follows:

$$EV^e = \int_{\underline{\psi}}^{\overline{\psi}} V(\psi, n^e) dG(\psi)$$

Therefore, a potential entrant will enter the market if and only if:

$$EV^e - c^e \geq 0$$

But then free entry implies that the net value of entry should not be positive, or else an infinite mass of entrants will enter the market, i.e.:

[19]The assumption that banks enter with fixed start-up capital provided by households is made to keep the model simple. This assumption essentially abstracts the model away from new equity issuance decisions on part of the banks and equity purchase decisions on part of the households. The motivation inducing households to provide additional capital (start-up capital) to entrants and not to incumbents follows from the concavity of the value function in net worth. It is due to this concavity that marginal value of bank capital is greater for smaller banks than for larger banks. Although in this paper I do not solve for n^e as an equilibrium outcome of the model where the household return on providing start-up capital is equalized to household return on deposits, I estimate n^e from the data.

$$EV^e - c^e \leq 0$$

The two conditions above imply that if the mass M of entrants is strictly positive, it must be true that $EV^e = c^e$. And, if the mass of entrants is zero, it must be the case that $EV^e < c^e$. Now I establish some analytical properties of the bank's problem.

3.5 Analytical properties

Proposition 3. *If $R_k > R$, then the Bellman Equation specifying the bank's problem is a Contraction Map, and therefore there exists a unique value function corresponding to the bank's problem.*

I prove that the Blackwell Conditions are verified, and ensure they are sufficient for a contraction map in the set up of my problem. *(Proof in appendix).*

Proposition 4. *Suppose there are no credit frictions. Suppose $(\psi, n) \ni h(\psi, n) > 0$. Then:*

1. $x_1(\psi, n) > 0$
2. $h_1(\psi, n) < 0$
3. $x_2(\psi, n) = 0$
4. $h_2(\psi, n) > 0$

The first proposition states that if a bank is more efficient, it takes on more leverage. This follows from the assumption that $F(\psi, \psi')$ is FOSD in ψ, which means that a more efficient bank enjoys better distribution of ψ', the next period efficiency. In other words, the possibility of being more efficient in the next period allows the bank to bump up expected net worth by taking on more leverage, with relatively less reduction in the probability of solvency.

The second proposition states that more efficient banks find it optimal to retain earnings by paying lower dividends, precisely because they face better prospects in the next period. The third proposition follows directly from the feature of the model in virtue of which the trade-off with respect to leverage does not depend on size. The fourth proposition states that larger banks pay more dividends. This follows from the concavity of the value function in size. For larger banks, the marginal value of retaining their earnings is relatively lower, rendering it optimal for larger banks to pay more dividends in the current period. (*Outline of proof in appendix*).

3.6 Bank distribution

Let $X = [0,1] \times [0, \bar{n}]$ be the state space of the model. Let $\mathcal{A} = [0,1] \times [1, \hat{\chi}]$ be the action space of the model. Then, $(x,h) : X \to \mathcal{Y}$ are the policy functions. With this notation, we define the distribution of banks over the state space. The distribution of banks in the *middle* of period t (i.e., after bank entry and exit has occurred but before period decisions have been made), is given by a measure $\mu : \mathcal{B}(X) \to \Re^{+}$, where $\mathcal{B}$ indicates the set of Borel subsets of X.[20] Abusing notation for μ in the obvious sense, let $\mu_t(\Psi, N)$ denote the time t distribution of banks with idiosyncratic shock $\underline{\psi} \le \psi \le \Psi$ and net worth $0 \le n \le N$. Then the evolution of bank distribution is given by the following expression:

$$\mu_{t+1}(\Psi, N) = M_{t+1} \mathbb{1}(n^e \le N) \int_{\underline{\psi}}^{\Psi} dG(\psi)$$

$$+ \int_0^{\bar{n}} \int_{\underline{\psi}}^{\overline{\psi}} \Bigg(\int_{\underline{\psi}}^{\Psi} \mathbb{1}\Big(g\big(((R_k'\psi' - R)x + R)(1-h)n\big) \le N \Big)$$

$$\mathbb{1}\bigg(\frac{R(x(\psi,n) - 1)}{x(\psi,n)} \le \psi' \bigg) F(\psi, d\psi') \Bigg) d\mu_t(\psi, n)$$

The first term indicates mass M_{t+1} of entrant banks that enter in period $(t+1)$, that have start-up capital less then N and a draw of efficiency less than Ψ. Basically, the first term represents mass of entrants who land up in the following subset of the state space: $[\underline{\psi}, \psi] \times [0, N]$. The second term represents the flow of incumbent banks into this subset of the state space, net of those which fail. Basically, we "count" all the banks from period t that, while following their respective policy functions, did not fail and landed in the following subset of the state space in period $(t+1)$: $[\underline{\psi}, \psi] \times [0, N]$.

The evolution of bank distribution can be written concisely using the notion of the transition function W on state space X. The transition function W is a map from $X \times \mathcal{B}(X) \to \mathcal{R}^{+}$, where $W\big((\psi,n),(\Psi,N)\big)$ denotes the probability of transition from point (ψ, n) in the state space to the following subset of the state space in the next period: $(\Psi, N) = \{(\psi, n) : \psi \le \Psi, n \le N\}$. The transition function and associated distribution evolution equation are as follows:

$$W\big((\psi,n),(\Psi,N)\big) = \int_{\underline{\psi}}^{\Psi} \mathbb{1}\Big(g\big(((R_k'\psi' - R)x + R)(1-h)n\big) \le N \Big)$$

[20]Not necessarily a probability measure

$$\mathbb{1}\left(\frac{R(x(\psi,n)-1)}{x(\psi,n)} \leq \psi'\right)F(\psi,\partial\psi')$$

$$\mu_{t+1}(\Psi,N) = M_{t+1}\mathbb{1}(n^e \leq N)\int_{\underline{\psi}}^{\Psi}\partial G(\psi) + \int_{\chi} W\big((\psi,n),(\Psi,N)\big)\partial\mu_t(\psi,n)$$

The latter can be concisely represented using the *distribution evolution operator* T as $\mu_{t+1}(\Psi,N) = T(\mu_t, M_{t+1}, O_t)$, where O_t is the aggregate state vector comprising equilibrium prices and quantities.

Induced Dividend payments Each banker holds shares in the banking industry. Shareholding is modeled as a continuous bijection from the *exogenously specified* mass of bankers to the *endogenously determined* mass of banks. The dividend income stream of a given banker is determined by the dividend payment stream of the banks whose shares the banker holds. In this sense, the bank dividend policy $h(\psi,n)$ and the bank transition function $W\big((\psi,n),(\Psi,N)\big)$ induce a dividend distribution transition function $\Pi(e,e') = \text{Pr}(\text{current dividend} = \text{e, future dividend} \leq e')$, where $e(\psi,n) = nh(\psi,n)$. This dividend income transition process is what governs the stream of idiosyncratic dividend income realization for bankers, and also governs a banker's expectation regarding future dividend income. The dividend distribution transition function $\Pi(e,e')$ is computed as follows, from which the dividend transition function $\pi(e,e') = \text{Pr}(\text{current dividend} = \text{e, future dividend} = e')$ can be computed as the marginal of the former:[21]

$$X' = \{(\psi,n) \in X \mid nh(\psi,n) = e\}$$

$$X'' = \{(\psi,n) \in X \mid nh(\psi,n) \leq e'\}$$

$$\Pi(e,e') = \int_{X'}\int_{X''} W\big((\psi,n),\partial(\Psi,N)\big)\partial(\psi,n)$$

4 Stationary General Equilibrium

I now focus on long-run stationary general equilibrium (SGE) in the model. Essentially, I consider the case of no aggregate uncertainty, and establish that equilibrium exists where the aggregate state vector and the bank distribution are time invariant. Although the long run equilibrium entails invariant aggregates, there are interesting dynamics at the individual bank and banker levels to be

[21] For the computations, I will discretize the state space for dividend payments, and compute the Markov Transition matrix following the approach laid out.

studied. I begin with the definition of SGE, followed by a discussion of its existence.

Definition A Stationary general equilibrium consists of (i) bank value function $V(\psi, n)$ and (ii) bank policy functions $x(\psi, n), h(\psi, n)$ (iii) invariant bank distribution $\mu(\psi, n)$ of net worth and efficiency shocks (iv) entrant mass M, (v) banker value function $U(e, d)$ and banker policy functions $d^b(e, d)$ and $c^b(e, d)$ (vi) banker distribution $\mu^b(h, d)$ (vii) worker consumption c^w and savings d^w(viii) aggregate consumption C, aggregate capital K, aggregate deposits D, aggregate labor demand L, Taxes T, Dividends H and (ix) wage rate W, interest rate R and dividends Z (these prices form the aggregate state vector O) such that given prices:

1. $V(\psi, n)$, $x(\psi, n)$ and $h(\psi, n)$ solve the bank's problem given prices
2. Free entry condition is satisfied: $EV^e \leq c^e$
3. Labor market clears at wage rate W: $L = 1 - \lambda$
4. Deposit market clears at interest rate R:
$$\int_0^{\bar{n}} \int_{\underline{\psi}}^{\overline{\psi}} d(\psi, n) \partial\mu(\psi, n) = (1 - \lambda) d^w + \lambda \int_{\underline{d}}^{\overline{d}} \int_0^{\bar{e}} d^b(e, d) \partial\mu^b(e, d);$$
5. Asset market clears at price Q:
$$K = \int_0^{\bar{n}} \int_{\underline{\psi}}^{\overline{\psi}} \Big(\int_{\underline{\psi}}^{\overline{\psi}} \psi' F(\psi, d\psi') \Big) s(\psi, n) \partial\mu(\psi, n) = S$$
6. Consumption goods market clears, where effective output is net of entry and operating costs:
$$Y = K^\alpha L^{1-\alpha} = C + I + M_t c^e + c^o;$$
$$c^o = \int_0^{\bar{n}} \int_{\underline{\psi}}^{\overline{\psi}} c(n) \partial\mu(\psi, n);$$
$$C = (1 - \lambda) c^w + \lambda \int_{\underline{d}}^{\overline{d}} \int_0^{\bar{e}} c^b(e, d) \partial\mu^b(h, d);$$
where the interpretation of aggregate investment I in SGE is that it includes 'recovery' of lost capital due to fixed capital depreciation at rate δ and idiosyncratic stochastic capital quality shocks, such that aggregate capital stock remains constant

7. The bank distribution is a fixed point of the distribution evolution operator: $\mu = T(\mu, M)$

8. $c^b(h,d), d^b(h,d)$ solve the banker's problem; c^w, d^w solve the worker's problem

9. Government budget constraint is satisfied: Taxes, net of transfers, equal shortfall in liabilities of failing banks:

$$\tau\Big(\lambda \int_{\underline{d}}^{\overline{d}} \int_0^{\bar{e}} (e + d^b(e,d)(R-1))\partial\mu^b(e,d)\Big) - T =$$

$$-\int_0^{\bar{n}} \int_{\underline{\psi}}^{\overline{\psi}} \Big(\int_{\underline{\psi}}^{\overline{\psi}} min(0, \psi' s(\psi,n) - Rd(\psi,n))F(\psi, d\psi')\Big)\partial\mu(\psi,n)$$

4.1 Existence and solution

I am interested in the SGE with positive entry ($M^* > 0$), the existence of which depends on the fixed parameters of the model, especially the entry cost. To this end, I first solve for equilibrium quantities that can be analytically pinned down. The first-order conditions from the representative banker's problem imply that $R_t = R^* = 1/\beta$ since consumption is a constant C^* in SGE. Similarly, since investment is constant from one period to another, there are no investment adjustment costs, and the capital firm's problem implies that $Q_t = Q^* = 1$ and $\Pi_t = \Pi^* = 0$. Next, deriving the equation from the firm's problem, which takes stock of capital and prices as given, aggregate labor demand L^* is given by:

$$L^* = \left(A\frac{(1-\alpha)}{W^*}\right)^{1/\alpha} K^* \tag{4.1.1}$$

Dividends are then given by the following relationship:

$$Z^* = \frac{Y^* - W^*L^*}{K^*} = \frac{\alpha Y^*}{K^*} = \alpha A(K^*)^{\alpha-1}(L^*)^{1-\alpha} \implies K^* = (\alpha A/Z)^{1/1-\alpha}L^* \tag{4.1.2}$$

Since labor supply equals $1-\lambda$, I can use aggregate labor demand to pin down wages as a function of dividends Z^*:

$$W^* = (1-\alpha)A^{1/1-\alpha}\left(\frac{\alpha}{Z^*}\right)^{\alpha/1-\alpha}$$

I now turn to the bank's problem. First I note that if Z, Q, R are known, then the bank's problem can be solved uniquely. Since R, Q are already pinned down, Z is the only equilibrium quantity to be solved for. Using the free entry condition, I can pin down Z as shown in the following proposition.

Proposition 5. *Given an entry cost $c^e > n^e$, equilibrium firm dividend rate Z^* is uniquely deter-*

mined.

Proof. Suppose Z is known, then R_k is also known, and the bank's problem can be solved. From a routine application of the envelope theorem to the Bellman equation for the bank's problem, and since the value function is increasing in net worth, I learn that the value function is strictly increasing in Z. I can then define:

$$g(Z) = \int_{\underline{\psi}}^{\overline{\psi}} V(\psi, n^e; Z) dG(\psi)$$

which is the expected value of entry as a function of dividend return Z on bank assets, where $g'(z) > 0$. Now if $Z \to 0 \implies V(\psi, n^e; Z) \to n^e \implies g(Z) \to n^e$, and if $Z \to \infty \implies g(Z) \to \infty$. Then, by the intermediate value theorem (IVT), I know that:

$$\exists\ Z^* \ni g(Z^*) = c^e$$

■

Once Z^* is solved for a given set of parameters of the model, I can solve the bank's problem uniquely, and also pin down w^* using Equation 4.1.2. Then, given an M, and given market prices, I can solve for the fixed point μ^* of T, eventually pinning down all other quantities. Essentially, an equilibrium can be uniquely described by (Z^*, M^*, μ^*). Based on these insights, the following solution algorithm is proposed:

1. Find Z^* using free entry:
$$\int_{\underline{\psi}}^{\overline{\psi}} V(\psi, n^e; Z^*) dG(\psi) = c^e;$$

2. Obtain the bank's value and policy functions, and W^*

3. Find that $L^* = 1 - \lambda$ from the labor market clearing condition, and obtain K^* from 4.1.1; also obtain output Y^* from production function

4. Assume M, and obtain the steady state bank distribution μ corresponding to M by iterating on T

5. Pin down M^* using the asset market clearing condition; Obtain μ^*

6. Using the bank distribution, obtain aggregates D, T

7. Back out the transition matrix for dividends

8. Solve the banker's income fluctuation problem, and the worker's problem

The idea for steps 5 is as follows. Given a mass of entrants M, the bank distribution implies aggregate assets. If there are too many entrants in the equilibrium, then the aggregate assets are higher relative to K^*. If there are too few entrants, the opposite is true. Based on this argument, I conjecture that the asset market can clear for some $M > 0$. Let $K^* - S'(M) = K^e(M)$ be the excess asset supply curve as a function of the mass of entrants M. Then a positive entry SGE exists if:

$$\exists\ M^* > 0 \ni K^e(M^*) = 0$$

The usual strategy for showing this is to use the intermediate value theorem (IVT) on the excess assets supply function. I will show this computationally.[22] Before I proceed to the data and model estimation section, the following proposition is useful for efficient computation in steps 4 and 5.

Proposition 6. *The distribution evolution operator T is linearly homogeneous in (μ, M) i.e.:*

$$\mu(M) = T(\mu(M), M, \theta) \implies M'\mu(M)/M = T(M'\mu(M)/M, M', \theta)$$

Proof. Let $\mu(M)$ be the stationary distribution corresponding to M. Then,

$$\mu(M)(\Psi, N) = M\mathbb{1}(n^e \leq N) \int_{\underline{\psi}}^{\Psi} dG(\psi) + \int_{\chi} W((\psi, n), (\Psi, N)) d\mu(M)(\psi, n)$$

Multiplying both sides by M'/M gives the following, where $(M'\mu(M)/M)$ is the transformed measure.

$$(M'\mu(M)/M)(\Psi, N) = M'\mathbb{1}(n^e \leq N) \int_{\underline{\psi}}^{\Psi} dG(\psi) +$$

$$\int_{\chi} W((\psi, n), (\Psi, N)) d(M'\mu(M)/M)(\psi, n)$$

■

The key to this linear homogeneity is that since $\mu(M)$ is a measure (not a probability measure), multiplication by a scalar simply scales the measure. Intuitively, when more entrants enter the

[22]In order for a stationary general equilibrium to exist where markets clear for some prices, the bank distribution must be continuous in the parameters of the model. This is because if the distribution does not move continuously, there might be 'holes' in the set of aggregate deposits and aggregate consumption obtained by integrating over the bank distribution, causing failure of market-clearing conditions. I do not prove this analytically, but show computationally that equilibrium exists.

market in equilibrium, each subset of the state space experiences a proportional increase in its equilibrium measure. Another way of looking at this is that when entry mass is higher in stationary equilibrium, the mass of incumbents must rationalize it - the mass of incumbents must be large enough so that it results in a higher failure mass that matches the higher entry mass. Note that *mass* is key here, since failure rate does not change merely due to a greater mass of entrants.

5 Data and Estimation

To estimate my model, I use data on US commercial and savings banks from (a) FDIC Call and Thrift Financial (CTR) Reports (b) the FDIC failed banks list and (c) FDIC reports of changes in the institution count. I source balance sheet, cash flow and income statement variables from CTR. I use the failed bank list to identify failed banks, and changes in the bank count list to compute failure and entry rates. Consistent time series estimates for individual banks are ensured by following the notes developed by Kashyap and Stein [2000].[23] Macroeconomic aggregates for the US are sourced from FRED and the World Bank Indicators database. I now discuss the functional forms and parameter estimation.

The operating cost function is chosen to be convex, so that it renders the bank's technology of converting assets today into assets tomorrow into a decreasing returns to scale technology. The functional form for operating costs is cn^{ζ}. Suppose $\bar{n} > 0$ is a scaling factor that represents maximum effective net worth on the grid used to solve the model. This factor is simply used to map the scale of the model to scale of the data. The parameters of the cost function are chosen such that effective net worth $n - c(n)$ is not greater than $\bar{n}$. This implies that $c = \bar{n}^{(1-\zeta)/\zeta}$. The investment adjustment cost is modeled as a quadratic cost $f(x) = (x-1)^2$. The process for achieving idiosyncratic efficiency is modeled as an AR(1) process with mean $\mu = 1$, persistence ρ, and variance σ:

$$\psi_{it} = \mu(1-\rho) + \rho\psi_{it-1} + \sigma\epsilon_{it}$$

Finally, the distribution of efficiency for entrants, $g(.)$ is the ergodic distribution of ψ. One period in my model corresponds to an year in the data. The values of some standard parameters are chosen based on the standard in literature. The discount factor β is set such that the interest

[23] Consistency of data is an issue with this data because of changes in variable definitions as well as bank mergers or disintegrations.

Type of financial asset	1995	1998	2001	2004	2007
Transaction accounts	87	90.5	91.4	91.3	92.1
Certificates of deposit	14.3	15.3	15.7	12.7	16.1
Savings bonds	22.8	19.3	16.7	17.6	14.9
Bonds	3.1	3	3	1.8	1.6
Stocks	15.2	19.2	21.3	20.7	17.9
Pooled Investment funds	12.3	16.5	17.7	15	11.4
Retirement accounts	45.2	48.8	52.2	49.7	52.6
Cash value life insurance	32	29.6	28	24.2	23
Other managed assets	3.9	5.9	6.6	7.3	5.8
Other	11.1	9.4	9.4	10	9.3
Stocks (Direct plus Indirect)	40.4	48.9	52.2	50.2	51.1
Any Financial Asset	91	92.9	93.4	93.8	93.9

Table 2: Financial asset ownership by families in the US, 1995 - 2007 *(Source: Survey of Consumer Finances)*

rate $R = 1/\beta$ in the model matches an annual risk free interest rate close to 4%. The Cobb-Douglas production function parameter α is set such that capital income share in the model implies a capital income share that is close to 0.33. For depreciation rate, a value of 0.025 is used, which is again common in the macroeconomic literature. Total factor productivity A is normalized to 1.

The fraction of bankers λ is set to match the pooled mean percentage of households in the US during 1995 - 2005 who held stocks, directly or indirectly, through investment funds or retirement accounts. This estimate is based on data from Survey of Consumer Finances 1995-2005. A break up of asset ownership by households, and stocks in particular, is given in Table 2. The above calibration is rationalized by the fact that bankers in my model are essentially the owners of capital stock, as bankers own banks, banks own firms, and firms own capital.

The benchmark leverage regulation $\hat{\chi}$ is calibrated based on the statutory risk-weighted capital regulation as there was no leverage regulation imposed on banks during 1995-2005. Then, as per Basel I, banks were required to have an amount of capital that is at least 8% of risk-weighted assets. I compute an approximate implied equivalent leverage regulation using the fixed statutory risk weights and actual capital ratios of banks. Specifically, since I do not have data on bank securities holdings, I first obtain a range of possible values for the implied leverage regulation parameter $\hat{\chi}$. To this end, I estimate the minimum and maximum capital charge for assets of each bank in the data during 1995-2005. Then the mean ratio of total assets to these capital charges gives the maximum and minimum implied leverage regulation that is equivalent to the statutory

Description	Symbol	Value
Standard parameters		
Coefficient of risk aversion	γ	1.5
Discount factor	β	0.96
Capital exponent in production	α	0.33
Capital depreciation rate	δ	0.025
Benchmark Leverage regulation	$\hat{\chi}$	29.58
Fraction of bankers	λ	0.475
Joint estimation using SMM		
Efficiency Persistence	ρ	0.8373
Efficiency Variance	σ	0.0072
Cost exponent	ζ	1.6853
Fixed cost of entry	c^e	3.4060
Asset diversion fraction	θ	0.1796
Entrant bank start up capital	n^e	0.2791
Income tax rate	τ	0.2416

Table 3: Summary of parameter values.

capital regulation during 1995-2005.

$$\hat{\chi}_{max} = 29.58; \quad \hat{\chi}_{min} = 27.37$$

These estimates are consistent with the observation that more than 99 percent of banks have a leverage of less than 25.30. I choose the more liberal estimate of $\hat{\chi}$ for the benchmark calibration.

The remaining parameters $\mathcal{P} = (\rho, \sigma, \zeta, c^e, \theta, n^e, \tau)$ are estimated jointly using the method of moments. As per this method, I begin with a guess of these parameters, and compute the stationary general equilibrium in the model. From SGE, I compute the target moments listed in Table 4. The vector of these model moments, say $\mathcal{M}^{model}(\mathcal{P})$, is a function of the parameters. The idea is to find the set of parameters $\mathcal{P}^*$ such that:

$$\mathcal{P}^* = arg \min_{\mathcal{P}} (\mathcal{M}^{model}(\mathcal{P}) - \mathcal{M}^{data})' W (\mathcal{M}^{model}(\mathcal{P}) - \mathcal{M}^{data})$$

where $\mathcal{M}^{data}$ is the vector of target data moments, and W is the weighting matrix. I use a pre-specified weighting matrix whose diagonal elements are the inverse of squared data moments. This is essentially equivalent to minimizing the sum of the squared relative difference between model and data moments. This weighting matrix gives me consistent parameter estimates,[24] which are

[24] One could do another step to back out the efficient weighting matrix using the co-variance matrix of initial parameter estimates. Using the efficient weighting matrix would give me the lowest possible standard error on my parameter estimates.

Moments	Data	Model
Failure rate	.0026	.0022
Ratio of size of entrant and incumbent	.2354	.2374
Correlation of Size and Failure rate	-.4238	-.3669
Leverage	10.38	12.28
Fraction of banks in quintile 1	.2000	.1962
Fraction of banks in quintile 2	.2000	.2395
Fraction of banks in quintile 3	.2000	.1754
Fraction of banks in quintile 4	.2000	.2146
Dividend Ratio	.0939	.0963
Income Gini	.3641	.3366

Table 4: 1995 - 2005 Data moments vs Model moments. For data moments (except for the quintiles and Gini index), standard errors are $[0.0004, 0.0414, 0.0915, 0.0247, 0.0009]$ respectively. Quintiles are based on size distribution in bottom 95% banks as of 2005:Q4.

summarized in Table 3. Target moments and estimation results are summarized in Table 4. The choice of these moments is such that key features of banking industry dynamics are captured, namely moments of the size distribution, leverage and dividend behavior, exit rate, size of entrants and relative failure rate of banks. The income tax parameter τ, targets the mean income GINI of US households during 1995-2005.

5.1 Computations

In this sub-section, I discuss the computational procedure I use to compute the stationary equilibrium, including the estimation of parameters using the Method of Moments. I use value function iteration with Howard policy improvement to solve for the value and policy functions of the bank.[25] The state space for banks is discretized as follows: (a) ψ has 21 equally spaced grid points in $[\underline{\psi}, \overline{\psi}]$; (b) n has 31 log-spaced grid points in $[0, \bar{n}]$. Spline interpolation is used to evaluate value function at off-grid points. Linear interpolation is used to evaluate policy function at off-grid points. A two-times-finer discretization of the state space is used to compute the stationary distribution for banks. On this finer grid, I construct the state space transition matrix using bank policy functions, and compute its ergodic distribution, which is the stationary distribution of banks. To handle entry and exit in this setting, I introduce a *dump* state which represent failures. Since in the stationary equilibrium, entry equals exit, transitions to the dump state represents failures, whereas transitions from the dump state represent entries. The ergodic measure of dump state is therefore interpreted as the failure rate. To solve the banker's income fluctuation problem, policy function iteration is

[25] I follow Heer and Maussner [2009] as the key reference for computational methods.

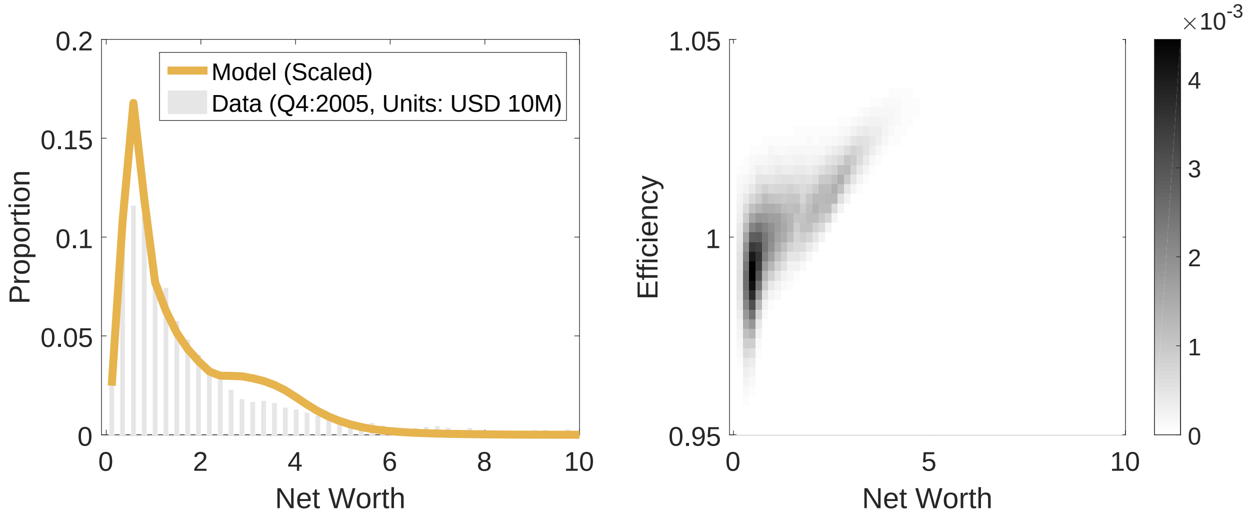


Figure 7: (Left) Distribution of banks over net worth in data and in model (Right) Distribution of banks over net worth and efficiency *(Distribution over state-space)*

employed. Stationary distribution is computed using a state transition matrix as before.

The Nelder-Mead simplex optimization algorithm is used to compute the Method of Moments parameter estimates.[26] Table 4 documents the result of this procedure. I target nine data moments based on pooled data from 1995 - 2005 to estimate six model parameters. The model does reasonably well at matching moments, except for mean leverage. One reason for not doing well in matching this target is that the actual implied leverage regulation is unknown. I obtain an approximation using the statutory capital regulation, which generates the possibility of a tighter leverage regulation than the one I use here.

Figure 7 compares the size distribution in the data with the one generated by the model. As also reflected in Table 4, the model captures the observed bank distribution reasonably well, except for the extreme right tail. The figure also shows that larger banks are more efficient in equilibrium. Intuitively, this follows from the dynamics of the industry: banks aim to grow larger by taking on more leverage, but only those that receive a better stream of efficiency shocks actually grow. In this sense, there is a selection effect which is operational. Bank value and policy functions are presented in Figure **??** in the appendix.

[26]Routine available in MATLAB in the function 'fminsearch'.

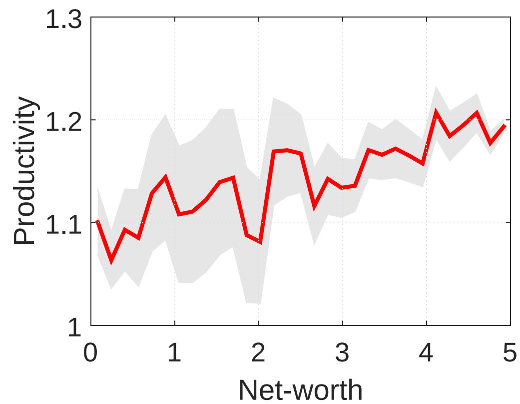


Figure 8: Equilibrium profile of mean bank productivity. The shaded band represents one standard deviation.

Bank productivity I define bank productivity as the rate at which a bank converts current period net-worth into next period net-worth, which is mathematically given as follows:

$$q(\psi, n) = E[n']/n = E\Big[g\Big(((R_k'\psi' - R)x(\psi, n) + R)(1 - h(\psi, n))n\Big)\Big]/n$$

The rationale for this measure is two-fold. First, I would like to combine the notion of technical efficiency of a bank (operating costs that increase with bank size), and asset management efficiency ψ, which is equilibrium, is higher for larger banks. Second, I would like to verify if my model is robust to the *economies of scale in banking* that several empirical papers have found, for example, Hughes and Mester [2013]. To this end, I compute the equilibrium profile of bank productivity in my model, which I graph in Figure 8. The shaded band represents one standard deviation. The model prediction that bank productivity is increasing in size of the bank is therefore consistent with empirical findings.

6 Counterfactual Experiment

In this section, I discuss the counterfactual experiments I conduct in my model. This exercise has two goals. The first goal is to understand the response of banking industry and macroeconomic aggregates to changes in bank leverage regulation. The second goal is to understand the trade-offs a planner faces while choosing leverage regulation to maximize welfare. This allows me to pin down the channels of welfare gains and losses that occur as the regulation regime changes.

The benchmark economy equilibrium features a size-independent leverage regulation regime

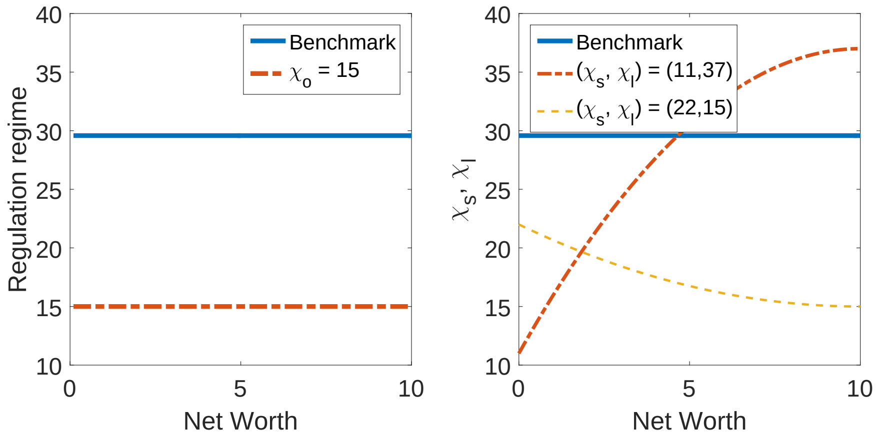


Figure 9: (i) *(Left)* Size independent regulation (ii) *(Right)* Size dependent regulation

where $\chi(n) = \hat{\chi} = 29.58$ i.e. all banks face the same limit $\hat{\chi}$ on the level of leverage they can assume. For the counterfactual experiment, I first define a regulation function as any continuous, differentiable and monotonic function of bank size $\chi(n)$ that satisfies the following conditions $\chi(n) \geq 1, \lim_{n\to\infty} \chi'(n) = 0$. I consider only monotonic regulation functions to rule out complicated policies. Moreover, optimization over the class of all continuous and differentiable functions is a more difficult problem to handle computationally. The other two conditions imposed on a valid regulation function are intuitive. First, a leverage regulation below 1 is absurd as it follows directly from the balance sheet identity that leverage cannot be less than 1. Second, the condition that regulation becomes size independent in the limit is needed to ensure that regulation does not blow to $+\infty$ or $-\infty$. This condition is not restrictive in the sense that the size independence of regulation kicks in for banks of very large size only.

As such, the class of regulation functions proposed above is large. For tractability, I consider the class of quadratic regulation functions, which are completely described by three parameters. The choice of quadratic regulation functions is not restrictive. It follows from the *Weierstrass Approximation Theorem*, according to which the class of quadratic regulation functions is representative of the class of all regulation functions up to an error that is bounded above.

I perform the counterfactual exercise in two steps. The size-independent exercise, where only degenerate regulation functions are considered, allows me to understand the effect of changes in the overall level of regulation, and characterize optimal *size-independent* regulation. The size-dependent exercise allows me to characterize optimal *size-dependent* regulation. I consider the following regulation functions, examples of which are given in Figure 9.

- Size-independent regulation: $\chi(n) \equiv \chi_0 \ni \chi_0 \geq 1$

- Size-dependent regulation:

$$\chi(n) = \chi_2 n^2 + \chi_1 n + \chi_0; \quad \chi(n) \geq 1, \lim_{n\to\infty} \chi'(n) = 0$$

$$\implies \chi(n) = (\chi_s - \chi_l)(n/\overline{n})^2 - 2(\chi_s - \chi_l)(n/\overline{n}) + \chi_s$$

Note that the quadratic regulation function is described by three parameters. The limiting condition imposes restrictions on these parameters such that two parameters are sufficient to describe a regulation function. In fact, the quadratic regulation function can be written in terms of regulation χ_s for banks of size zero, limiting regulation χ_l, and a choice parameter $\overline{n}$ that controls the upper limit of the grid for bank net worth.

A note about the incidence of regulation is due here. The statutory regulation function is specified in terms of bank size. However, the incidence of regulation on banks depends on the equilibrium distribution of the banks. For example, consider the second alternate regulation function in the left panel in Figure 9. The smallest banks face regulation close to 22, while the largest banks face regulation close to 15. In this regime, if the equilibrium distribution of banks has a support of, say, $[4, 6]$, the incidence of regulation would only vary in an interval around 17.

For each alternative specification, I compute the stationary general equilibrium, and compare banking industry and macroeconomic aggregates. The response of the banking industry to changes in the level of regulation can be summarized as follows.

- Distribution: The distribution of bank net worth moves to the right as regulation becomes tighter (weaker leverage regulation), which leads to larger mean bank size, as shown in Figure 10. This result is in line with the observation in the data that as regulation has become tighter over time, the distribution has moved to the right. The variance in size distribution first increases, but eventually decreases as regulation tightens.

- Leverage and exit rate: As shown in Figure 11, mean leverage of banks is lower, and so is the overall failure rate. Wage paid to workers is lower as capital stock is lower, but dividend payment received by bankers are higher.

- Concentration: The concentration of the banking industry, as measured by a decrease in the Herfindahl index (computed assuming 5000 banks) decreases as regulation becomes tighter.

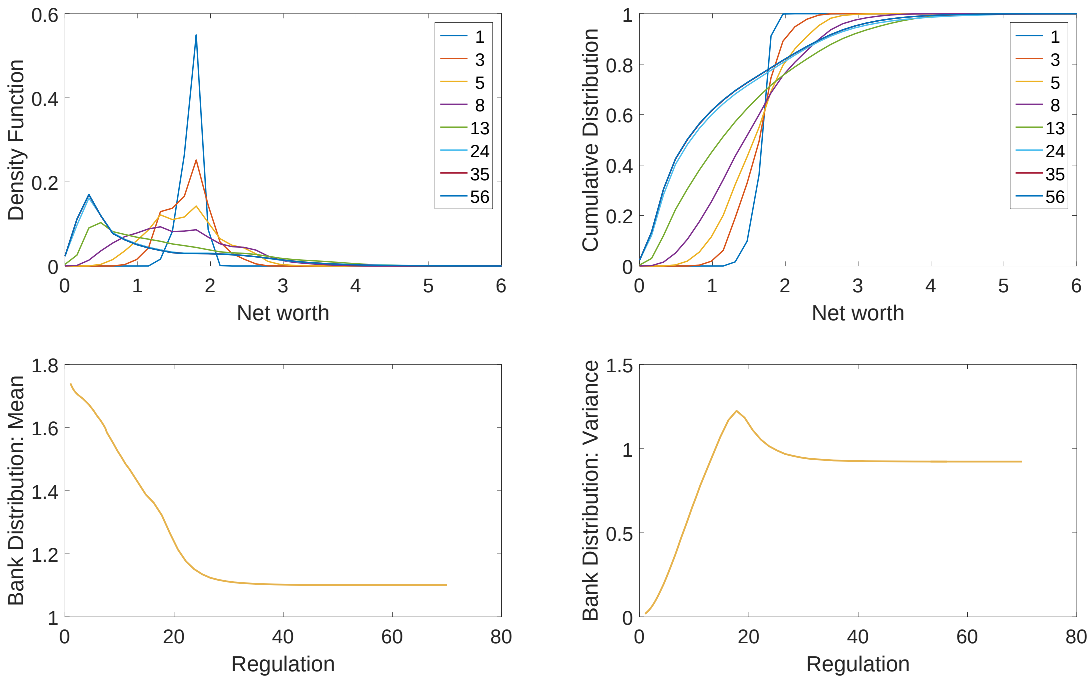


Figure 10: Response of bank distribution to changes in regulation *(In the figures in top row, the legend lists alternate regulation regimes)*

Also the mass of banks in equilibrium (relative to the mass of households, which is normalized to unity) increases, pointing towards a less concentrated industry.

6.1 Optimal regulation

In this section, I characterize optimal leverage regulation, relative to the benchmark economy where leverage regulation is $\hat{\chi} = 29.58$. To this end, I choose the regulation function $\chi(n)$ that maximizes the utilitarian welfare measure, i.e., the mass-weighted utilities of workers and bankers:

$$\max_{\chi \in \mathcal{C}^1(\mathbb{R}+)} \mathcal{W}(\chi) \quad \equiv \max_{\chi \in \mathcal{C}^1(\mathbb{R}+)} (1-\lambda)\frac{u(c_w(\chi))}{(1-\beta)} + \lambda \int_{\underline{d}}^{\bar{d}} \int_0^{\bar{h}} U(h, d; \chi) \partial \mu^b(h, d; \chi)$$

The optimal *size-independent* regulation is characterized by an overall more stringent regulation relative to the benchmark. Quantitatively, the optimal *size-independent* leverage regulation is 3.55 whereas the benchmark leverage regulation is 29.58. Assuming a 100% risk weight for the only (risky) asset in my model, the former translates into a capital ratio requirement of 28%, which is very aggressive relative to the consolidated 10.5% level proposed in Basel III. However, this

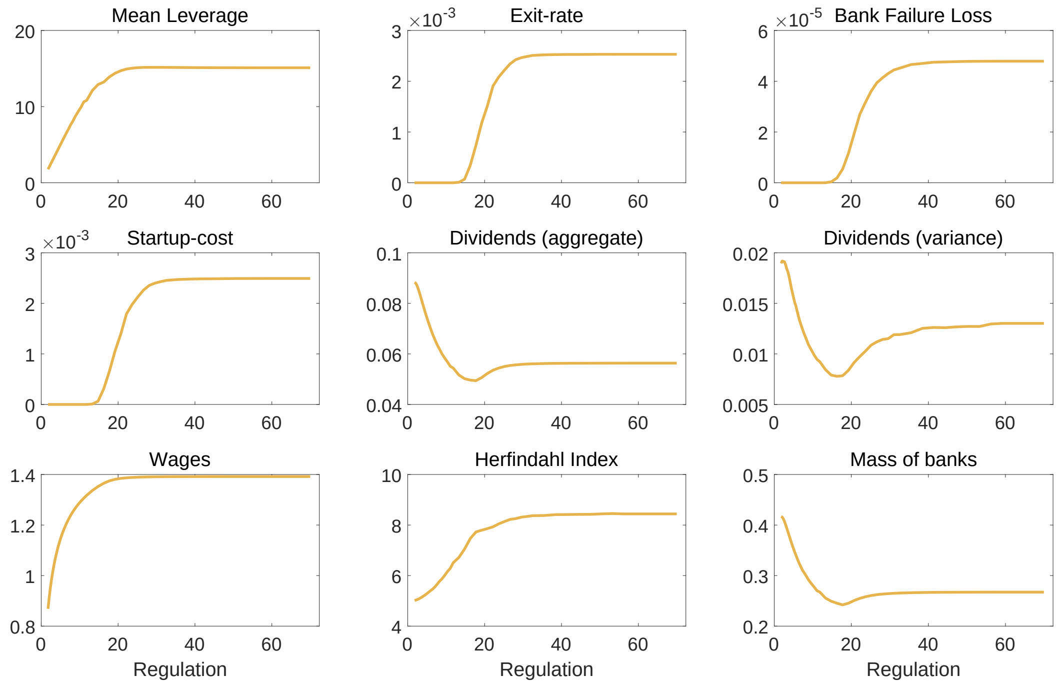


Figure 11: Response of industry to changes in regulation

aggressive capital ratio requirement is in line with that proposed in Admati and Hellwig [2014].

The optimal *size-dependent regulation* is characterized as being more stringent for large banks relative to small banks. Quantitatively, the optimal *size-dependent* leverage regulation is $\chi_s^* = 4.12$ for the smallest banks, and $\chi_l^* = 3.17$ for the largest banks, while varying quadratically between those extremes *(see Table 5 and Figure 12)*. Qualitatively, this result reinforces the proposition in Basel III regarding capital surcharge for SIFIs. The general rule for size-dependent regulation in fact complements the Basel proposition in the following sense. In determining the set of SIFIs, a regulator needs to factor in multiple, dynamically evolving characteristics of banks, and also review the incidence of regulation as the banking industry changes in response to regulation. The formula-based regulation proposed in this paper is not subject to these dynamic considerations.

Channels of welfare gain The rationale for an optimum that implies an overall tighter regulation is multi-dimensional. First, loss due to industry turnover is lower as less levered banks fail less often. This also lowers the start-up costs borne by bankers as a smaller mass of entrants enters the market. Second, the mean dividend income accruing to bankers is higher, primarily due to a larger equilibrium mass of banks, which dominates the welfare loss from lower wages for workers. Tighter

Statutory		Incidence					
Smallest	Largest	Smallest	10 percentile	Median	90 percentile	Largest	Mean
4.44	1.10	4.12	3.87	3.67	3.50	3.17	3.63

Table 5: Optimal Leverage Regulation: Statutory value, its incidence on banks of different sizes, and its mean incidence.

regulation results in higher aggregate bank net worth as banks are better capitalized, even though mean bank size is smaller. With regards to the profile of size dependent regulation, the following remarks summarize why tighter regulation of large banks relative to small banks improves welfare. I use the median bank size to classify banks as small or large.

Proposition 7. *Cleansing effect: The mean efficiency of failed banks is lower than the mean efficiency of entrant banks.*

Proof. The mean efficiency of entrants is 1 by calibration of the efficiency process. The mean efficiency of failed banks is given as:

$$m_\psi = \int\int\Big(\int_{\underline{\psi}}^{R(x(\psi,n)-1)/x(\psi,n)} \psi' f(\psi,\psi')\partial\psi\Big)\partial\mu(\psi,n)$$

Since the upper limit on leverage for all banks is equal to $\chi_s^* = 4.12$, this implies that $x(\psi,n) \leq \chi_s^* \;\; \forall \; \psi, n \implies R(x(\psi,n)-1)/x(\psi,n) < 1$. It then follows trivially from the integral that $m_\psi < 1$. ■

Intuitively, for a given level of current period efficiency, bank failure probability is increases with leverage. Therefore, by allowing small banks to be more highly levered, these banks experience a higher failure rate. However the ones that fail are less efficient relative to incumbents. Moreover, the entrants that replace these failed banks draw their efficiency from the ergodic distribution, and have higher mean efficiency. The higher turnover rate among small banks entails real costs, but optimally, this improves welfare since less efficient banks are eliminated, and failed small banks are, on average, immediately replaced by entrants which are also smaller.

Proposition 8. *Selection effect: For banks with positive expected premium $E[R_k\psi' - R|\psi]$ on intermediation, the expected growth rate for small banks is higher relative to the expected growth rate of large banks.*

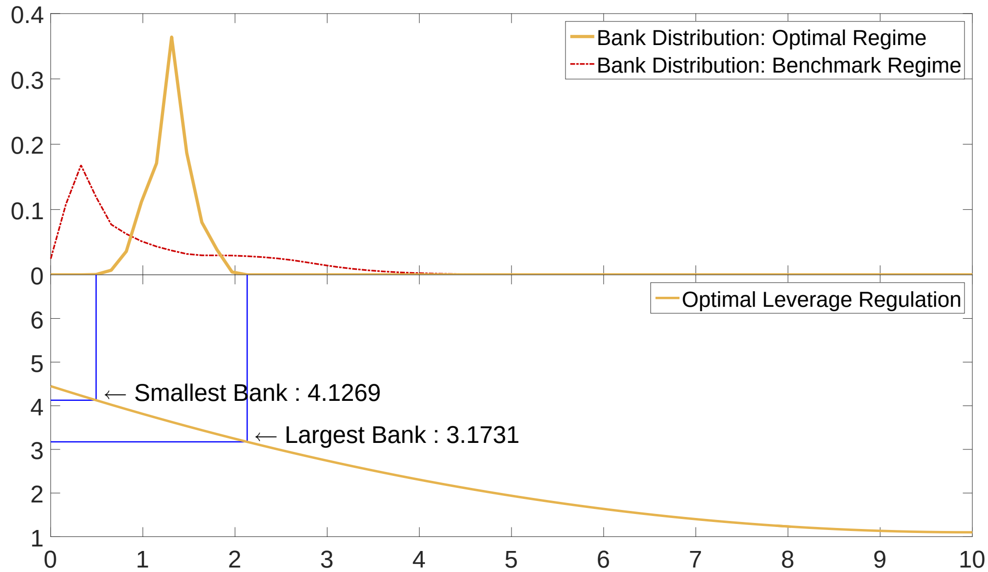


Figure 12: *(Top)* Bank size distribution in the benchmark and the optimal regimes *(Bottom)* Optimal leverage regulation, and its incidence on smallest and largest banks

Proof. The expected growth rate $l(\psi, n)$ of banks of size n can be defined as:

$$l(\psi, n) = \big[(R_k E[\psi'|\psi] - R)x(\psi, n) + R\big](1 - h(\psi, n))$$

$$\implies \partial l(\psi, n)/\partial n = \big[(R_k E[\psi'|\psi] - R)\partial x(\psi, n)/\partial n\big](1 - h(\psi, n)) +$$
$$\big[(R_k E[\psi'|\psi] - R)x(\psi, n) + R\big](-\partial h(\psi, n)/\partial n)$$

From the profile of optimal regulation, I know that $\partial x(\psi, n)/\partial n < 0$, and from comparative statics, I know that larger banks pay higher dividends, i.e., $\partial h(\psi, n)/\partial n > 0$. Then, since $(R_k E[\psi'|\psi] - R) > 0$, I have $\partial l(\psi, n)/\partial n < 0$. ■

Intuitively, by allowing the smaller banks to take on more leverage, the regulator allows the small banks to take on more risk, and grow faster *conditional* on survival. Naturally, more small banks end up failing this way, but the failure of small banks is less costly, as discussed above, and hence improves welfare.

Proposition 9. *Stabilization effect: The probability of failure for large banks is lower than the probability of failure for small banks.*

Proof. Bank failure probability is given as:

$$\hat{p}(\psi, n) = \int_{\underline{\psi}}^{R(x(\psi,n)-1)/x(\psi,n)} f(\psi, \psi')\partial\psi'$$

From the profile of optimal regulation, I know that $\partial x(\psi, n)/\partial n < 0$. Therefore, since the upper limit of the integral is increasing in $x(\psi, n)$, the proposition follows trivially.[27] ■

By tightening the regulation for large banks, the regulator prohibits them from taking on more risk, therefore rendering them stable. The downside is that large banks are more efficient in financial intermediation in equilibrium, and limiting the size of their intermediation can be costly. However, their failures are also costly because they are difficult to replace immediately. In particular, an entrant bank needs time to grow large to actually replace a failed large bank, and in each period during this time, additional operating costs are borne.

The gain in welfare from imposing regulation can be expressed in terms of consumption equivalence. Let A denote the benchmark regime (denoted by red marker in Figure 13) and let B denote the optimal regime (denote by blue marker). Then welfare gain in consumption equivalent (CE) units is *defined* as the percentage increase in consumption all agents would receive should they live in the optimal regime forever. Formally, the gain ω_U is defined as one that satisfies the following equation.

$$(1-\lambda)\frac{u((1+\omega_U)c_w^A)}{(1-\beta)} + \lambda\int_{\underline{d}}^{\overline{d}}\int_0^{\bar{h}} E_0 \sum \beta^t u((1+\omega_U)c_{bt}^A(h,d))\partial\mu_b^A(h,d) =$$

$$(1-\lambda)\frac{u(c_w^B)}{(1-\beta)} + \lambda\int_{\underline{d}}^{\overline{d}}\int_0^{\bar{h}} E_0 \sum \beta^t u(c_{bt}^B(h,d))\partial\mu_b^B(h,d)$$

The steady state gain in welfare in CE terms for a size-independent regulation is 2.2%, while that for a size-dependent regulation is 10.5%. These are large gains relative to, for example, the costs of business cycles, but comparable to, for example, the contemporaneous GDP cost of the 2008 financial crisis, which is 6.2%, and the lifetime GDP cost in the range of $40-90\%$ as estimated by Luttrell et al. [2013]. In other words, these welfare gains are not implausible from a steady state comparison perspective. If the transition of the economy from the benchmark regime to the optimal regime is taken into account, the gain in welfare will be lower. Intuitively, in the initial

[27] In fact, the proposition is reinforced due to the property that distribution of ψ' is first order stochastically increasing in ψ, and the fact that in equilibrium, larger banks have a first order dominant distribution over ψ relative to smaller banks.

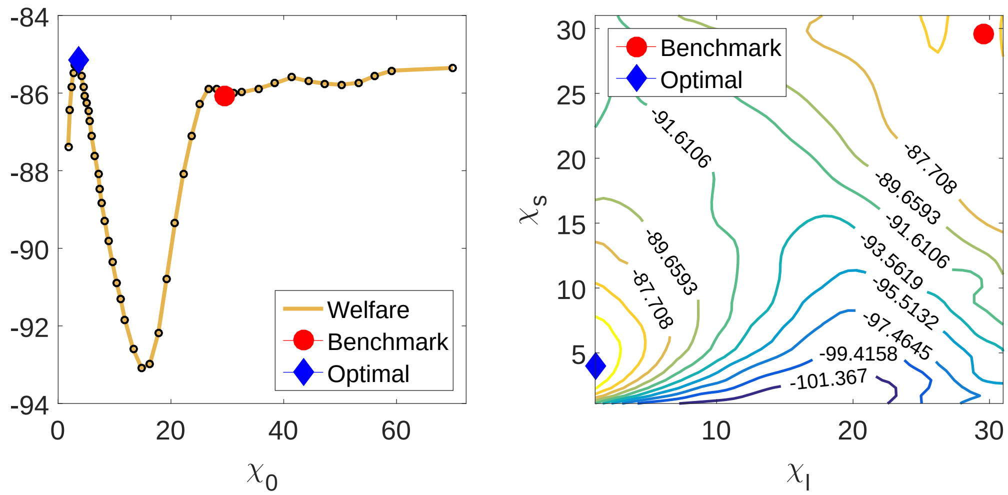


Figure 13: Welfare measures corresponding to *(Left)* Size-independent regulation regimes and *(Right)* Size-dependent regulation regimes (contour plot)

periods after the sudden imposition of optimal regime, the regulatory constraint binds the tightest as banks require time to re-optimize. Given that re-optimization requires several periods to achieve, the private decisions of economic agents are sub-optimal during transition, resulting in aggregate inefficiency and welfare loss. This phenomenon provides one of the explanations why a *myopic* economy might choose not to adopt such a policy change, even though eventually it can be in a much better state in terms of welfare.

6.2 Sensitivity Analysis

In this section, I examine the sensitivity of optimal leverage regulation to alternative calibrations. Optimal regulation is specifically sensitive to the relative risk aversion of households γ, the fraction of bankers and workers λ, and the tax rate on bankers τ. I vary these parameters on a discrete set containing the baseline calibration.[28] The result of this exercise is documented in Table 6.

A decrease in relative risk aversion implies a relatively liberal leverage regulation. Intuitively, as households become less risk averse, the welfare loss from dividend income volatility is smaller. Since dividend payment volatility increases with bank leverage, the optimal regulation is relatively liberal in an economy with lower risk aversion. An increase in risk aversion relative to benchmark does not result in a significant change in either the statutory regulation or its incidence. If the fraction of bankers, who are the households directly affected by banking industry dynamics, is smaller, the

[28]Since the computation of optimal regulation for each calibration requires close to 40 hours, I can perform only a limited number of these experiments.

Parameter Perturbed (Benchmark Value)	Perturbed Value	Statutory		Incidence		
		Smallest	Largest	Smallest	Largest	Mean
Benchmark Regime		4.44	1.10	4.12	3.17	3.63
γ (1.5)	1	6.66	1.10	6.47	4.40	5.36
	2	4.44	1.39	4.15	3.28	3.69
λ (0.475)	0.4	4.74	1.10	4.51	3.35	3.86
	0.5	4.14	1.10	3.85	2.98	3.40
τ (0.2416)	0.2	4.92	1.10	4.67	3.46	4.00
	0.3	3.97	1.10	3.69	2.87	3.26
θ (0.1796)	0	4.93	1.10	4.56	3.46	3.92

Table 6: Sensitivity analysis: A regime is characterized by the value of the parameter that is perturbed relative to benchmark. The shaded row represents benchmark. In the first set of alternative regimes, *ceteris paribus*, γ is increased and decreased relative to its benchmark value of 1.5. The tightest allowed leverage regulation is 1.10, ensuring a non-trivial problem for the banks.

weightage of bankers in the welfare function is smaller. Since tighter regulation improves welfare for bankers and liberal regulation improves welfare for workers, a smaller fraction of bankers shifts optimal regulation in the liberal direction. Similarly, a larger fraction of bankers shifts optimal regulation in the tighter direction.

Finally, as taxes on bankers shift consumption from bankers to workers, lowering tax makes bankers better off. This means that a more liberal regulation would make workers better off while not making bankers much worse off. Consequently, optimal regulation is relatively liberal. Similarly, higher taxes on bankers make them worse off, shifting the optimal regulation in their favor. Overall, this exercise shows that the quantitative value of optimal leverage regulation is sensitive to the exact calibration, and intuitively appealing. However, for small changes in the calibration in the beighbourhood of the benchmark calibration, the change in optimal regulation is small, and in this sense, the optimal regulation estimated in this paper is robust.

Case without fire sale issue To test the progressivity of size-dependent bank leverage regulation, I consider a version of the model where banks do not have the option to divert assets, and therefore, they are not subject to the incentive constraint that arises due to the limited committment problem. For the purpose of this sensitivity analysis, I do not recalibrate the model, leaving parameters at their baseline values. However, I do recompute the stationary equilibrium and as-

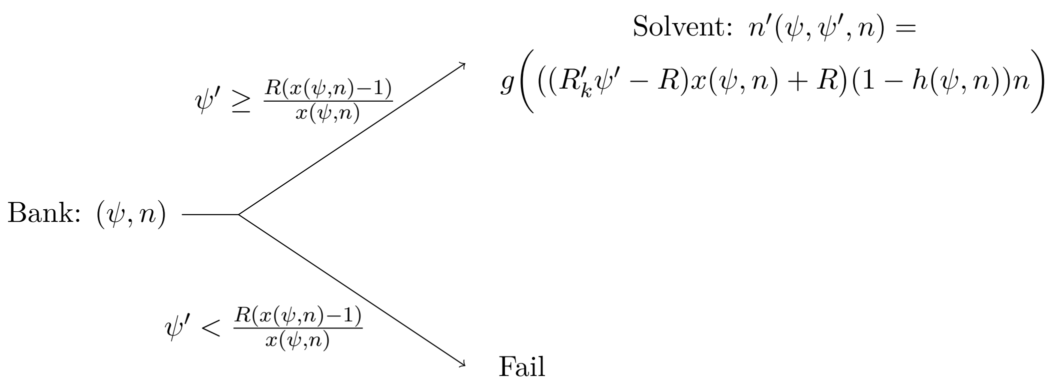


Figure 14: Dynamics of an individual bank

sociated welfare for the counterfactual policy regimes. The finding from this exercise is reported in Table 6. The conclusion is that qualitatively, the optimal regulation is size-dependent, and is tighter for large banks relative to small banks, as in the benchmark. Quantitatively, the optimal regulation is less stringent relative to the benchmark, a result that is driven by the absence of the fire sales issue.

6.3 Aggregate shocks

Although there is no aggregate uncertainty in my model, in this subsection, I consider a simple exercise wherein banks receive an unexpected aggregate shock to their efficiency ψ at date 1. Assuming that the economy is in the steady state at date 0, I compute the mass of bank failures that occur at the beginning of date 1 as a result of this shock. I then compare the incremental bank failures that occur due to the aggregate shock across the benchmark and optimal regimes. The objective of this exercise is to show that the optimal regulation reduces the incidence of an aggregate shock by reducing the failure rate of banks, and also by reducing the *distress rate* of (surviving) banks as measured by the expected percentage decrease in net worth as a result of the aggregate shock. In other words, this exericse shows that a less levered banking industry responds less aggresively to an aggregate shock.

In the benchmark economy, the steady state failure rate of banks is 0.0022. An unexpected aggregate shock of $\epsilon = 5\%$ at the beginning of date 1 would result in a higher failure rate as more number of banks will find their leverage too high to sustain the shock and remain solvent i.e. maintain a positive net-worth. I must clarify here that using this approach I can only compute the instantaneous mass of failures that occur at the beginning of date 1. To compute the response of the

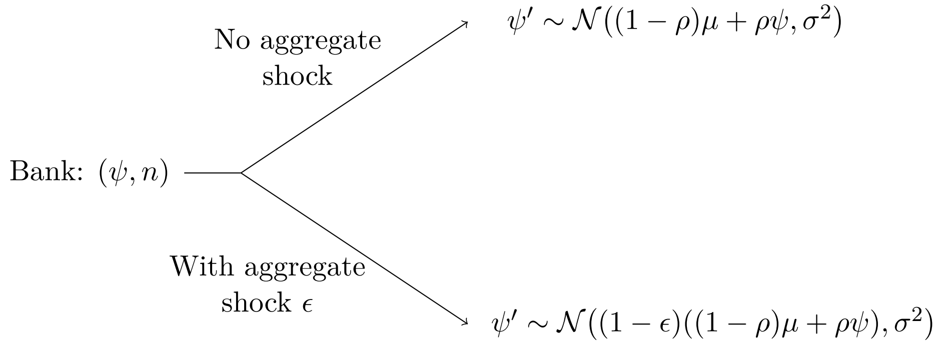


Figure 15: Distribution of realized shocks

economy thereafter, I would need to compute the full transition dynamics of the economy back to the steady state, taking into consideration the persistence of the shock. However, this computation is outside the scope of this paper. In order to compute the effective failure rate at date 1, I begin with the steady state equilibrium, and compute the next period solvency and net-worth positions of individual banks as shown in Figure 14. Without an aggregate shock, ψ follows a standard AR(1) process with mean $\mu = 1$. With an aggregate shock that affects the mean of the conditional distribution of bank efficiency, the resulting distribution looks as shown in Figure 15. In Figure 16, I document the failure and distress rates in the economies with the benchmark regulation (left panel) and the optimal regulation (right panel).

7 Conclusion

In this paper I characterize optimal size-dependent leverage regulation for banks. To this end, I develop a stochastic general equilibrium macroeconomic model with a heterogeneous financial sector. A key feature of the model is that bank leverage decisions play a critical role in determining banking sector dynamics and macroeconomic aggregates. Financial frictions render bank leverage choices inefficient, which rationalizes leverage regulation policy in the model. I show that generally tighter regulation generates welfare gains relative to the benchmark. In particular, I show that the optimal regulation is characterized by tighter regulation for large banks relative to small banks. In this sense, the paper qualitatively reinforces the Basel III proposition that systemically important financial institutions, which are generally the larger banks, must face tighter regulation relative to smaller banks. Quantitatively, relative to what Basel III proposes, the paper points towards an overall tighter regulation for all banks.

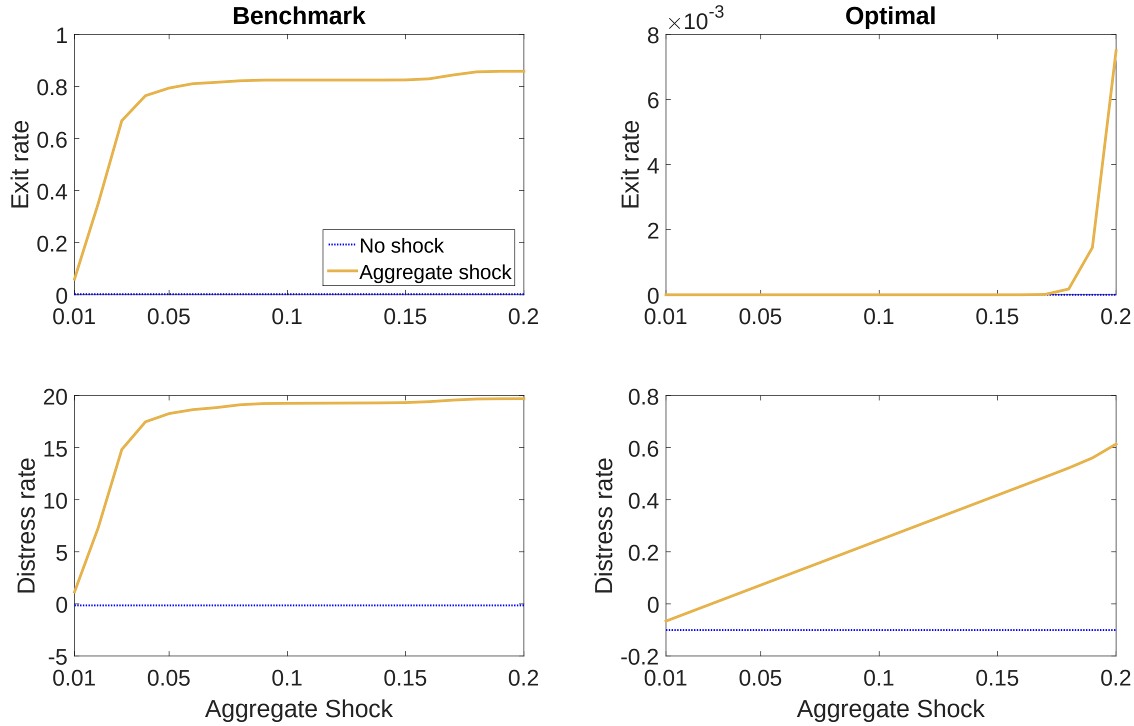


Figure 16: *(Top)* Exit rate; *(Bottom)* Distress rate *(percentage decrease in net worth)*

The paper contributes to the existing literature in two ways. Firstly, this is one of the few papers that take up the question of size-dependent bank regulation in a macroeconomic setting. Secondly, the model developed here captures the two main mechanisms behind banking industry dynamics, namely leverage and financial frictions, something existing models of firm dynamics do not focus on. One reason that this element is missing from the literature is that for non-financial firms, leverage is not the most critical element in the nature of their business, whereas for banks, leverage is central to their business.

The analysis in this paper can be extended in at least two ways. First, it would be interesting to study the welfare implications of the size dependent regulation when there are other policies present in the model, for example, liquidity regulation or risk based capital regulation. Since the model features an endogenously determined size distribution of banks, independent analysis of size-dependent policies like branching regulation *(for example, explicit constraints on bank size)*, and anti-trust regulation is also feasible in this framework. Second, time-invariant leverage or capital regulation can have pro-cyclical effects, and can exacerbate recessions. It would therefore be an insightful extension of my work to introduce aggregate uncertainty to this model, and study countercyclical capital regulation, something we pursue in a different context in a representative bank setting in Agarwal and Goel [2016]. Of course, introducing aggregate uncertainty in the current setup would necessitate solution techniques developed in Krusell and Smith [1998] or Reiter [2009], as the distribution of banks would become part of the state space. As a first step in this direction, in a supplement to this paper which is currently in preparation, I solve the deterministic transition dynamics of the model following a one time unexpected shock.

Chapter 2

Financial Contagion in Multiple-edge Interbank Networks

1 Introduction

There is growing consensus among academics and practitioners that the nature of interconnectedness in the financial system has a major role to play in shaping the magnitude of a financial crisis. On one hand, linkages between financial institutions offer avenues for diversification of risk, but on the other hand, these linkages can also serve as channels for transmission of negative shocks from one bank[29] to the other. A growing literature contributes to the study of this *stability* - *fragility* trade-off as a financial network becomes more *complete*. Allen and Gale [2000] and Freixas et al. [2000] are among the first papers to study this issue. They show using stylized models that a complete financial network is more resilient to contagion compared to a relatively to an incomplete network. In contrast, Acemoglu and Jensen [2015] show that it is not generally true that more complete financial networks are less prone to financial contagion. In particular, they show that for densely connected financial networks in case of sufficiently small negative shocks, financial contagion is weak, but for larger negative shocks, financial contagion can be disproportionately stronger - exhibiting a form of phase transition.

A common feature of these and several other studies of contagion in financial networks is that they consider financial network linkages to be *single-edged*, representing at most one of the following linkages possible between banks: standard debt contract (e.g. Gai et al. [2011]), cross holding of deposits (e.g. Dasgupta [2004]), collateralized debt contract etc. In this paper, I study *double-*

[29]The term 'bank' is used to indicate financial institutions in general.

edge financial networks, where the linkage between a given pair of banks consists of two distinct weighted edges, each representing a fundamentally distinct relationship between banks. The goal of pursuing this extension is to understand whether the presence of more than one edge between banks renders the network more stable or more fragile, and whether contagion is non-monotonic in the completeness of the network.

The motivation behind studying double-edge financial networks is that banks engage in several fundamentally different direct or indirect relationships with other banks. Some examples are the overnight federal funds market, short term borrowing and lending market (maturities greater than a week, but less than three months), interbank forex market (a key market in both domestic and international financial systems), securitized borrowing and lending market, and cross holding of assets (like REPO). The conventional wisdom is that these financial relationships provide avenues for diversification of risk. However, the 2008 financial crisis revealed that these relationships can also serve as channels for the transmission of shocks. In this paper, the goal is to understand the latter effect when more than one interbank relationships are considered together. In particular, I focus on the following interbank relationships.

The first relationship I consider is short term debt contracts, which engenders the possibility of the bank not able to roll over its debt in the event of a shock to the funding liquidity of the bank. One reason why this might be possible is that the lenders, which are often other banks, are themselves constrained. The second relationship I consider is cross-holding of assets due to *securitization*, which engenders the possibility of shocks to return on assets originated by one bank, impacting the balance sheet of other banks. The mechanism through which this impact is transmitted from one bank to the other is the *fire sale* of assets. Banks that need to sell assets at fire sale prices due to exogenous shocks, can induce asset price depressions, which in turn can deteriorate the balance sheet of other banks which hold the same asset. Fire sales can trigger another mechanism that causes transmission of the initial shock in the case of *collateralized debt* contracts *à la* Kiyotaki and Moore [1997]. The fire sale of asset depresses the price of asset which also serves as debt collateral, therefore reducing the borrowing capacity of all banks that hold that asset as collateral.

These two relationships by which distress propagates from one bank to the other, are not independent. Banks which are unable to roll over their debt due to some exogenous shock, can be forced to reduce their balance sheet size by selling assets, possibly in a fire sale. In a second round effect, fire sales that depress asset prices, deteriorate balance sheet of other banks which hold the same asset, possibly leading to their insolvency. To capture these insights, I build upon the framework of Allen and Gale [2000], and incorporate double-edged linkages in the financial network. There are three periods and four regions in the model. Each region has a continuum of banks and households. Households deposit their endowment in the bank, which in turn, invests in long and short assets. To induce banks in one region to transact with banks in other regions, I have two types of idiosyncratic shocks in the model – the standard Diamond and Dybvig [1983] style liquidity preference shock, and an asset return shock. Banks are ex-ante identical, and to hedge against these shocks, exchange deposits and assets at date 0. Liquidity shocks are realized at date 1, when deposit cross holding may be liquidated. When unexpected liquidity demand shock hits a bank in the network, the bank can be forced to liquidate it long assets before maturity. However, with large enough liquidity demand shocks, the bank may not have enough long assets to fulfill liquidity demands even after liquidation, which can induce a run on the bank and lead to its bankruptcy. Given interbank linkages, the bankruptcy of one bank has implications for solvency of other banks in the network. In particular, the two-dimensional linkage between banks serves as two channels of not only risk sharing but also channels for transmission of contagion. I derive parameterizations for which the financial network becomes more fragile, and show that with reasonably probable parameterization, the contagion problem is worsened in multiple-edge networks relative to the corresponding single-edge network.

Related literature A problem related to financial contagion is interbank freeze, where banks become unwilling to roll over debt contracts, and possibly engage in credit rationing. This hampers the functioning of financial intermediation. Allen et al. [2009] show that liquidity in the interbank market is key to the efficient functioning of the market, and that in the event of relatively large aggregate shocks, interbank markets may freeze. Müller [2006] shows that interbank credit line is an important channel that contributes to interbank freeze. In summary, these studies look at either the contagion, or the interbank freeze problem in a single-edge network model, trying to identify channels of contagion or interbank freeze, or characterize optimal networks resilient to these problem. Along similar lines, Babus [2007] develops a model of network formation, where banks form linkages to insure against the risk of contagion. Caballero and Simsek [2009] develop the notion of complexity externality to show how a bank's lack of knowledge about the network of its neighbors can contribute to financial fragility.

Cifuentes et al. [2005] study two channels of financial contagion - one is the direct channel i.e. balance sheet linkages, and the other is indirect channel that operates via asset price changes. In this sense, their paper is closely related to this paper, the key difference being that I focus on liquidity risk as the trigger for financial contagion, whereas Cifuentes et al. [2005] focus on credit risk. My focus on liquidity freeze is guided by the large empirical literature that points towards the *liquidity freeze* problem to be at the heart of the 2008 financial crisis (e.g. Acharya [2011], Domanski and Turner [2011]).

The paper is organized as follows. In section 2, I discuss multiple-edge networks linkages in general, non-financial framework developed in Morris [2000]. I show that the contagion properties of a multiple-edge network are different from those of a single-edge network. In particular, I show that relative to a single-edge network, it is not obvious whether contagion becomes weaker or stronger in a multiple-edge network. This result reinforces the relevance of the study of financial contagion in multiple-edge financial networks, a model that is empirically motivated by the existence of multi-dimensional interbank relationships. I present the empirical evidence on interbank relationships in Section 3. Section 4 builds upon the model of Allen and Gale [2000] to include double-edge linkages between banks. Section 5 compares the contagion property of single-edge versus double-edge networks, and section 6 concludes.

2 Contagion in multiple-edge networks

I extend Morris [2000] to allow for double-edge linkages between agents. This purpose of this extension is to show in a general non-bank context that it is not obvious whether double-edge networks are either more or less prone to contagion relative to single-edge networks (as opposed to the prior one might have that more linkages always implies lower probability of contagion). In this sense, this extension serves to reinforce the motivation behind working with double-edge network models when the actual phenomenon being studied features multiple-edge linkages between economic agents. To this end, I consider an infinite population of strategically interacting agents. Each agent chooses an action from a binary set to maximize its payoff. The payoff for an agent is a function of its own action and the action of its finitely many neighbors. The linkage between two agents consists of two undirected weighted edges. I study the contagion property of arbitrary double-edge networks, and contrast the result with corresponding single-edge networks. I begin by formally defining contagion in this context.

Contagion is defined as the spread of an action to the whole network from a finite subset of the network. Morris [2000] studies local interaction games on a network to characterize the set of network parameters where contagion is possible. He considers single-edge networks where an agent is connected to another agent by means of a single, undirected and unweighted edge. In this section, I aim to characterize contagion properties in a multiple-edge network where an agent is connected to another agent by means of two undirected but weighted edges. In particular, I am interested in understanding whether contagion becomes relatively easy or difficult as one moves from a single-edge network to a multiple (double) edge network.[30]

In moving from a single-edge network to a double-edge network, I replace each edge with a pair of edges that have weights whose sum is normalized to unity. I consider the pair of edges to be ordered. The action chosen by an agent (the node) is applied to the first edge in the pair, and the complement of the chosen action is applied to the second edge in the pair. Each edge in the pair has a weighted contribution to the overall payoff an agent realizes from its linkage to another agent.

The best response strategy for an agent is a function of her belief about what her neighbors play, the payoffs from both edges of each linkage, and also the edge weights. As we will see, the best response strategy does not necessarily shift in one direction compared to the single-edge network case – In particular, I prove the proposition that it is not necessary that contagion becomes more or less probable, and that it depends on network weights and interaction game parameters. In this sense, the contagion properties of a single-edge network does not translate naturally to it double-edge counterpart.

2.1 Model

Let $\mathcal{N}$ denote a countably infinite set of agents and let $\sim$ be a binary relation on $\mathcal{N}$ that represents the *neighborhood* relation i.e. $x \sim x' \implies x$ is directly connected to x'. $(\mathcal{N}, \sim)$ is referred to as the *local interaction system*. The linkage between x and x' could comprise of a single edge, in which case the network is single-edged, or comprise of two ordered and weighted edges, in which case the network is said to be double-edged. In the latter network, first edge in any linkage between two nodes has weight $w \in (0, 1)$ and second edge has weight $1 - w$. Each agent can choose an action $e \in \{0, 1\}$. This action entails to an effort e applied to the first edge, and a complementary effort

[30] There are multiple ways in which a single-edge network can be extended to a multiple-edge connected network. The extension considered in this paper is the simplest natural extension from Morris [2000] so as to be able to compare the implications for contagion.

Figure 17: (Left) single-edge (Right) multiple-edge

$1 - e$ applied to the second edge. The agent chooses action e to maximize aggregate payoff from all its linkages. The payoff from a linkage depends on the actions of the agent and its neighbor, and is the weighted sum of payoffs from the two edges that constitute that linkage.

The payoff from an edge is a function of the actions taken by the two agents connected by that edge. The payoff is given by the following game with a symmetric payoff matrix M:

Action	0	1
0	$u(0,0), u(0,0)$	$u(0,1), u(1,0)$
1	$u(1,0), u(0,1)$	$u(1,1), u(1,1)$

I assume that both edges in any linkage correspond to the same game represented by payoff matrix M. Then $(\mathcal{N}, \sim, M)$ can be referred to as the *double-edge local interaction game*. When the underlying network is single-edged, I refer to it as a *single-edge local interaction game*. In order to have two strict Nash equilibria of the game, it is assumed that $u(0,0) > u(1,0)$ and $u(1,1) > u(0,1)$. For convenience, I rename payoffs and retain only the row agent's payoffs in the payoff matrix.

Action	0	1
0	a	b
1	c	d

I will refer to a, b, c, d as payoff parameters of the game. Now in order to pin down the best response strategy of an agent, the payoff from both edges of each linkage of the agent needs to be considered. The augmented payoff matrix for a linkage, considering both edges, with relative weight w, is given below.

Action	0	1
0	$wa + (1-w)d = A$	$wb + (1-w)c = B$
1	$wc + (1-w)b = C$	$wd + (1-w)a = D$

I will refer to A, B, C, D as the augmented payoff parameters of the game. Note that the condition for two strict Nash equilibria is satisfied in the augmented payoff matrix i.e. $A > B$ and $D > C$.

Since total payoff for the agent is the sum of payoffs from each of the linkages, an agent's best response strategy would depend on its belief about what effort each of its neighbors are exerting. In particular, if the agent believes that a proportion, say q, of its neighbors are exerting effort 1, then the expected payoff of the agent by choosing action $e = 0$ (which means effort 0 on first edge and effort 1 on second edge) is $qB + (1-q)A$. Similarly, the expected payoff of the agent by exerting effort $e = 1$ (on first edge and 0 on the second edge) is $qD + (1-q)C$. The agent would exert effort 1 if its expected payoff from effort 1 is greater than the expected payoff from effort 0:

$$\begin{aligned} qD + (1-q)C &\geq qB + (1-q)A \\ q &\geq \frac{(A-C)}{(D-B)-(A-C)} := q^{double} \end{aligned}$$

This means that if an agent believes that at least q^{double} of its neighbors are choosing action 1, then it is optimal for the agent to choose action 1 as well. I refer to q^{double} as the belief parameter of the *double-edged local interaction game* defined by payoff parameters a, b, c, d and weight parameter w, and is given primitively by the following expression:

$$q^{double} = \frac{w(a-c) + (1-w)(d-b)}{(a-c)+(d-b)}$$

Following the analysis of Morris [2000], note that for a *single-edge local interaction game* with the same payoff parameters, the belief parameter is given by the following expression:

$$q^{single} = \frac{(a-c)}{(a-c)+(d-b)}$$

Since the belief parameter determines the best response function of an agent, the payoff parameters can be normalized as follows for mathematical convenience:

$$(a-c) := q; \quad (d-b) := (1-q)$$

Henceforth, instead of referring to a, b, c, d as the payoff parameters of the game, I refer to q as the payoff parameter of the game. Consequently, the belief parameters of a local interaction game on single and double edged networks are given by the following expressions:

$$\begin{aligned} q^{single} &= q \\ q^{double} &= wq + (1-w)(1-q) \end{aligned}$$

The interpretation of belief parameter is that it is the minimum fraction of neighbors that must be choosing a certain action in order for the agent in question to choose that action. I now consider the contagion problem in this network, which aims to characterize conditions on parameters for which a given action spreads to the entire network starting from a finite set of agents. First note that an action spreads across a network via best response strategy of the agents. As seen in last section, best response strategy depends on the belief parameter of the local interaction game. For example, if the belief parameter is too large, then an agent needs a large fraction of her neighbors playing an action to induce her to play that action. This implies that spreading of an action may be more difficult if the belief parameter is larger, and therefore, I can define the contagion threshold as follows:

Definition 1. *The contagion threshold ξ of a network is defined as the largest belief parameter of a local interaction game for which contagion is possible.*

This means that for any local interaction game with belief parameter q, where the contagion threshold of the underlying network is ξ, contagion is not possible if $q > \xi$, and contagion is possible otherwise. Morris [2000] showed that $\xi \leq 1/2$ for any local interaction game. I use this property to prove the following proposition. First I note that the belief parameter of a *single-edged local interaction game* is not equal to the belief parameter of a *double-edged local interaction game*. This leads to the following lemmas:

Lemma 1. *If $q > 1/2$ then $q^{double} < q^{single}$ $\forall w$*

When payoff parameter q is larger than half, belief parameter becomes smaller as we move from a single-edge network to a double-edge network for any weight w.

Lemma 2. *If $q < 1/2$ then $q^{double} > q^{single}$ $\forall w$*

When payoff parameter q is smaller than half, belief parameter becomes smaller as we move from a single-edge network to a double-edge network for any weight w. With these lemmas, it is straightforward to prove the following proposition.

Proposition. *Consider a local interaction game with payoff parameter q on a single-edge network with contagion threshold ξ. (i) Next, suppose that contagion is possible in the single-edge network. Then there exists a double-edge network corresponding to the double-edge network*[31] *with appropriate weights and the same local interaction game where contagion is impossible ($q^{double} > \xi$). (ii) Finally, assume that $q : 1 - q < \xi$ and suppose that contagion is not possible in the single-edge network ($q = q^{single} > \xi$). Then there exists a corresponding double-edge network with appropriate weights and the same local interaction game where contagion is possible ($q^{double} < \xi$).*

Proof. Part (i) Recall that $q^{single} = q$ and $q^{double} = wq + (1 - w)(1 - q)$. Suppose that contagion is possible in the single-edge network, then $q^{single} = q < \xi$. Then, $\xi < 1/2 \implies q < 1/2$ and $1-q > \xi$. Using above relations, I get that for any $w < ((1-q)-\xi)/(1-2q)$, $q^{double} = wq+(1-w)(1-q) > \xi$ *i.e.* contagion is not possible on the double-edge network.

Part (ii) Assume that $q : 1 - q < \xi$ and suppose that contagion is not possible in the single-edge network $\implies q = q^{single} > \xi$. Then, $\xi < 1/2 \implies 1 - q < 1/2$ and $q > 1/2 > \xi$. Using the above, I get that for any $w < (\xi - (1-q))/(2q-1)$, $q^{double} = wq + (1 - w)(1 - q) < \xi$ *i.e.* contagion is possible on the double-edge network.

■

This section shows that contagion properties of a single-edge local interaction game are different from those of the corresponding double-edge local interaction game. In particular, I showed that with appropriate choice of weights, the possibility of contagion on a double-edge network could be of the opposite nature compared to that of a single-edge network. The takeaway is that it is not obvious that contagion always becomes easier or more difficult as one moves from one model setup to the other - it depends on the weight parameter of the double-edge network and on the game payoff parameters. Consequently, I claim that it is efficient to consider double-edge network models when studying contagion properties of real life phenomenon where any two agents have more than one linkage among them. This paper shows one application of this result to the notion of financial contagion in interbank market where banks form relationships with each other on multiple markets.

[31] That is, replacing all single edged linkages with double edged linkage

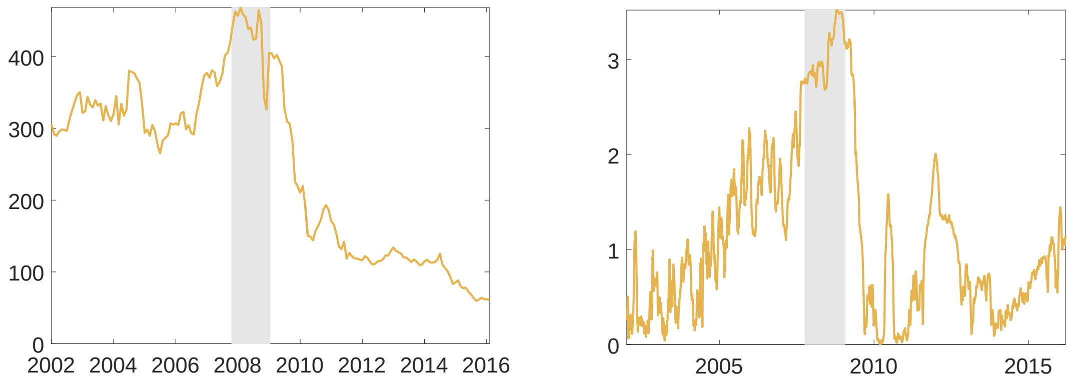
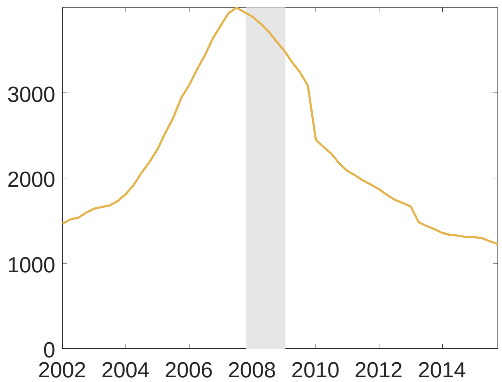


Figure 18: (Left) Interbank Loans: US Commercial Banks (Billion USD); Source: FRED (Right) Interbank Liquidity Spread; Source: Cleveland Financial Stability Index (CFSI)

3 Empirical evidence

Banks fund themselves using a range of products, including deposits, corporate bonds, repurchase agreements (REPO), commercial paper, etc. The counter-party in some of these funding arrangements is another bank. For example, banks borrow overnight from each other in the federal funds market, or for longer duration in the REPO or corporate bond market. Stress on the balance sheet of the lender, or increased risk perception about the borrower, are two important drivers of the breakdown of an interbank market, leading to a liquidity freeze problem. The volume of interbank loans (all maturities) rose substantially in the run upto the 2008 financial crisis, and fell substantially during and following the 2008 financial crisis (see Figure 18). Interbank liquidity spreads show a similar pattern, indicating build-up of risks in the run upto the crisis, and crash soon after that.

Banks also engage in cross-holding of assets, which I incorporate in the model in this paper as a potential driver of financial contagion. Assets (loans) originated by one bank can end up being held by other banks through the process of securitization and trade. The process of securitization and trade can assist in the diversification of risk, and render a financial system more stable. However, it is also possible that *bad* assets securitized along with a range of other assets can remain undetected and mis-priced for a long period, leading to a cascade effect once bad assets are identified by the market. Shin [2009] proposes a mechanism to explain this latter effect. While focusing on a demand side explanation, he shows that banks in order to fill in the expanding balance sheets, need to find borrowers, in the pursuit of which, the bank might originate loans to the *bad* borrowers with a high probability. Asset-backed-securities (ABS), including mortgage-backed-securities (MBS) are prime



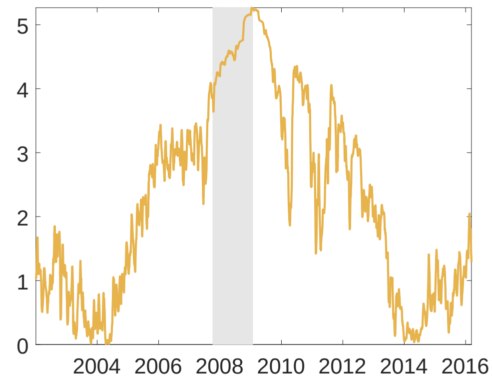


Figure 19: (Left) US Asset (loan)-backed securities (Billion USD); Source: FRED (Right) US Asset-backed-securities spread; Source: CFSI

examples of securitization. Agarwal et al. [2010] discuss how the ABS volumes skyrocketed leading up to the financial crisis (see Figure 19). I also document how the change in risk perception and risk materialization lead to the skyrocketing of yields, minuting the incentive to issue further ABS, and leading to the drying up of ABS and its associated credit markets.

4 Model

The model framework is closely based on the framework in Allen and Gale [2000], with two main departures. First is that I allow for stochastic return on assets, in addition to existing stochastic liquidity demand. In this sense, banks are subject to idiosyncratic shocks on both asset and liability sides of its balance sheet. This generates the motive for banks to engage in double-edged interbank linkages in order to diversify the two dimensional idiosyncratic risk.

The model is as follows: there are three dates, t = 0, 1, 2, and the economy consists of four regions A, B, C, D. Each region consists of a unit continuum of identical households and a unit continuum of identical banks. The regions are $ex-ante$ identical. $Ex-post$, the regions are different due to idiosyncratic shocks. Within each region i, the representative household is endowed with a unit of consumption good at date 0. There is no endowment at later dates. Household preferences are Diamond and Dybvig [1983] style:

$$U(c_1, c_2) = \begin{cases} u(c_1) & \text{w.p. } \omega_i \\ u(c_2) & \text{w.p. } (1-\omega_i) \end{cases}$$

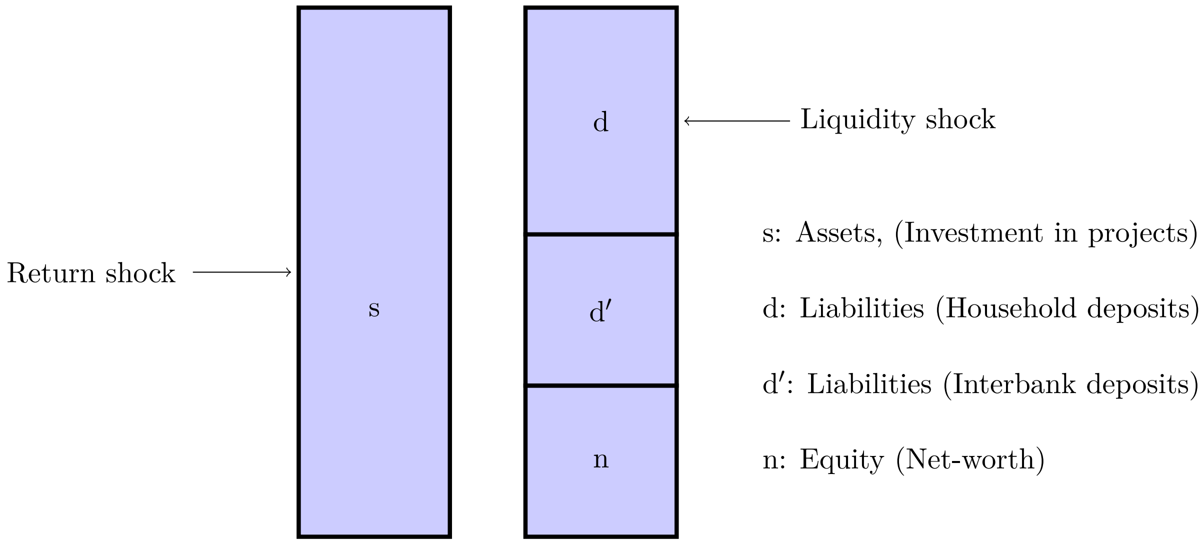


Figure 20: Stylized Bank Balance Sheet

where c_t is consumption at date $t = 1, 2$, ω_i is the fraction of early households and $1 - \omega_i$ is the fraction of late households. I assume that $u(.)$ is the log utility function. Household type in each region is revealed at date 1, and this is private information for the household. Households deposit their endowment in the representative bank in their region at date 0. At date 1, early households withdraw. Late households withdraw at date 1 only if they perceive that they will be better off withdrawing at date 1 instead of date 2.[32] Households are served pro-rata.

The representative bank in each region accepts unit deposit from the representative household and invests into two assets, of which one is liquid and the other is illiquid. The liquid asset is a short term asset that pays a return of 1 between date 0 and date 1. The illiquid asset is a long asset that pays a stochastic return $R > 1$ between date 0 and date 2. The illiquid asset, can be liquidated earlier at date 1 at a discounted return $r < 1$. Banks offer deposit contracts (d_1, d_2) at date 0 for withdrawals d_1 at date 1 and d_2 at date 2. A stylized bank balance sheet is given in Figure 20.

4.1 Shocks

There are two types of shocks in this paper. One is a liquidity demand shock, and the other is an asset return shock. Liquidity demand shock is realized at date 1, while asset return shock is realized at date 2. Liquidity demand shock may have aggregate uncertainty, but asset return

[32]Late households are assumed to have access to a storage technology so that they can withdraw at date 1 and consume at date 2.

	A	B	C	D
S_1	ω_H	ω_L	ω_H	ω_L
S_2	ω_L	ω_H	ω_L	ω_H

Table 7: Liquidity demand shock: No aggregate uncertainty

	A	B	C	D
S_1	ω_H	ω_L	ω_H	ω_L
S_2	ω_L	ω_H	ω_L	ω_H
$\bar{S}$	$\gamma + \epsilon$	γ	γ	γ

Table 8: Liquidity demand shock: Aggregate uncertainty

shock is assumed to have no aggregate uncertainty throughout this paper. Each region, receives idiosyncratic shocks of either type.

Liquidity demand shock I assume that in each region, the fraction of early households, ω, can take two values *i.e.* $\omega \in \{\omega_L, \omega_H\}$, $0 < \omega_L < \omega_H < 1$. Each region learns about the type of its households at date 1. Depending on the vector of shocks across regions, there are two possible states for the economy at date 1, as shown in Table 7. The states are assumed to be equi-probable. In order to study contagion, a small aggregate uncertainty is added to the model, as shown in Table 8, by means of introducing the possibility of occurrence of a new state $\bar{S}$ which has zero probability of occurrence at date 0. In this sense, this is an unexpected shock to the economy. The liquidity demands in this state are perturbed around $\gamma = (\omega_H + \omega_L)/2$.

Asset return shock Each region, at date 2, receives an idiosyncratic shock to the return on assets originated in that region. I assume that asset return can take two values $\{R_L, R_H\}$, $1 < R_L < R_H$. Since banks in a region may become insolvent or bankrupt at date 1, the possible states for the economy at date 2 depends on the state of the economy at date 1. The state of the economy at date 2 is given by the vector of asset return across regions. When banks in all regions are solvent at date 1, there are two equi-probable states for the economy at date 2, as shown in Table 9. In this case, the state of economy at date 2 is independent of the state of economy at date 1.[33]

[33] The assumption of independence between state of the economy at date 1 and date 2 can be easily relaxed in a more realistic model. I suspect that results will not change qualitatively. Nevertheless, if the assumption is to be relaxed, a reasonable approach is where the resolution of uncertainty with respect to liquidity demand partially reveals information with respect to asset return. For example, it could be more probable to move from state S_1 to say state N_2 etc. A state transition probability matrix could characterize this interdependence conveniently.

	A	B	C	D
N_1	R_H	R_L	R_H	R_L
N_2	R_L	R_H	R_L	R_H

Table 9: Asset return shocks: No aggregate uncertainty

However, if banks in any particular region go bankrupt at date 1, then for the assets that were originated in that region, return is discounted by α; return vector for other regions remains unaltered. For example, if banks in region A fail, the new equi-probable states of the world are as given in Table 10. The rationale behind this assumption is that once a bank becomes bankrupt at date 1 and information is revealed about the poor operations of the bank, the quality of its assets as perceived by the market is also affected adversely. A mechanism through which this assumption operates, for example, is the downgrading of rating of assets originated by a failed bank. Nevertheless, discounting of asset return by α is a reduced form approach to introducing this drop in expected return on assets originated by a bankrupt bank, and is for simplicity of exposition. A micro-founded approach, for example, would be to assume an ex-post change in relative probability of occurrence of high and low return on the assets originated by a bankrupt bank. These two approaches are, however, qualitatively equivalent.

	A	B	C	D
N_1^A	$R_H - \alpha$	R_L	R_H	R_L
N_2^A	$R_L - \alpha$	R_H	R_L	R_H

Table 10: Asset return shocks: Region A bankrupt

In addition to asset return adjustment, bankruptcy of a bank should have a negative impact on the liquidation value r of the assets originated by that bank. However, for simplicity, I continue to assume r to be non-stochastic and non-responsive to bankruptcies.[34]

[34]Liquidation value could be treated as stochastic, and be revealed at date 1. For example, if $r \in \{r_L, r_H\}$, then it would be a reasonable assumption that low liquidation values in a region are perfectly correlated with high liquidity demand in that region as households may look for other investment avenues. In other words, each region could be in one of the following states (ω_H, r_L) or (ω_L, r_H) at date 1. This ties back to the partial revelation of uncertainty in asset return discussed earlier, where if a region has low liquidation value at date 1, then it would be more probable that asset return in that region at date 2 would also be low.

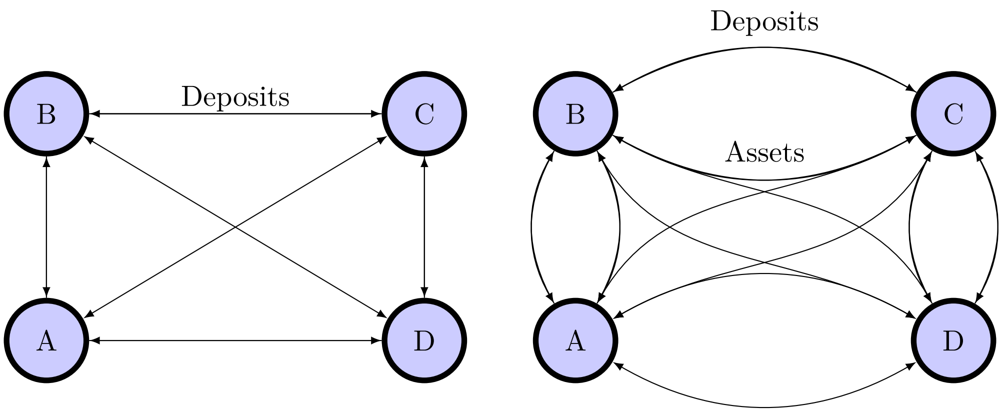


Figure 21: Complete Networks (i) single-edge (ii) multiple-edge

4.2 Interbank Networks

The four regions in this model can be inter-linked in multiple ways. A linkage from bank A to bank B in this model implies that A can make a deposit at A, and that A can buy assets from B – in this sense, a linkage in this model comprises of two edges. Since the direction of linkage matters, A linked to B does not imply that B can make a deposit at A, or that B can buy assets from A.

Following Allen and Gale [2000], two cases of the interbank network are considered below. First is the case of a complete network, as shown in Figure 21. Banks in each region are linked with banks in all other regions. Also note that the linkages comprise of both directions.

Allen and Gale [2000] show that a complete network is more resilient to contagion compared to an incomplete network shown in Figure 22. In particular, they show that there exist parameter values for which any equilibrium of an incomplete market involves contagion whereas some equilibrium of the complete market does not involve contagion. In this sense, the analysis of a multiple-edge network in complete and incomplete settings is non identical. In this paper, I consider the incomplete network and investigate how contagion changes as we move from singly to multiple-edge networks. I begin with the contagion problem fixed asset return case, followed by the stochastic asset return case.

4.3 Single-edge network: fixed asset return case

This is the case studied in Allen and Gale [2000]. Assets provide a fixed return of R, and the only shock in the model is a liquidity demand shock. To solve for the optimal and equilibrium solutions, the shock structure from Table 7 is considered.

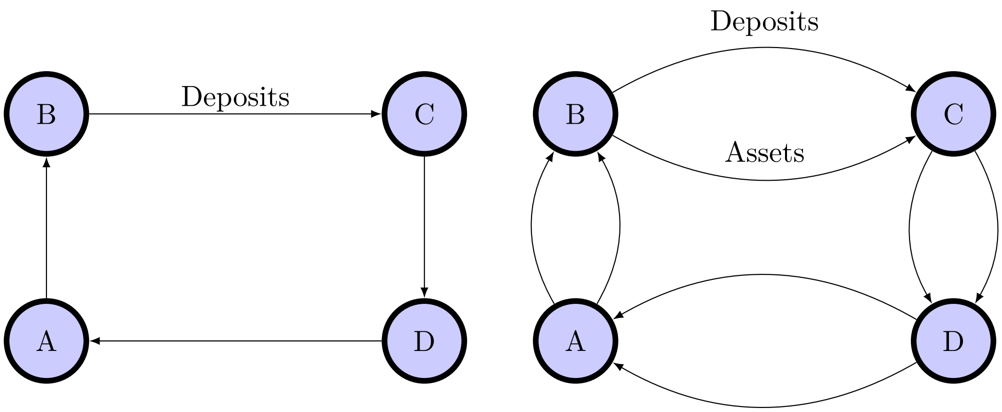


Figure 22: Incomplete Networks (i) single-edge (ii) multiple-edge

Planner's solution Since there is no aggregate uncertainty, and since the regions are ex-ante identical, the planner's allocations is identical across regions. Let x and y denote investment in long asset and short asset respectively in each region. Let c_1 and c_2 denote the consumption of early and late consumers respectively. For the economy as a whole, fraction of early consumers is given by $\gamma = (\omega_H + \omega_L)/2$. Then the following feasibility constraints must hold.

$$
\begin{aligned}
x + y &\leq 1 \quad \text{(date 0)} \\
\gamma c_1 &\leq y \quad \text{(date 1)} \\
(1-\gamma)c_2 &\leq Rx \quad \text{(date 2)}
\end{aligned}
$$

The planner's objective is to maximize ex-ante expected utility of households, which is given by:

$$\gamma u(c_1) + (1-\gamma)u(c_2)$$

For the rest of this paper, let (x, y, c_1, c_2) denote the first best allocation obtained from the above maximization problem subject to the feasibility constraints. The first best solution satisfies the incentive constraint $c_1 \leq c_2$, which ensures that the late households find it weakly optimal to withdraw at date 2 and not pretend to be early households and withdraw at date 1. The first best allocation involves transfer of resources among regions depending on the state of the economy. Intuitively, since there is no aggregate uncertainty, allocations can be made at date 0 with aggregate resources in mind, with necessary re-distributions at date 1.

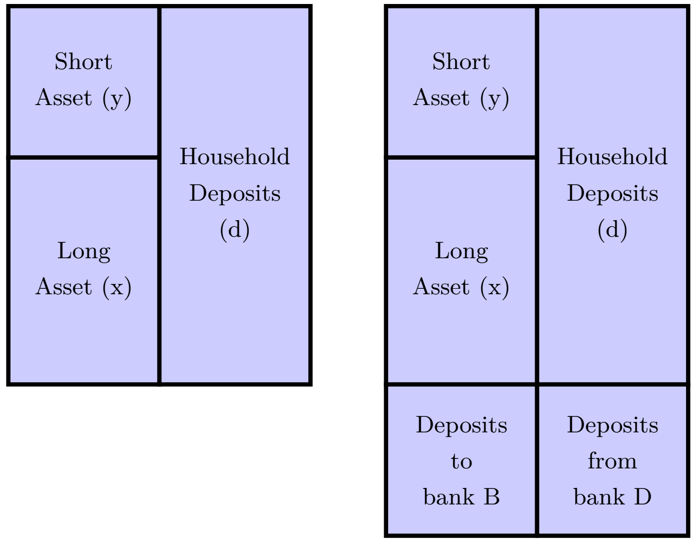


Figure 23: Balance sheet (i) no interbank market (ii) single interbank market

Decentralized Equilibrium In a decentralized equilibrium, the first best allocation is not feasible. For example, in a region where liquidity demand is high, ω_H, the short asset will not meet liquidity demand since $y \leq \gamma c_1 < \omega_H c_1$. Liquidation of long asset is possible to meet liquidity demand at date 1, but that would lead to unaccounted liquidity demand at date 2. Therefore, although there are enough resources in the economy as whole, the distribution is not efficient. The need for distribution paves way for an interbank market where banks exchange deposits. For the single-edge incomplete interbank network in Figure 22 (i), the representative bank in region A forms linkage with bank B, and so on. Specifically, bank A (i.e. the representative bank in region A) holds z units of deposits in region B, bank B holds z units of deposit in bank C, and so on. A typical balance sheet after interbank trade is given in Figure 23. If a bank receives high liquidity demand shock, it calls upon its deposits in other banks at date 1, otherwise at date 2. Allen and Gale [2000] show that with $z = \omega_H - \gamma$, the equilibrium coincides with planner's solution.

Aggregate uncertainty in liquidity demand shock In this section, in order to study contagion, I introduce aggregate uncertainty as depicted in Table 8. The first best allocation stands unaffected since the new state $\bar{S}$ which represents aggregate uncertainty has zero probability at date 0. Even the decentralized equilibrium allocation at date 0 remain unchanged. However, the continuation equilibrium is different when $\bar{S}$ is realized at date 1.

In state $\bar{S}$, bank A needs to liquidate part of its long asset to meet the extra demand for liquidity

ϵc_1. But the long asset can be liquidated only to an extent that late customers receive at least c_1 at date 2, otherwise they will run on the bank at date 1. The maximum extra liquidity (i.e. liquidity in excess of that provided by the short asset) that a bank can generate at date 1 by liquidating some of the long asset is is referred to as the *buffer* and is given by

$$b(\omega) = r\left[x - \frac{(1-\omega)c_1}{R}\right] \tag{4.3.1}$$

where ω is the liquidity demand at date 1. Next, let q^i be the value of deposits in region i at date 1. Note that at date 1, either $q^i = c_1$, when only early customers withdraw, or $q^i < c_1$, when all depositors at bank i withdraw. In the latter case, the value for bank A for example is given as follows:

$$q^A = \frac{y + rx + zq^B}{1+z} \tag{4.3.2}$$

The numerator represents assets and denominator represents liabilities; q^A equates the two quantities, given that the bank is liquidated. For bank A to remain solvent, the following must be true:

$$b(\gamma + \epsilon) \geq \epsilon c_1 \tag{4.3.3}$$

Suppose this is not true. Then $q^A < c_1$ and there will be a run on bank A. Bank A would be bankrupt, and would call upon its deposits with bank B. All regions will liquidate cross holdings of deposits. The run on bank A induces bank D to run on the deposits it had placed with bank A. As bank D withdraws from bank A, the maximum value it can get is:

$$\bar{q}^A = \frac{y + rx + zc_1}{1+z} \tag{4.3.4}$$

This is the main channel of contagion. Whether it leads bank D to become insolvent or bankrupt, depends on the parameter values. For bank D to remain solvent, the following must be true

$$b(\gamma) + z\bar{q}^A \geq zc_1 \tag{4.3.5}$$

This condition states that bank D's buffer and value of deposits at bank A should be enough to pay off bank C's deposits held at bank D. In conclusion, for bankruptcy of bank A to spillover to bank D in this particular network, both conditions 4.3.3 and 4.3.5 should be violated.

4.4 Multiple-edge network: stochastic return case

In this section, I allow asset return to be stochastic, but assume that there is no aggregate uncertainty in asset return unless in case of bankruptcy in a given region. The asset return shocks follow the description in Table 9 and Table 10. In order to compare the fixed and stochastic asset return cases, I assume that expected asset return *i.e.* $(R_H + R_L)/2$ in the stochastic case equals the return R in fixed return case. Liquidity shocks are the same as before (Table 7).

Planner's solution With no aggregate uncertainty in liquidity demand shock or asset return shock, and with ex-ante identical regions, the planner's allocation is identical across regions. Then, with notation as before, we have the following feasibility constraints.

$$\begin{aligned} x + y &\leq 1 \quad \text{(date 0)} \\ \gamma c_1 &\leq y \quad \text{(date 1)} \\ (1-\gamma)c_2 &\leq (R_H + R_L)x/2 = Rx \quad \text{(date 2)} \end{aligned}$$

The planner's objective is to maximize ex-ante expected utility of households, which is given by:

$$\gamma u(c_1) + (1-\gamma)u(c_2)$$

Given the assumption that $(R_H + R_L)/2 = R$, the first best in fixed return case is exactly equal to the first best in stochastic return case. The incentive constraint is also satisfied. One key difference is that to achieve the first best, the planner would need to undertake resource redistribution at both date 1 and date 2. In particular, resources from high return regions would be transferred to regions with low return. I continue to use (x, y, c_1, c_2) to denote the first best allocation.

Decentralized Equilibrium As in the previous case, in a decentralized equilibrium, the first best allocation is not feasible. But now, since there are two types of shocks, banks engage in two different interbank markets to hedge against the corresponding shocks. Bank A holds $z = (\omega_H - \gamma)$ units of deposits in region B, bank B holds z units of deposit in bank C, and so on. In addition, bank A buys assets from bank B. Given the possible states of the economy at date 2, the return on assets bought by bank A (bank B assets) is perfectly negatively correlated with the return on bank A's original assets. This is a perfect hedge for bank A against idiosyncratic asset return shocks if bank A holds the following portfolio of assets at date 0, (y, x^A, x^B), y units of short assets,

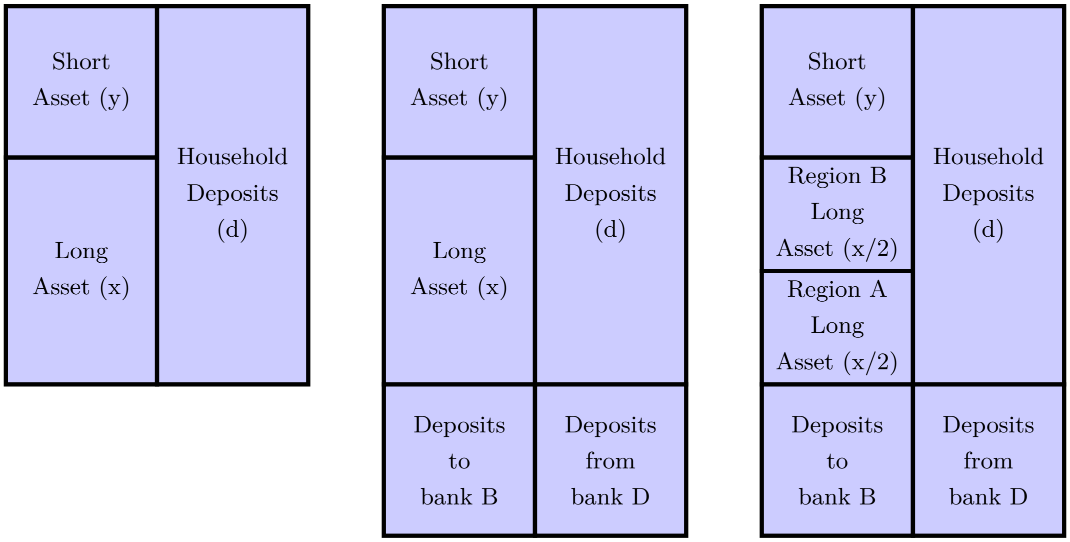


Figure 24: Balance sheet (i) no interbank market (ii) single interbank market (fixed asset return) (iii) two interbank markets (stochastic asset return)

$x^A = x/2$ units of assets originated in region A, and $x^B = x/2$ units of assets originated in region B. Similarly, the assets of bank D are given by (y, x^D, x^A). A typical balance sheet after interbank trades described here, is given in Figure 24.

The date 0 interbank purchase and sale of ex-ante identical assets that is described above, leads to a date 0 expected return on long assets for each bank to be equal to $(R_H + R_L)/2 = R$,[35] irrespective of the state of nature revealed at date 2. The cross holding of deposits is exactly as before i.e. $z = \omega_H - \gamma$, and the equilibrium coincides with planner's solution.

Aggregate uncertainty in liquidity demand shock Now suppose that there is aggregate uncertainty in liquidity demand shock as per Table 8. $\bar{S}$ has zero probability at date 0. But at date 1, if $\bar{S}$ is realized, bank A would need to liquidate some its long assets in order to meed the excess liquidity demand. Bank A has two long assets, x^A and x^B, which are *ex-ante* identical. So bank A will liquidate both these assets in equal proportion. With probability 1/2, the return on (x^A, x^B) would be (R_H, R_L), and with probability 1/2, the return would be (R_H, R_L); the expected return on long asset is still R. Liquidation value r is assumed to be the independent of region of origin of

[35]Recall that asset return vector across the four regions can take two values with probability 1/2 each.

an asset. Given a liquidity demand ω at date 1, the buffer function is as follows:[36]

$$\begin{aligned} \tilde{b}(\omega) &= r\left[x^A - \frac{(1-\omega)c_1}{2R}\right] + r\left[x^B - \frac{(1-\omega)c_1}{2R}\right] \\ &= r\left[x - \frac{(1-\omega)c_1}{R}\right] \end{aligned} \tag{4.4.1}$$

For bank A to be solvent, the necessary condition is

$$\tilde{b}(\gamma + \epsilon) \geq \epsilon c_1 \tag{4.4.2}$$

Now suppose that the ϵ shock is large enough to violate Equation 4.4.2 and cause bank A's bankruptcy. This alters the return on assets originated by bank A as per Table 10, where expected return on assets originated by bank A is lower. Since bank D holds assets originated in region A, this downgrade has implications for the portfolio and solvency of bank D.

At date 1, Bank D has a portfolio of long assets as (x^D, x^A). For this portfolio, the state contingent returns at date 2 are $(R_L, R_H - \alpha)$ in state N_1^A and $(R_H, R_L - \alpha)$ in state N_2^A. The decision of liquidation of long assets in order to meet the unexpected liquidity demand at date 1 is very different for bank D compared with that of bank A. Since expected liquidation cost of x^A, $(R - \alpha)/r$ is less than the expected liquidation cost of x^D, R/r, x^D will not be liquidated before x^A. But since the two long assets together also provide a hedge against return shock, depending on model parameters, it is possible that the assets are liquidated in equal amounts. These observations lead to three possibilities in which assets may be liquidated **without inducing a run on the bank**. Recall that liquidation must be such that late customers are guaranteed at least c_1 unit of consumption at date 2.

Case (i) Only a fraction of x^A can be liquidated at date 1

Suppose $l^A \in (0, x^A]$ units of x^A can be liquidated. x^D is not liquidated, so $l^D = 0$. The resources that will be available at date 2 are:

$$x^D R_L + (x^A - l^A)(R_H - \alpha) \quad \text{(in state } N_1^A\text{)}$$
$$x^D R_H + (x^A - l^A)(R_L - \alpha) \quad \text{(in state } N_2^A\text{)}$$

[36]For this case, the expression is identical to the one for $b(\omega)$, but will be different for later cases.

In order to prevent a run, the resources in either state must be enough to guarantee provision of c_1 units of consumption to late households. Since resources in state N_1^A are always less than resources in state N_2^A, it must be true that:

$$x^D R_L + (x^A - l^A)(R_H - \alpha) \geq (1 - \omega)c_1$$

Alternatively, the expected utility for a late customer by waiting should be higher than the utility by withdrawing c_1 immediately. With equi-probable states at date 2, this condition translates to the following condition:

$$u((x^D R_L + (x^A - l^A)(R_H - \alpha))/(1 - \omega))/2 + u((x^D R_H + (x^A - l^A)(R_L - \alpha))/(1 - \omega))/2 \geq u(c_1)$$

I adopt the first formulation for mathematical convenience. Since buffer corresponds to maximum possible liquidation, I pin down l^A as follows:

$$\begin{aligned}
& l^A = \max\Big\{l > 0 | x^D R_L + (x^A - l)(R_H - \alpha) \geq (1 - \omega)c_1\Big\} \\
\implies \quad & l^A : x^D R_L + (x^A - l^A)(R_H - \alpha) = (1 - \omega)c_1 \\
\implies \quad & x^D R_L < (1 - \omega)c_1
\end{aligned}$$

The last implication simply means that for case(i), in either state of the economy at date 1, x^D by itself is not enough to serve late households, and hence a fraction of x^A must be be liquidated. Finally, the buffer function $\tilde{b_1}(w)$ is given as follows:

$$\begin{aligned}
\tilde{b_1}(w) &= rl^A + rl^D \\
&= r\Big[\frac{x^D R_L + x^A(R_H - \alpha)}{(R_H - \alpha)} - \frac{(1 - \omega)c_1}{R_H - \alpha}\Big] \\
&= r\Big[x\frac{R - \alpha/2}{R_H - \alpha} - \frac{(1 - \omega)c_1}{R_H - \alpha}\Big]
\end{aligned} \tag{4.4.3}$$

(recall that $x^A = x^D = x/2$ and $R_H + R_L = R$)

Case (ii) All of x^A and some of x^D can be liquidated at date 1

Suppose $l^D \in [0, x^D]$ units can be liquidated. x^A is fully liquidated, so $l^A = x^A$. The resources that will be available at date 2 are:

$$(x^D - l^D)R_L \quad \text{(in state } N_1^A\text{)}$$
$$(x^D - l^D)R_H \quad \text{(in state } N_2^A\text{)}$$

In order to prevent a run, since resources in state N_1^A are always less than resources in state N_2^A, it must be true that:

$$(x^D - l^D)R_L \geq (1-\omega)c_1$$

To pin down largest possible liquidation l^D without inducing a run, I use:

$$(x^D - l^D)R_L = (1-\omega)c_1$$

The corresponding buffer function $\tilde{b_2}(w)$ is given as:

$$\begin{aligned}
\tilde{b_2}(w) &= rl^A + rl^D \\
&= rx^A + r\left[x^D - \frac{(1-\omega)c_1}{R_L}\right] \\
&= r\left[x - \frac{(1-\omega)c_1}{R_L}\right]
\end{aligned} \tag{4.4.4}$$

Case (iii) An equal fraction of x^A and x^D are liquidated at date 1

Suppose $l^A \in (0, x^A]$ and $l^D \in (0, x^D]$ are the liquidation amounts, with $l^A = l^D = l$, *say*. The resources that will be available at date 2 are:

$$(x^D - l^D)R_L + (x^A - l^A)(R_H - \alpha) \quad \text{(in state } N_1^A\text{)}$$
$$(x^D - l^D)R_H + (x^A - l^A)(R_L - \alpha) \quad \text{(in state } N_2^A\text{)}$$

Such a portfolio would be risk-less, and would give a sure return of:

$$(x/2 - l)(R_L + R_H - \alpha)$$

Then to prevent a run, the following must be true that:

$$(x/2 - l)(R_L + R_H - \alpha) \geq (1-\omega)c_1$$
$$\implies l = x/2 - \frac{(1-\omega)c_1}{2R - \alpha}$$

The latter equation pins down l, and the buffer function $\tilde{b_3}(w)$ is given by:

$$\begin{aligned} \tilde{b_3}(w) &= rl^A + rl^D \\ &= \left[x - \frac{(1-\omega)c_1}{R - \alpha/2}\right] \end{aligned} \tag{4.4.5}$$

Comparison of buffer functions The particular case of liquidation that is applicable depends on the parameter values. I begin with a comparison of the buffer function in the fixed asset return case to the buffer functions in three cases discussed above (where asset return is stochastic).

Proposition 10. *Suppose the downgrade α is small enough to satisfy $\alpha < (R_H - R_L)$. Then for all $\omega \in (0,1)$, the buffer value in fixed asset return case is higher than the buffer value in any stochastic asset return case. In other words, for all $\omega \in (0,1), b(w) > \tilde{b_i}(\omega)$ for $i = 1, 2, 3$.*

Proof. The proof is trivial for buffer functions in cases (ii) and (iii) above. For case (i), I use the condition on α, and the fact that in the first best, $x = (1-\gamma)c_1$. ∎

In analyzing the liquidation cases under parameter condition, $\alpha < (R_H - R_L)$, liquidation as per case (iii) will not occur since for all $\omega \in (0,1), \tilde{b_2}(\omega) > \tilde{b_3}(\omega)$. That is, since liquidation as per case (ii) always gives as much as liquidation as per case (iii), case (iii) will not be used at all. So the only possible liquidation cases are (i) and (ii). The case that arises depends on the following conditions. If $\tilde{b_2}(\gamma) \geq \tilde{b_1}(\gamma)$, then liquidation case (i) prevails and only a fraction of x^A can be liquidated. Else, if $\tilde{b_1}(\gamma) < \tilde{b_2}(\gamma)$, then liquidation case (ii) prevails and all of x^A and a fraction of x^D can be liquidated.

Finally, to pin down the contagion mechanism, consider a scenario where bank A is bankrupt.[37] As a result of this, all deposit cross holdings are liquidated. In particular, bank D liquidates deposits

[37] Note that the criteria for bankruptcy of bank A does not vary across fixed and stochastic asset return cases. This is critical because I want to focus on how the contagion varies across the two cases, and not the overall financial fragility of the system.

held at bank A, and bank C liquidates deposits held at bank D. In order for bank D to remain solvent at this stage, the following condition must hold, which states that bank D should be able to remunerate bank C using its buffer and liquidation of deposits at bank A, where the latter is valued at the maximum value of deposits at bank A, $\bar{q}^A$:

$$max(\tilde{b_1}(\gamma), \tilde{b_2}(\gamma)) + z\bar{q}^A \geq zc_1$$

Assuming bank B is solvent and the value of its deposits is $q^B = c_1$, $\bar{q}^A$ is given as follows:

$$\bar{q}^A = \frac{y + rx + zc_1}{1 + z}$$

Since $z = \omega_H - \gamma$, the condition for bank D to be solvent is given by:

$$\max\left(\tilde{b_1}(\gamma), \tilde{b_2}(\gamma)\right) \geq (\omega_H - \gamma)\left(c_1 - \frac{y + rx + (\omega_H - \gamma)c_1}{1 + \omega_H - \gamma}\right) \tag{4.4.6}$$

This is the condition for bank D's solvency. If this condition is violated, bank D goes bankrupt, and contagion occurs. Using the proposition above where I proved that $b(\gamma) > \max\left(\tilde{b_1}(\gamma), \tilde{b_2}(\gamma)\right)$, the corollary follows.

Corollary 1. *Suppose the downgrade α is small enough to satisfy $\alpha < (R_H - R_L)$. Then given liquidity demand shock parameters ω_H, ω_L, contagion is easier in multiple-edge interbank network relative to single-edge interbank network.*

This is the key result of this paper, which shows that under some conditions, contagion is easier in a multiple-edge interbank model, compared to a single-edge interbank model. Although the result is conditional on parameter values ($\alpha < (R_H - R_L)$), the condition itself is not stringent. The condition rules out large downgrades relative to the asset return possibilities. Nevertheless, if this condition is violated, the buffer in case (i) becomes larger relative to single-edge network, whereas the buffer in other cases becomes smaller. This means that when case (i) holds, contagion becomes weaker in a multiple-edge interbank network model.

5 Conclusion

In studying the financial contagion problem, I extend the model in Allen and Gale [2000] to allow for more than one edge between two banks in a network. In a single-edge incomplete network *à la*

Allen and Gale [2000] where contagion is possible, I show that considering a multiple-edge network has important implications for contagion.

In particular, I derive conditions under which contagion becomes easier in moving from a single-edge network to a multiple-edge network. Although this result is conditional on the parameterization, the condition itself is not stringent - it rules out large relative downgrades in asset return. I show that even if this condition is not satisfied, the result holds for two of the three liquidation cases. For the remaining case, contagion becomes weaker. The main insight is that single-edge networks behave differently compared to multiple-edge networks, and the implications for financial contagion in these networks can be qualitatively opposite. The model can be extended to endogenize the asset liquidation value r and include partial revelation of asset return uncertainty at date 1. In general, the model can be embedded in standard business cycle models to study the Macro-prudential policies in the presence of network externalities.

Chapter 3

Countercyclical capital regulation: Rules versus Discretion (with Isha Agarwal)

1 Introduction

A time invariant risk-weighted capital-ratio requirement (RWCR) has pro-cyclical effects on the economy through the bank lending channel, that is, RWCR reinforces business cycle fluctuations (see, for example, Repullo and Suarez [2013]). During business cycle upturns, estimated risk weight of bank assets is lower, resulting in lower capital charge and subsequently lower incidence of RWCR. This allows banks to take on additional leverage, resulting in increased bank lending. On the contrary, during business cycle downturns, estimated risk weight of bank assets is higher, resulting in higher capital charge and subsequent tightening of RWCR. This forces banks to de-leverage, resulting in lower bank lending. Even a time invariant non-risk-weighted capital-ratio requirement, henceforth referred to as the leverage-ratio requirement (LR), has pro-cyclical effects on the economy. During business cycle upturns, increasing asset prices and lower credit losses result in higher bank leverage-ratios and subsequently lower incidence of the LR. This allows banks to take on additional leverage, resulting in increased bank lending. Similarly, during business cycle downturns, falling asset prices and higher credit losses result in lower bank leverage-ratios and subsequent tightening of the LR. This forces banks to de-leverage, resulting in lower bank lending.

To deal with these issues, countercyclical capital-ratio requirement (CCR) has been proposed by the Basel Committee for Banking Supervision (BCBS) as part of the Basel III accords. CCR is currently in the process of being adopted by central banks / regulatory authorities around the world, for example, in US by the Federal Reserve (see FRB [2016]):

> *"The proposed Policy Statement describes a set of principles for translating judgmental assessments of financial-system vulnerabilities into specific levels of the countercyclical capital requirements, a set of empirical models used as inputs to the judgmental process that distill and translate quantitative indicators of financial and economic performance into potential settings for the CCR ... "* (US Federal Register, February 3, 2016)

The objective behind imposing the CCR is twofold. First is to avoid the creation of credit bubbles by imposing higher capital-ratio requirement on banks when there are indications of build-up of excessive credit in the economy. Higher capital-ratio requirement in periods of excessive credit growth also enable banks to build enough capital to sustain future credit losses.[38] Second objective of the CCR is to impose lower capital-ratio requirement during business cycle downturns so that the capital built during business cycle upturns can be released and be used to absorb credit losses during business cycle downturns when the incidence of credit losses is higher.

There are several papers that assess the implications of CCR in a macroeconomic setting. Benes and Kumhof [2011] model capital requirement similar to the Taylor rule in the New-Keynesian framework. They show that a time varying capital-ratio requirement that responds to loan gap is welfare improving relative to a fixed capital-ratio requirement. Christensen et al. [2011], who study interaction of monetary and macro-prudential policy in a New-Keynesian framework with financial friction due to information asymmetry, show that the benefits of a countercyclical capital-ratio requirement depends on the contemporaneous stance of monetary policy. Zhu [2007] shows that risk-weighted capital requirements are welfare improving. Although, he emphasizes that the capital buffer banks maintain for operational efficiency offsets the counter-cyclicality of capital regulations. Repullo and Saurina Salas [2011] assess the current Basel III CCR proposal, and argue that in order for CCR to address pro-cyclicality, it must be pegged to GDP growth and not to credit-GDP growth. Their intuition is that in periods of high GDP growth relative to credit growth (which is a business cycle upturn) the credit-GDP based CCR would be lower, and therefore, would exacerbate the pro-cyclical effect of capital-ratio requirement on the business cycle. They propose alternative

[38] With regards to indicators of build-up of excessive credit in the economy, deviation of the credit-GDP ratio from its long term trend has been shown in Drehmann et al. [2010] and in Drehmann et al. [2011] to be the best performing leading statistic, and in this sense, a reliable measure of the *financial cycle*. Bonfim and Monteiro [2013] support this conclusion, but also suggest that a broader set of macroeconomic indicators must be taken into consideration when implementing CCR.

GDP-growth based CCR rules. In summary, though the above discussed strand of literature seems to generate a consensus that some form of CCR must be used to address the pro-cyclicality of time-invariant capital-ratio requirement, and that CCR can reduce the variance of investment and output, and increase welfare, the implementability of CCR is a question that remains un-addressed.

In this paper, we take up the question whether CCR should be implemented strictly as a rule under commitment, or whether regulators should have discretion with respect to the timing and magnitude of changes in CCR along the financial cycle. We believe that this question is important because CCR implementation is subject to the following trade-off which has welfare implications. On one hand, discretionary policy enables the regulator to respond to unexpected shocks, but on the other hand, relative to a rule-based policy, it increases policy uncertainty which can induce precautionary and inefficient behavior by banks. We prove that this trade-off arises in a simple three-period economy as well as a finite-horizon economy due to information asymmetry and unexpected shocks. The motivation behind information asymmetry is that policy makers have access to confidential bank balance sheet and interbank transaction data, and they conduct stress tests, both of which provide them with more information about the build-up of credit and systemic risk - something individual banks do not have access to. The motivation behind unexpected shocks is guided by the limited ability of the policy maker (and private agents in general) to foresee all future contingencies.

Intuitively, the result that information asymmetry leads to higher policy uncertainty for banks as measured by the variance of future CCR in a discretionary regime relative to a rule-based regime, is proved as follows. In a rule-based regime, the state-contingent capital-ratio requirement (CR) is known to the bank, so the variance in future CR is limited to the variance in future states of the world. In a discretionary regime, the regulator solves for CR at the beginning of each period after that period's uncertainty is resolved. Therefore, the banks do not know future CR exactly. In order to estimate future CR (so that current period decisions comply with future CR), the banks need to solve a *version* of the regulator's problem in the next period. However, due to information asymmetry, banks cannot solve the *true version* of the planner's problem — they can only solve for a distribution of the possible levels of CR in each future state.

The paper is organized as follows. In section 2, we discuss and compare the two possible implementations of CCR. In particular, we discuss the literature on time-inconsistency, and document how the same arises in the context of rule-based CCR. We then discuss the policy uncertainty literature, and assess how discretionary CCR can lead to an increase in policy uncertainty. In section

3, we propose a three-period model and prove the existence of increase in policy uncertainty in discretionary regime. We also propose a finite-horizon model based on the insights from three-period model, integrated with the notion of financial frictions. Section 4 summarizes this paper.

2 Rules versus Discretion

Broadly, two possible implementations of CCR are assessed. First is the *rule-based* CCR where at date zero, the regulator commits to a (possibly state-contingent) rule that governs the capital-ratio requirement for banks in all foreseeable future states of the world. Second is the *discretionary* CCR where the regulator cannot commit to any rule at date zero, and instead, solves for the capital-ratio requirement at each date. We begin with a discussion of the time-inconsistency and policy uncertainty issues while focusing on CCR.

2.1 Time inconsistency

The time inconsistency problem arises when a principal makes a promise to induce expectation in an agent in order to elicit certain behavior by the agent, but once that behavior is elicited, the principal finds it optimal to renege on the promise. The issue of time-inconsistency of macroeconomic policy rules has been studied extensively, starting from the seminal paper by Kydland and Prescott [1977] who showed that a policy rule can either be optimal, or time consistent, but not both. Their intuition is that since there is no mechanism to induce future policy makers to care about the effect of their policy on current decision of agents via the expectation mechanism, the optimal rule is time inconsistent. In other words, if there are no commitment devices, the optimal rule will not work as intended as agents will know the incentive of future policy makers to deviate from the optimal rule, and will adjust their expectations accordingly. However, the time inconsistency issue is resolved if there are commitment devices, like *loss of reputation or strong institutions*, precisely because policy making is a repeated game.

In the context of monetary policy, several papers including Kydland and Prescott [1977] have shown that a rule that promises low inflation is time inconsistent, because the central bank has an incentive in future to inflate the economy at a rate greater than the promised rate. Intuitively, the promise of low inflation "buys" the regulator expected inflation $\pi^e = 0$, which it can use to obtain $\pi - \pi^e > 0$ by reneging on the promise of low inflation ex-post. Regulator wants to renege since reducing unemployment is part of the objective function, and because the Phillips curve implies that surprise inflation can bring down unemployment. However, if the central bank fears that deviating from the promise can jeopardize its reputation for a long time into the future, the central bank might not renege on its promise. Barro and Gordon [1983] formalize this insight by imposing a cost of reneging by the policy maker.

In the context of macroprudential policy in general, Bianchi and Mendoza [2015] show that the optimal policy rule (under commitment) for tax on debt is time inconsistent, and find the optimal time consistent policy by solving the recursive problem of a policy maker, taking as given future policies. They show that during a period with negative shock, the regulator, in order to increase current asset price and cause the relaxation of collateral constrain, promises low future consumption to increase future marginal propensity of consumption. However, ex-post, reneging is optimal since higher consumption serves to increase welfare. The time inconsistency arises because once the promise of low future consumption has been used to increase current prices through the expectation channel, the promise is not useful any more, and in fact, it is costly to fulfill the same.

In the context of capital-ratio requirement in particular, Kowalik [2011] documents the following time inconsistency problem. The regulator, after several periods of positive growth, might lower the estimate of the probability of a crisis in the near future, and be tempted to not increase capital-ratio requirement as per the preset rule. Alternatively, the regulator might face pressure from politicians to focus on short term growth, and therefore it might not be able to increase the capital-ratio requirement as per the preset rule. Another reason why the time-inconsistency problem might arise is because bank lobbyists might pressure to delay the increase in the capital-ratio requirement dictated by the preset rule. The commonality between these cases that engenders the time-inconsistency issue is an ex-post possibility that is not built into the rules. In the first case, time inconsistency arises due to the possibility of Bayesian updating of the Markov transition probabilities that lowers the probability of a crisis after several periods of positive growth. In the second and third cases, the time inconsistency arises because unexpected, unanticipated aggregate phenomenon can possibly hit the system, and because rules do not internalize the effect of these phenomenon. Of course, as we discussed earlier, if there are commitment devices available, then the rules will be immune to these considerations, and the time inconsistency problem will not bite.

In this paper, we document another time inconsistency issue with respect to rule-based CCR. Consider an economy that is in the expansionary or build-up phase of the financial cycle. The policy rule would dictate banks to start building up capital, such that the higher capital-ratio requirements are met within one year of this announcement (this is similar to the actual CCR implementation proposed by the Basel Committee for Banking Supervision (BCBS)). Given that banks have considerable time to build their capital-ratios, they prefer doing so by issuing new equity as opposed to doing so by reducing leverage.[39] But then, at the end of one year when banks have raised equity and have increased their capital-ratios to meet the regulatory minimum, the regulator might be tempted to bring down the CCR so as to spur bank lending and growth. The overall process which includes deviating from the rule (and therefore constituting the time inconsistency issue) results in the regulator achieving the following two objectives. First, due to higher CCR ex-ante, the banks build-up capital and have strong balance sheets, so bank failures have low probability. Second, due to lower CCR ex-post banks are left with excess capital relative to the minimum required capital-ratios, giving them leeway to engage in lending, and possibly spur growth. In other words, the promise of higher capital requirement in future "buys" the regulator stronger bank balance sheets. The reneging of the promise buys the regulator possibility of increased lending and output.

Even though rule-based CCR is subject to the time-inconsistency problem, it does not necessarily bite in a way it does in case of monetary policy because banks cannot act strategically given the incentive of a regulator to deviate from the rule. In other words, even if the bank knows that the regulator can deviate from the preset rule, it cannot choose to act differently because actions of the bank are ex-post verifiable and punishable. This is in contrast with the issue of time-inconsistency in monetary policy, where private agents (banks) can change their expectations without punishment given the policy maker's incentive to cheat.

To summarize, there are several reasons behind why a rule-based CCR can be time-inconsistent. We argued why time-inconsistency of CCR does not necessarily bite in a way it does in the case of monetary policy. Nevertheless, we will assume that the policy maker cares for loss of reputation which acts as the commitment device that ensures that the policy rule is implemented as intended. In other words we assume that the policy maker can commit to a CCR rule in the *rule-based* regime. We now review the literature on policy uncertainty, and discuss how a *discretionary* CCR regime can lead to an increase in policy uncertainty.

[39] Among other papers, Choe et al. [1993] and Covas and Haan [2011] have shown that the pro-cyclicality of equity issuance empirically as well as theoretically. Intuitively, raising equity during expansions is easier because bank managers can justify the need for fresh equity, and also because lower credit losses during expansions reduce the perceived riskiness of banks.

2.2 Policy uncertainty

Policy uncertainty may be defined as the variance in future policy path. The policy uncertainty literature focuses on the following key aspects. First is an understanding of the origins of policy uncertainty, like upcoming elections, change in government, information asymmetry, and unexpected macroeconomic shocks. Second is the study of effect of policy uncertainty on the decision making of private agents. We jointly review these aspects below.

Elections:– Giavazzi and McMahon [2012] show using German data that policy uncertainty arising in the run-up-to national elections can significantly increase household savings due to precautionary motives (for example, see Carroll and Kimball [2007]). The increase in policy uncertainty can stem from the possibility of reversal of reforms post election. Disagreement on policies between political parties can also lead to policy uncertainty as private agents cannot be certain of the policy directive, and this phenomenon can become more pronounced in the run-up-to an election when the incumbent party is keen on pushing on more number of reforms. In a similar study, Julio and Yook [2012] show that corporate investment cycle is closely associated with timing of national elections across several nations. The paper provides evidence in support of the hypothesis that political uncertainty leads firms to wait until resolution of political uncertainty before taking up hiring or investment. As discussed in Bernanke [1983] and in Pindyck [1988], the channel through which uncertainty in general affects firm decisions, is more pronounced if investments are irreversible or if hiring and firing is costly. The trade-off private agents face is between committing to an investment today in order to generate returns earlier, and waiting for more information about the return on investment.

Change in government:– Closely related with how elections can lead to policy uncertainty, change in government can also lead to policy uncertainty as new governments might take time to learn about the macro-economy and formulate policies, or when preferences of new government are not known to the private agents. Pastor and Veronesi [2012] and Pástor and Veronesi [2013]) formalize this notion as *political uncertainty: what will the government do*, and contrast it with the notion of *impact uncertainty: what is the impact of a new policy.* The impact uncertainty arises in their model because firm profit follows a stochastic process whose mean is affected by the policy, and this effect is unknown to firms, something they learn over time using Bayesian updating. In a developing economy context, Rodrik [1991] assesses the impact of the possibility of reform reversal on capital flight. He assumes that the policy reform can increase the return on capital, but that there is a fixed probability that the reform will be reversed.

Information uncertainty:– Another important source of policy uncertainty is information asymmetry, wherein, private agents might not have access to the kind of information the policy maker has. For example, financial regulators conduct several stress tests on banks, giving the regulator informational advantage. It also might have unique access to bank data, for example, interbank transaction data in the US (Fed-wire data).

Unexpected shocks:– If a policy maker is exposed to an unexpected or unforeseen macroeconomic situation, he/she might need time to assess the situation and take the right policy action. This might include debating in the parliament and formal analysis of the situation. It is also possible that the right policy action is not determinable or is not unique, leading to policy uncertainty when inaction is not an option. This view has gained momentum after the recent financial crisis that posed new challenges for monetary, fiscal and regulatory authorities. For instance, FOMC [2009] and IMF [2012] suggest that policy uncertainty about health-care, tax, finance and environment contributed a great deal to the *wait-and-watch* approach followed by several businesses, which delayed recovery after the financial crisis. Fernandez-Villaverde et al. [2015] test this narrative and show that there exists considerable amount of tax and government spending volatility in the US. They extend the macroeconomic model in Christiano et al. [2005] to include policy uncertainty as time varying variance in the noise in government policy rule. After a two standard deviation shock to the volatility of capital income tax, key macroeconomic variables remain low for several quarters. On the other hand, using a similar DSGE model where there is time varying variance of noise in tax, spending or monetary policy, Born and Pfeifer [2014] show that output loss due to policy uncertainty is small. Their intuition is that the wait-and-watch effect of policy uncertainty is counteracted to some extent by the precautionary effect which induces private agents to work more and build buffer capital stock.

There are several contexts in which policy uncertainty can arise due to one or more of the above sources. Hassett and Metcalf [1999] study the effect of uncertainty in investment-tax-credit policy on the timing and level of investment by firms. Rodrik [1991] studies the effect of uncertainty about success of structural reforms on the level and timing of investment – he shows that the level drops and the timing is delayed. In similar vein, Handley and Limao [2012] model policy uncertainty as a probability that the policy maker is able to change the policy based on current state variables, whereas with complementary probability, the policy remains unchanged. In studying the impact of this kind of trade policy on the decision of firms to invest in exports, they show that with sunk export costs that render export investment irreversible, a firm's export investment is lower when trade policy uncertainty is higher. Likewise, Born and Pfeifer [2014] and Fernandez-Villaverde et al. [2015] study policy uncertainty in the context of government taxes, spending, and monetary policy.

In the context of monetary policy, Christiano et al. [1999] document three reasons why there might be exogenous shocks to monetary policy that contribute to policy uncertainty. First is the shock to regulator's preference, which might emanate from a change in government, or from a change in welfare weight of certain types of households. Second is the regulator's desire to avoid social costs of disappointing private agents' expectations. For example, if most of the economy feels an interest rate cut is called for, it might be socially very costly for the Fed to not cut interest rate and de-rationalize the expectation of private agents which can lead them to forgo planned investment opportunities. Third is the measurement error in preliminary data that is used to inform policy decisions.

In the context of discretionary implementation of countercyclical capital-ratio requirements, information asymmetry and unexpected shocks can contribute to policy uncertainty. Under discretion, the regulator imposes higher capital-ratio requirement when it feels that there is a need to stem the build-up of excessive credit and systemic risk in the economy, or if it feels that the economy is comfortably out of a recession and that it needs to start creating buffer for a future shock. However, both these rationales for activating higher capital-ratio requirements can contribute to policy uncertainty. First, the regulator might have more information than individuals banks about the build-up of systemic risk. Such information asymmetry can arise because regulators have access to private and confidential information of individual financial institutions, and they might be able to aggregate this information across institutions to form conjectures about the build-up of excessive credit or risk in the system, something individual private agents / institutions cannot pursue. Second, it is also possible that the regulator has preferences that are not perfectly known to the banks – for example, the regulator might have a bias against recessions and it might be willing to forgo credit creation and output in order to bring down the probability of future recessions. Third, should unforeseen shocks hit the economy, the regulator might not be able to respond contemporaneously as it might be unable to come up with the right policy response. For example, consider a situation when the credit-GDP ratio is falling, and the GDP is increasing. If the regulator is focusing on the former indicator of build-up of credit, it might be tempted to not activate CCR, but then, it would render the CCR policy pro-cyclical, which defeats the purpose of CCR. The right policy action in this situation is debatable, resulting in policy uncertainty. To summarize, there are various sources of policy uncertainty in discretionary CCR, and in the rules versus discretion debate with regards to CCR, these must be factored in.

2.3 Rules versus Discretion

In this section, we organize the debate on whether CCR must be rule-based or discretionary. In assessing this question, we draw insights from the discussion in the previous sections to make three assumptions. First is the existence of commitment devices that mitigate any time-inconsistency issue associated with a rule-based CCR regime. Second is the existence of information asymmetry between regulator and banks, which is the source of policy uncertainty in a discretionary CCR regime. Third is the existence of unexpected aggregate shocks.

On one hand, rules are desirable because they facilitate expectation formation by banks about future capital requirements. Also, since rules specify a formula to compute capital requirement based on publicly available information, changes in capital requirement along the business or financial cycle do not reveal new information – in a discretionary CCR regime, increase of capital requirements might signal build-up of excessive credit in the economy. In this sense, rules leave little room for misinterpretation of, or negative reaction to, CCR announcements. Rules also eliminate much potential for policy uncertainty which can lead to underinvestment and inefficient risk taking by banks.

In contrast, discretion is desirable as it allows the regulator to respond to new information that becomes available over time, information that is not available at the time for formulating the (state contingent) rules. For example, shocks like the housing bubble burst and its subsequent amplification may not be incorporated into the rules as these shocks are either unanticipated or their unraveling process is too complex to be fully known at the time of formulating the rules. This explains why the 2004 Basel II accord failed to incorporate the possibility of the 2008 financial crisis, and was inadequate to address the crisis. If the regulator had the discretion to temporarily increase capital requirements based on higher credit-GDP ratios in the run up to the crisis, the 2008 financial crisis could have been a smaller recession. It is in this sense that rules are undesirable as they tie the hands of the regulator in response to unexpected shocks.

Nevertheless, discretionary CCR entails to policy uncertainty. It can create noise about the future path of regulation, which can induce banks to behave in a precautionary manner i.e. take less leverage and less risk relative to what is optimal. Unexpected announcements about higher capital requirement when the regulator feels that there is excessive build-up of credit, can induce

$$\Pi = \begin{bmatrix} \pi_L & 1-\pi_L \\ 1-\pi_H & \pi_H \end{bmatrix}$$

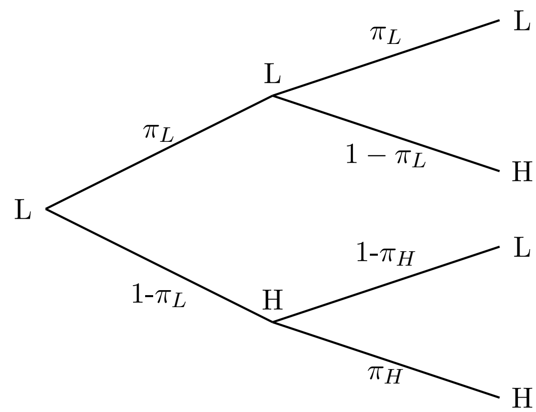


Figure 25: Aggregate Uncertainty

investors to deduce increase in risk in the economy, and can make it difficult for banks to raise capital by issuing equity. As a result, banks may be forced to increase capital ratios by reducing lending and leverage, which could actually initiate a downturn.

Overall, there is a strong case for rules but also for discretion. With discretionary CCR *a'la* Basel III still in the process of being phased in, there is not enough data to empirically assess its success and its limitations. This debate reminds us how a complex policy like the advanced internal review based (IRB) approach to capital regulation failed to signal the build-up of systemic risk before the 2008 financial crisis. And, it reminds us of the need for simple rules like backstop leverage ratio requirement which are not subject to model risk or measurement risk. In the next section, we propose a model that captures some of the above discussed trade-offs a regulator faces while debating rule-based CCR versus discretionary CCR.

3 Model

In developing the model, we focus on policy uncertainty in discretionary CCR regimes *viz-a-viz* inability of regulator to respond to unexpected shocks in rule-based regimes.

3.1 Main result in a simple setting

In this section, we develop a simple three period model to illustrate the above trade-off that a regulator faces while choosing rule-based *viz-a-viz* discretionary regulation. There are three dates 0,1,2, and there is aggregate uncertainty. The economy consists of the following agents: a representative household, a representative bank, a representative firm and a regulator. The household owns the bank and receives dividends in return. It saves via the deposit market. The bank uses deposits and net-worth (internal equity) to fund investment in the firm. Firm equity is held by the bank, and the bank receives firm dividend in return. The regulator sets state-contingent bank capital requirements (CCR) to maximize welfare.[40] CCR is modeled as a state-contingent lower bound on the capital-ratio of banks.

The household maximizes expected present value of the utility, subject to budget constraints, given initial wealth W_0, and bank dividends E_t. After date-2 the household dies, and therefore at date-2 it consumes everything and does not save for the future.

$$
\begin{aligned}
\max_{C_0,C_1,C_2,D_0,D_1} \; & u(C_0) + \beta \mathbb{E} u(C_1) + \beta^2 \mathbb{E} u(C_2) \\
C_0 + D_0 &= W_0 + E_0 \\
C_1 + D_1 &= R_0 D_0 + E_1 \\
C_2 &= R_1 D_1 + E_2
\end{aligned}
$$

The firm operates a simple production technology $Y = K$ that uses only capital. The firm pays dividend per unit of capital given by $Z = Y/K = 1$. The bank maximizes expected present value of the dividends it pays to the household, taking as given initial net-worth N_0.

$$
\begin{aligned}
\max_{E_0,E_1,E_2,S_0,S_1,D_0,D_1} \; & E_0 + \beta \mathbb{E} E_1 + \beta^2 \mathbb{E} E_2 \\
E_t + S_t = N_t + D_t \quad & t = 0,1; \quad E_2 = N_2 \\
(E_t + S_t) x_t \leq N_t \quad & t = 0,1 \\
N_t = (Z_t + 1 - \delta)\psi_t S_{t-1} & - R_{t-1} D_{t-1} \\
-\mathbf{1}\{t=1\}\chi\Big(\big(R_{t-1}D_{t-1} - (Z_t + 1 - \delta)\psi_t S_{t-1}(1 - x_t)\big)_+ \Big) & \quad t = 1,2
\end{aligned}
$$

The first constraint is the balance sheet constraint. The second constraint denotes the fact that at date-2, all net-worth is paid as dividend. The third constraint denotes the contemporaneous capital requirement imposed as a limit on debt as a multiple of net-worth i.e. ($D_t \leq N_t(1 - x_t)/x_t$). The fourth constraint denotes the evolution of net-worth based on the return on bank assets (equity holding in the firm), net of debt cost and regulatory cost χ. The return on bank assets depends on the stochastic component ψ. Evolution of net-worth is subject to future capital requirement, and is

[40] We abstract away from financial frictions in this version of the model for mathematical convenience. Financial frictions rationalize capital requirements, and will be introduced in the general model.

the key mechanism to induce current bank decisions to internalize future capital requirement. We assume that future regulation is imposed as a convex cost $\chi(.)$ if capital requirement is violated, and the cost is proportional to the degree of violation. This regulatory cost is motivated from dividend distribution constraints imposed on banks that violate capital-ratio requirements. Future capital requirement arises naturally in this model as it addresses bank failures at the beginning of next period. There are no capital requirements to be satisfied at date-2 when the bank pays all its net-worth as dividends, and there is no debt issued or investment made.

We assume that the aggregate state of the world that lies in (L, H), and that ψ takes values (ψ_L, ψ_H) in the two states respectively. The state transition probability matrix is given in figure 25. The intra-period time-line for the economy is given in figure 26. We consider two policy regimes depending on whether the CCR is implemented as rule-based or discretionary. To enable us to study the trade-offs between these regimes, recall that we made two important assumptions. First is the information asymmetry problem between policy maker and private agents. Second is the possibility of aggregate shocks that have zero probability at date-0. The motivation behind first assumption is that policy makers have access to confidential bank balance sheet and interbank transaction data, and they conduct stress tests, both of which give them more information about the build-up of credit and systemic risk - something individual banks do not have access to. A similar assumption in the context of monetary policy can be found in, for example, Stulz [1986] who models policy uncertainty by assuming that households are rational but they do not know the true dynamics of money supply. The motivation behind the second assumption is guided by the limited ability of the policy maker (and private agents in general) to foresee all future contingencies. An example is the 2008 financial crisis. The fact that there was a need to change financial regulation policies (Basel III) after the crisis implies that financial regulation policies existing at the time of the crisis did not internalize the possibility of a crisis. Another approach to motivating the *unexpected* aggregate shock assumption is the limited ability of a policy maker to foresee how the economy might unravel in response to not-unexpected shocks in the presence of amplification mechanisms like fire sale and network externalities. These two assumptions are cast into the model economy as follows.

Information asymmetry:– Information asymmetry is assumed in that the regulator knows the exact transition probabilities for ψ, while the household and the bank have unbiased priors (η_L, η_H), such that $\mathbb{E}\eta_L = \pi_L$, $\mathbb{E}\eta_H = \pi_H$. Suitable prior distributions in this case are given by the beta distributions $\beta(a_L, b_L)$ and $\beta(a_H, b_H)$.[41] The state of economy which is revealed at the beginning

[41] Another approach to model information asymmetry is to assume that both the central bank and the private

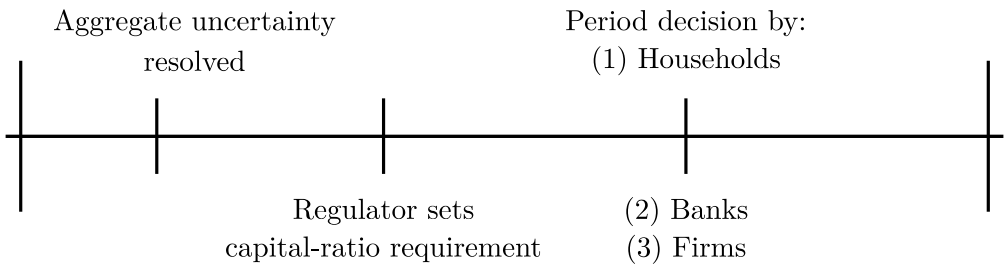


Figure 26: Timeline

of each period is assumed to be visible to both the regulator and the bank.

Unexpected shocks:– The possibility of an unexpected aggregate shock that is not known to agents at date-0, is modeled as the possibility of a third state at date-1. In particular, we assume a crisis state ψ_C which denotes lowest return on capital relative to other states.[42]

In light of these assumptions, in the rule-based regime, CCR is state contingent, and is solved for by the regulator based on the date-0 welfare maximization problem of the regulator. The rule-based CCR is not contingent on crisis state ψ_C, and the rule associated with ψ_L which is the closest to ψ_C, is assumed to be applicable in state ψ_C. Banks know the state contingent rule, and they know that the regulator is committed to the same. Banks form expectations about the states next period, and accordingly they form expectations about the rule applicable next period. Policy uncertainty in this case is limited to the uncertainty about the state that is realized in the next period.

In the discretionary regime, the regulator computes the policy for each period after observing the state revealed at the beginning of the period. Therefore, discretionary CCR can respond to ψ_C as opposed to rule-based CCR. However, in this case banks need to obtain estimates of CCR for all states of the economy in the next period. Given the information asymmetry problem, banks cannot exactly predict the state-contingent CCR for next period, which leads to an increase in the policy uncertainty. The uncertainty in level of future capital-ratio requirement matters because it enters current bank decision problem through the regulatory cost. In the rule based regime, banks know the exact future path of the state-contingent policy, and there is no need for them to estimate

agents infer the state of the economy by observing a signal, and that the signal has lower variance and is therefore more informative for the regulator. Alternatively, one can assume that the regulator's objective function is imprecisely known to the private agents e.g. parameter uncertainty in Blake and Zampolli [2011].

[42] The issue that unexpected shocks serve as a disadvantage of rules has been raised in various policy contexts. For example, William Dudley, President of Federal Reserve Bank of New York, mentioned in his October 3, 2015 speech that it is hard to anticipate episodes where the standard beliefs about what causes booms turn out to be false, and that such changes cannot be hardwired in a rule. For example, rules that are determined on the basis of the knowledge of the aggregate uncertainty process, can fail to anticipate the amplification that bad shocks can lead to.

future capital-ratio requirement. A consequence of this *relative* increase in policy uncertainty is that banks show precautionary behavior in a discretionary regime. We state and prove the policy uncertainty insight formally in the following proposition.

Proposition. *Under information asymmetry, the variance of future capital-ratio requirement for banks, conditional on the future state of the economy, is higher in a discretionary regime relative to the rule-based regime.*

Proof. We focus on the bank's decision problem at date-0. Without loss of generality, we assume that the date-0 state of the economy is L. In the rule based regime, the regulator commits to a declared state-contingent capital requirement $\bar{x}_L$ and $\bar{x}_H$ for states L and H respectively.[43] In this sense, the state-contingent capital requirement is known to the bank. The bank's priors about the state transition probabilities from state L at date t to state L and state H at date $t+1$ are given by random variables η_L and $1-\eta_L$ respectively. Therefore, policy uncertainty faced by the bank at date-0 in a rule-based regime can be stated as:

$$x_{1,rule} = \begin{cases} \bar{x}_L & w.p.\ \mathbb{E}\eta_L = \pi_L \\ \bar{x}_H & w.p.\ 1-\mathbb{E}\eta_L = 1-\pi_L \end{cases}$$

In a discretionary regime, the bank has same priors about state transition probabilities, but does not know the level(s) of capital requirement in the two states of the economy at date-1. The bank estimates the latter by solving an equivalent of the regulator's welfare maximization problem at each state in date-1. The regulator's welfare maximization problem is setup in a primal approach, which involves (a) establishing the implementability conditions from the competitive equilibrium for a given regulation, and (b) using the implementability conditions as constraints in the regulator's problem. We first establish conditions for the competitive equilibrium at date-1 in state L.[44] The bank's problem at date-1 in state L for a given regulation x_L is as follows, where the objective function is based on the expected prior of transition probabilities. N_1, and x_1 which is the capital-ratio requirement at date-1, are taken as given. We note that there are no capital-ratio requirements at date-2.

[43]We do not need to compute $\bar{x}_L$ and $\bar{x}_H$ for this proof.

[44]The choice of state L is without loss of generality because the analysis for state H follows similar steps.

$$\max_{S_1,D_1,E_1,E_{2L},E_{2H}} E_1 + \beta\pi_L E_{2L} + \beta(1-\pi_L)E_{2H}$$

$$E_1 + S_1 = N_1 + D_1$$

$$(E_1 + S_1)x_1 \leq N_1$$

$$E_2 j = N_2 j \quad j \in \{L, H\}$$

$$N_2 j = (Z + 1 - \delta)\psi_j S_1 - R_1 D_1 \quad j \in \{L, H\}$$

$$\implies E_1 = 0 \tag{3.1.1}$$

$$S_1 = N_1/x_L \tag{3.1.2}$$

Assuming that dividends cannot be negative, and that $\beta(Z+1-\delta)(\pi_L\psi_L + (1-\pi_L)\psi_H) > 1$, the solution to bank's problem is given by equations 3.1.1 and 3.1.2. The household problem at date-1 in state L is:

$$\max_{C_1,D_1,C_{2L},C_{2H}} u(C_1) + \beta\pi_L u(C_{2L}) + \beta(1-\pi_L)u(C_{2H})$$

$$C_1 + D_1 = R_0 D_0 + E_1$$

$$C_{2j} = R_1 D_1 + E_{2j} \quad \forall j \in \{L, H\}$$

$$\implies u'(C_1) = \beta R_1(\pi_L u'(R_1(R_0D_0 + E_1 - C_1) + E_{2L}) +$$

$$(1-\pi_L)u'(R_1(R_0D_0 + E_1 - C_1) + E_{2H})) \tag{3.1.3}$$

Equation 3.1.3 specifies the first-order condition for the household problem in terms of C_1 and R_1. Finally, the market clearing conditions are:

$$S_1 = K_1(1-\delta) + I_1, \quad \textit{Asset market} \tag{3.1.4}$$

$$Y_1 = K_1 = C_1 + I_1, \quad \textit{Goods market} \tag{3.1.5}$$

Equations 3.1.1 - 3.1.5 specify the equilibrium conditions for the competitive equilibrium in terms of C_1, R_1, E_1, S_1, I_1, which also constitute the implementability conditions. The version of planner's problem at date-1 in state L that the bank would solve to obtain $x_L(\eta_L)$ for each possible level of the transition probability prior η_L is given below. The bank takes N_1, K_1, R_0, D_0 as given. We define $R_k = (Z + 1 - \delta)$, and also simplify the implementability conditions:

$$\max_{x_L} u(C_1) + \beta\eta_L u(C_{2L}) + \beta(1-\eta_L)u(C_{2H})$$
$$s.t.\ \ C_1 + S_1 = K_1(Z+1-\delta),\ S_1 = N_1/x_L,\ E_1 = 0$$
$$u'(C_1) = \beta R_1\big(\pi_L u'(R_1(R_0 D_0 + E_1 - C_1) +$$
$$E_{2L}) + (1-\pi_L)u'(R_1(R_0 D_0 + E_1 - C_1) + E_{2H})\big),$$
$$where\ E_{2j} = R_k\psi_j S_1 - R_1 D_1, C_{2j} = R_1 D_1 + E_{2j}\ and\ D_1 = E_1 + S_1 - N_1 \quad j \in \{L, H\}$$

The problem can be further simplified to express the regulator's problem in terms of x_L, state variables and the parameters as follows:

$$\max_{x_L} u(K_1(2-\delta) - N_1/x_L) + \beta\eta_L u(R_k\psi_L N_1/x_L) + \beta(1-\eta_L)u(R_k\psi_H N_1/x_L)$$
$$\implies u'(K_1(2-\delta) - N_1/x_L) =$$
$$\beta\Big(\eta_L\psi_L u'(R_k\psi_L N_1/x_L) + (1-\eta_L)\psi_H u'(R_k\psi_H N_1/x_L)\Big) \tag{3.1.6}$$

In case of CRRA utility with parameter γ, the regulation is given as:

$$x_L(\eta_L) = \frac{N_1}{K_1(2-\delta)}\left(1 + \frac{\beta R_k^{\gamma-1/\gamma}}{[\eta_L\psi_L^{1-\gamma} + (1-\eta_L)\psi_H^{1-\gamma}]^{1/\gamma}}\right) \implies var(x_L(\eta_L)) > 0 \tag{3.1.7}$$

Equation 3.1.6 characterizes the solution of the planners problem solved by the bank based on prior η_L. Since x_L is a function of η_L, the former is stochastic and inherits a distribution that depends on the distribution of η_L. Similarly, x_H is a function of η_H. In case of CRRA utility, we note that the function $x_L(\eta_L)$ given by equation 3.1.7 is a bijection, which means we can invert $x_L(\eta_L)$ (and similarly x_H) to express the distribution function of future capital requirement explicitly:[45]

$$F_x(\epsilon) = \pi_L F_L \cdot x_L^{-1}(\epsilon) + \pi_H F_H \cdot x_H^{-1}(\epsilon)$$

■

The proposition shows that in case of discretionary policy, the variance of capital requirement in each state at date-1 is higher relative to the case of rules when this variance is zero since the exact state-contingent capital requirement in known. To summarize, the model displays the following trade-off between a rule-based and a discretionary CCR. On one hand, discretionary CCR gives the regulator the ability to respond to unexpected shocks, but on the other hand it increases policy

[45] We assume that F_L stands for the cumulative distribution function of the random variable η_L, similarly for η_H.

uncertainty. On the contrary, rule-based CCR ties the hand of the regulator in case of unexpected shocks, but it entails to lower policy uncertainty.

3.2 General finite economy

In this section, we extend the model from previous section to a finite horizon setup with T periods, and also incorporate financial frictions *à la* Gertler and Kiyotaki [2010]. At date-T, all output and capital is consumed, and no savings and investments are undertaken. Capital requirements are imposed at all dates expect date-T. The agents in the economy include a unit continuum of households, a representative bank, a representative goods producing firm, a representative capital producing firm, and a regulator that imposes state-contingent bank capital-ratio requirements. State-contingency of regulation is key as we focus on countercyclical capital-ratio regulation that varies as credit conditions in the economy vary. Relative to Gertler and Kiyotaki [2010], we abstract away from the notion of interbank markets, but introduce the following features. First is the possibility for banks to raise fresh equity. This is a crucial feature because in order to increase (decrease) capital-ratio, a bank must have the option to issue (sell) equity in addition to the option of reducing (increasing) lending. Second is the exogenously driven stochastic process that governs the tightness of the borrowing constraint on banks. This process captures the credit condition of the economy, and serves as one of the variables that CCR is made contingent on.

Household The household utility maximization problem subject to the budget constraint is given below, where C_t is consumption, D_t is deposits made at date t, W_t is wage rate, Π_t is profit of capital firms and E_t is the dividend income of bankers which accrues to the household since banks are owned by household. R_{t-1} is the interest rate between dates $t-1$ and t. The household labor supply is perfectly inelastic and is in unit supply.

$$\max_{C_t, D_t} E_t \sum_{i=0}^{T} \beta^i \frac{C_{t+i}^{1-\gamma}}{1-\gamma}$$

$$s.t. \;\; C_t + D_t = W_t + \Pi_t + R_{t-1} D_{t-1} + E_t$$

Firms The representative goods firm uses a standard Cobb-Douglas production function to produce output using labor and capital. It owns capital, and purchases labor in the labor market. Firm equity is held by the bank, who receives dividends from the firm. The firm's profit maximization problem is static, and is given as follows, where A_t is the technology shock:

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

$$K_{t+1} = \psi_{t+1}[I_t + K_t(1-\delta)]$$

$$Z_t = \frac{Y_t - w_t L_t}{K_t}$$

Capital is subject to a quality shock ψ_t which we assume follows an AR(1) process. Since firm equity is held by banks, the capital quality shock is a reduced form exogenous driver of bank asset price dynamics. One interpretation of this shock is that it is the ability of the bank to assess the quality of its equity investment in the firm. In this sense, a higher ψ implies better investment that is expected to generate higher return on assets for the bank. Z_t is the dividend per unit of capital that the firm pays its equity-holders.

Capital producing firms Capital producing firms are owned by households. They produce capital for sale to firms at price Q_t. Their production function is subject to an investment adjustment cost, and their objective is given below, where we assume that the investment adjustment function is given by $f(x) = (x-1)^2$, such that $f(1) = 0$ and $f'(1) = 0$.

$$\max_{I_t} E_t \sum_{\tau=t}^{\infty} \Lambda_{t,\tau} Q_t I_\tau - [1 + f(I_\tau / I_{\tau-1})] I_\tau$$

Banks Banks are managed by bankers, and since bankers are members of the household, banks are owned by the household. The objective of the bank is to maximize expected present value of the dividends E_t they return to the household:

$$\mathbb{E}_t \sum_{i=0}^{\infty} \Lambda_{t,t+i} E_{t+i} \tag{3.2.1}$$

The bank balance sheet identity is as follows, where Q_t is the price of firm equity, S_t is quantity of firm equity, N_t is bank's net worth, and D_t is deposits. Bank dividend payment can take negative values to incorporate the case of equity issuance:

$$Q_t S_t + E_t = N_t + D_t \tag{3.2.2}$$

Bank net worth evolves due to gain in assets from last period, which includes dividend gains and *capital* gains net of depreciation and capital quality shocks, minus the liability to depositors, and minus cost of regulation χ_t, which we elaborate later. Equivalently, current period net worth equals previous period net worth minus interest payment on deposits and cost of regulation, plus dividend and capital gains on risky assets. Net-worth is defined in terms of previous period quantities and current period prices so that it is a sufficient statistic for current period bank decisions and is a state variable.

$$N_{t+1} = [Z_{t+1} + (1 - \delta)Q_{t+1}]\psi_{t+1}S_t - R_t D_t - \chi_t \tag{3.2.3}$$

Credit friction:– There is no credit friction between banks and firms in this model. However, there is credit friction between banks and households. At the end of each period t, the bank is assumed to be able to divert a fraction θ_t of its assets $Q_t S_t$. Upon diverting, the bank defaults on its debt, and is shut down. Depositors claim the residual fraction $1 - \theta_t$ of bank assets. Since depositors realize the bank's incentive to divert, they limit the amount of deposits they are willing to place with the bank. This limit manifests itself as a constraint for the bank in the sense that in equilibrium, the bank should have no incentive to divert assets – the value to a banker from diverting assets should be less than or equal to the value to the banker from continuing banking operations. This incentive constraint is as follows:

$$V_t \geq \theta_t Q_t S_t \tag{3.2.4}$$

where V_t is the value of the bank at the end of period-t. We assume that the diversion fraction θ_t is an exogenously driven stochastic process, a two state Markov chain with support $\{\theta_L, \theta_H\}$ and transition probabilities given by the transition matrix:

$$\Pi = \begin{bmatrix} \pi_L & 1 - \pi_L \\ 1 - \pi_H & \pi_H \end{bmatrix}$$

θ is realized at the beginning of every period, and it is observed perfectly by all agents in the economy. However, Π is known only to the regulator, whereas banks have beta priors η_L, η_H for the transition probabilities π_L, π_H. This last assumption implies the information asymmetry problem.

Regulation:– Regulation is imposed as a state-contingent minimum capital-ratio requirement, which must be satisfied when current period decisions are taken. The requirement imposes a lower bound

$(x_t/1-x_t)$ on the ratio of debt to net-worth, and translates into the following constraint:

$$N_t \geq (Q_tS_t+E_t)x(O_t) \tag{3.2.5}$$

$$\chi_{t+1} = \chi\Big((R_tD_t-(Z_{t+1}+1-\delta)\psi_{t+1}S_t(1-x(O_{t+1}))_+\Big) \tag{3.2.6}$$

Regulation also imposes a fee χ_t on the bank if capital-ratio at the beginning of next period violates the minimum requirement set for next period. The cost is assumed to be a convex function of the deviation from the minimum capital-ratio, and is given by 3.2.6. Intra-period timing for the economy is given in figure 27).

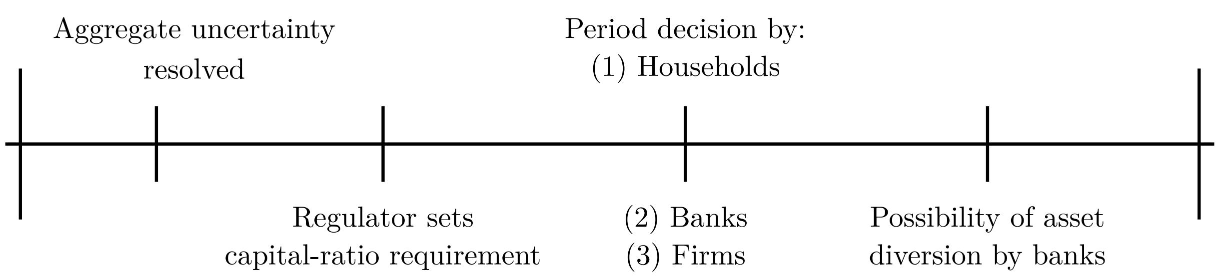


Figure 27: Timeline

Recursive formulation: The bank state variables are ψ_t, N_t. Let the aggregate state vector be O_t, both of which are determined at the beginning of the period. The objective function of the bank can be written recursively as a Bellman equation:

$$V(\psi_t, N_t; O_t) = \max_{E_t,S_t,D_t} E_t + \mathbb{E}_t\Lambda_{t,t+1}V(\psi_{t+1}, N_{t+1}; O_{t+1})$$

subject to 3.2.2, 3.2.3, 3.2.4, 3.2.5 and 3.2.6. We now define the competitive equilibrium for the finite horizon economy.

Definition *A competitive equilibrium for a given state-contingent policy* $\{x_t\}_{t=0}^T$ *is defined by quantities* $\{C_t, D_t, \Pi_t, E_t, Y_t, K_t, L_t, I_t, N_t\}_{t=0}^T$ *and prices* $\{Q_t, R_t, W_t, Z_t\}_{t=0}^T$ *such that*

1. *Given prices,* $\{C_t, D_t\}_{t=0}^T$ *solve household utility maximization problem;*

2. $\{L_t\}_{t=0}^T$ *solves firm profit maximization, given wages;*

3. *Capital good firm's problem is solved by* $\{I_t\}_{t=0}^T$*, given asset prices;*

4. *Given asset price, return on firm shares, return on government debt, and deposit interest rate, $\{E_t, S_t, D_t\}_{t=0}^{T}$ solves bank's problem;*

5. *Labor market clears at wage rate W_t: $L_t = 1 \; \forall \; T \geq t \geq 0$;*

6. *Asset market clears at asset price Q_t: Firm share holding by banks equals capital held by firms $S_t = I_t + (1-\delta)K_t \; \forall \; T \geq t \geq 0$;*

7. *Goods market clears: $Y_t = C_t + I_t(1 + f(I_t/I_{t-1})) \; \forall \; T \geq t \geq 0$;*

The following proposition hold in this economy, the proof for which we provide in the the appendix.

Proposition. *For all $t \in 0, 1, ...T-2$, under information asymmetry, the variance of date-$(t+1)$ capital-ratio requirement for a bank at date-t, conditional on the date-$(t+1)$ state of the economy, is higher in a discretionary regime relative to the rule-based regime.*

4 Conclusion

This paper discusses the trade-off a regulator faces while deciding whether countercyclical capital-ratio requirement (CCR) must be implemented as a rule that it commits to, or if the regulator must retain discretion with regards to the timing and magnitude of change in CCR. We review the literature on time-inconsistency of rule-based policies, and document how this issue can arise in the context of rule-based CCR. We argue that unlike the case of monetary policy, ex-post verifiability of bank decisions disables them to behave strategically in response to regulation, because of which, the time-inconsistency problem fails to manifest in case of CCR. We review the literature on policy uncertainty, and document how the same can arise in the context of discretionary CCR. Using a three-period model we prove that under information asymmetry, discretionary policy leads to an increase in policy uncertainty relative to the rule-based regime — a welfare *decreasing* effect with regards to the regulator's policy implementation trade-off. We use insights from the three period model to prove a similar proposition for the finite-horizon model. The flipside of the policy implementation trade-off is that discretionary CCR enables the regulator to respond to unexpected shocks as opposed to rule-based CCR — a welfare *increasing* effect.

A Appendix to Chapter 1

A.1 Proofs

Propostion 3: Existence of a unique value function I reproduce the problem of a bank in an economy without aggregate uncertainty. The aggregate state is degenerate, so we will not write it specifically.

$$V(\psi, n) = \max_{x\in[1,\overline{x}],h\in[0,1]} \left(hn + E\Lambda' \int_{\frac{R(x-1)}{x}}^{\overline{\psi}} V(\psi', n')F(\psi, d\psi') \right)$$

$$n' = g\Big(((R_k'\psi' - R)x + R)(1-h)n \Big)$$

$$V(\psi, n) \geq \theta(1-h)nx$$

I will follow Stokey [1989] to establish the proof. I will prove that Blackwell conditions are satisfied by the bank's problem, following which, by the Contraction Mapping Theorem (CMT), a unique value function will exist. Some notation is as follows. Let $Y = [1, \phi] \times [0, 1]$ be the action space. Let the exogenous state space of efficiency shocks ψ be $[\underline{\psi}, \overline{\psi}] = \Psi$, and let the endogenous state space of net worth be $\mathcal{N} = [0, \overline{n}]$. Then the state space for the dynamic problem can be written as $\chi = \Psi \times \mathcal{N} \subset \Re^2$. By imposing some structure on the operating cost function $c(.)$, we will ensure that the state space is bounded so that the reward function hn is bounded. This is for mathematical convenience in ensuring that Blackwell Conditions are sufficient for the Bellman mapping to be a contraction. Some more notation is useful. Let $\Gamma : \chi \to Y$ be the policy correspondence (i.e. at each point of the state space, what actions a bank can take). Let $g : \chi \times Y \times \Psi \to \mathcal{N}$ be the law of motion for n (given a point in the state space, and given some action, and given a realization of shock tomorrow, where the bank would transit to). Now I verify the assumptions required for Blackwell conditions to be sufficient for CMT.

Firstly, to ensure a bounded state space, no bank must grow larger than $\overline{n}$ for any starting net worth, leverage, or shock. Since $g(.)$ is increasing and concave, the condition for a bounded state space can be written as follows:

$$\lim_{\omega\to\infty} g(\omega) \leq \overline{n} \implies \lim_{\omega\to\infty} \omega - c(\omega) = \omega - c\omega^{\zeta} \leq \overline{n}$$

This condition imposes a constraint on the set of parameters c, ζ of the cost function and $\overline{n}$. The idea of this condition is that there is an upper limit to how large a bank can grow as for very large sized banks, operating costs grow one to one with net worth. This can be imposed on the parameters as follows:

$$c\zeta\omega^{(\zeta-1)} = 1; \quad \omega - c\omega^{\zeta} = \overline{n}$$

The lower bound on net worth is zero because a bank exits the market as soon as its net worth is less than or equal to zero. This implies a bounded state space for the bank's problem. Secondly, the action space is also bounded as already discussed in the exposition of bank's problem.

The above ensures the following two conditions. (SLP A9.4) $\mathcal{N}$ is a closed, bounded, convex subset of R. (SLP A9.5) Ψ is closed, bounded (hence compact) subset of R. Moreover, F, the transition function for exogenous state ψ should satisfy the Feller property (i.e. maps $C(\Psi) \rightarrow C(\Psi)$), which I ensure by assuming a simple AR(1) process for ψ with Gaussian noise. (SLP A9.6) The policy correspondence Γ is non-empty, compact valued and continuous, which is satisfied in this case because Y is compact. (SLP A9.7) The reward function $H : \chi \times Y \rightarrow \mathcal{R}; H(\psi, n, x, h) = hn$ is bounded because state space is bounded, and is continuous. Also, the discount factor $\beta < 1$. (SLP A9.16) Y is convex and compact subset of R^2. (SLP A9.17) g is continuous in all arguments.

Hence I have verified all assumptions required for Blackwell conditions to be sufficient. I begin with the Bellman equation corresponding to the bank's problem[46] . Let $C(\chi)$ be the space of continuous bounded functions on χ with sup norm. Then, for any $q \in C(\chi)$, the Bellman operator from bank's problem is given as:

$$T(q)(\psi, n) = \max_{\substack{x\in[1,\phi], h\in[0,1] \\ \theta(1-h)nx \leq q(\psi,n)}} \left(hn + \beta \int_{\underline{\psi}}^{\overline{\psi}} q(\psi', n') \mathbb{1}(\psi' \geq \frac{R(x-1)}{x}) F(\psi, d\psi') \right)$$

$$where : n' = g\Big((R'_k\psi' - R)x + R)(1-h)n \Big)$$

Since q is continuous, the maximand is continuous. The action set is closed and bounded, hence compact. For every continuous function on a compact set, a maximum exists, and is attained. Also, since q is bounded, since F satisfies Feller property, and since the reward function H is bounded, $T(q)$ is bounded. Finally, continuity of H and g imply continuity of $T(q)$ by Theorem of the Maximum (SLP Theorem 3.6). Thus I showed that T is a map from space of continuous bounded functions to itself. Now I verify the following Blackwell sufficient conditions (BSC) to conclude that T is a contraction mapping.

[46] Which is different from the kind of Bellman equation employed in SLP, and which makes this proof necessary.

- **Monotonicity:** $\forall f, q \in C(\chi)$ such that $f(x) \leq q(x)\ \forall x \in \chi \implies$

$$T(f)(x) \leq T(q)(x)\ \forall x \in \chi$$

- **Discounting:** $\exists \beta \in (0,1)$ such that $\forall f \in C(\chi),\ \forall a \geq 0,\ \forall x \in \chi \implies$

$$T(f+a)(x) \leq T(f)(x) + \beta a$$

Monotonicity Suppose $f(x) \leq q(x)\ \forall x \in \chi$. Let (x^f, h^f) and (x^q, h^q) be the maximands of $T(f)$ and $T(q)$ respectively at the point $y = (\psi, n)$. This means the following conditions are satisfied:

$$\theta(1-h^q)nx^q \leq q(\psi, n); \quad \theta(1-h^f)nx^f \leq f(\psi, n) \leq q(\psi.n)$$

This means that (x^f, h^f) is a feasible point for the Bellman operator at q. Let $y'(x,h) = (\psi', n')$. Note here that n' is a function of y, x, h, ψ', and this is implicit here. Then,

$$T(q)(y) = \left(h^q n + \beta \int_{\underline{\psi}}^{\overline{\psi}} q(y'(x^q, h^q))\mathbb{1}\left(\psi' \geq \frac{R(x^q-1)}{x^q}\right)F(\psi, d\psi')\right)$$

Using definition of the maximum, and the fact that (x^f, h^f) is a feasible point for the Bellman operator at q, we get:

$$\implies T(q)(y) \geq \left(h^f n + \beta \int_{\underline{\psi}}^{\overline{\psi}} q(y'(x^f, h^f))\mathbb{1}\left(\psi' \geq \frac{R(x^f-1)}{x^f}\right)F(\psi, d\psi')\right)$$

$$\implies T(q)(y) \geq \left(h^f n + \beta \int_{\underline{\psi}}^{\overline{\psi}} f(y'(x^f, h^f))\mathbb{1}\left(\psi' \geq \frac{R(x^f-1)}{x^f}\right)F(\psi, d\psi')\right)$$

$$\implies T(q)(y) \geq T(f)(y)$$

Discounting

$$T(f+a)(y) = \max_{\substack{x\in[1,\phi], h\in[0,1] \\ \theta(1-h)nx \leq (f+a)(y)}} \left(hn + \beta \int_{\underline{\psi}}^{\overline{\psi}} [f(y'(x,h)) + a]\mathbb{1}\left(\psi' \geq \frac{R(x-1)}{x}\right)F(\psi, d\psi')\right)$$

$$\implies T(f+a)(y) \leq T(f)(y) +$$

$$\max_{\substack{x\in[1,\phi],h\in[0,1]\\ \theta(1-h)nx\leq(f+a)(y)}} \int_{\underline{\psi}}^{\overline{\psi}} \beta a \mathbb{1}\left(\psi' \geq \frac{R(x-1)}{x}\right) F(\psi, d\psi')\bigg)$$

which follows because maximization of a sum of functions is at least as small as maximization of individual functions. Using $x = 1, h = 1$, which is a feasible point, in the second maximization, gives us

$$\implies T(f+a)(y) \leq T(f)(y) + \beta a$$

Hence the Bellman operator is a contraction map, implying that there exists a unique fixed point of the Bellman operator. This fixed point is the value function for bank's problem. Consequently, I can use value function iteration to obtain arbitrarily close estimate of the value function.

Proposition 4: Comparative statics. The steady state formulation of bank's problem without the borrowing constraint is given below. The proposition holds for the case with borrowing constraint, but the proof is omitted here as it is more involved.

$$V(\psi, n) = \max_{x\in[1,\overline{x}],h\in[0,1]} \left(hn + \beta \int_{\frac{R(x-1)}{x}}^{\overline{\psi}} V(\psi', n') F(\psi, d\psi')\right)$$

$$n' = g\bigg(((R_k\psi' - R)x + R)(1-h)n\bigg)$$

The optimization problem has four boundary conditions on the choice variables, namely $x \geq 1, x \leq \overline{x}, h \geq 0, h \leq 1$. The comparative statics of the policy functions is trivial if any of the boundary conditions bind.[47] I will therefore derive comparative statics for the general case where none of these conditions bind. For convenience in writing the FOCs, I define the following terms. A numerical subscript denotes derivative with respect to corresponding argument.

$$A(\psi') = R_k\psi' - R$$

$$B(x) = R(x-1)/x; \quad B_1(x) = R/x^2;$$

$$G(x,h,n) = g\bigg((A(\psi')x + R)(1-h)n\bigg); \quad G_1(x,h,n) = g'(.)A(\psi')(1-h)n$$

$$G_2(x,h,n) = -g'(.)(A(\psi')x + R)n; \quad G_3(x,h,n) = g'(.)(A(\psi')x + R)(1-h);$$

[47]In fact, from the structure of the problem, I know that only the second or third constraints might bind in an equilibrium.

$$N(x,h,n) = g\Big(\big(A(B(x))x + R\big)(1-h)n\Big) = g\Big(R(R_k-1)(x-1)(1-h)n\Big);$$

The first order conditions and the envelope conditions are as follows:

$$\begin{aligned}
&[x]: \quad \beta \int_{B(x)}^{\overline{\psi}} V_2(\psi', n')G_1(x,h,n)f(\psi,\psi')d\psi' - \beta V\big(B(x), N(x,h,n)\big)f(\psi, B(x))B_1(x) = 0 \\
&[h]: \quad n + \beta \int_{B(x)}^{\overline{\psi}} V_2(\psi', n')G_2(x,h,n)f(\psi,\psi')d\psi' = 0 \\
&[\psi]: \quad V_1(\psi, n) = \beta \int_{B(x)}^{\overline{\psi}} V(\psi', n')f_1(\psi,\psi')d\psi' \\
&[n]: \quad V_2(\psi, n) = h + \beta \int_{B(x)}^{\overline{\psi}} V_2(\psi', n')G_3(x,h,n)f(\psi,\psi')d\psi'
\end{aligned}$$

Since $-(1-h)G_2(x,h,n) = nG_3(x,h,n)$, $[Env2]$ and $[h]$ imply:

$$V_2(\psi, n) = 1 \implies V(\psi, n) = n + C(\psi) \quad (*)$$

where $C(\psi)$ is some unknown function of ψ. I can use this expression to simplify the FOCs $[x]$ and $[h]$:

$$\begin{aligned}
&[x]: \quad \beta \int_{B(x)}^{\overline{\psi}} G_1(x,h,n)f(\psi,\psi')d\psi' - \beta\Big(N(x,h,n) + C(B(x))\Big)f(\psi, B(x))B_1(x) = 0 \\
&[h]: \quad n + \beta \int_{B(x)}^{\overline{\psi}} G_2(x,h,n)f(\psi,\psi')d\psi' = 0
\end{aligned}$$

This gives me a system of two equations $[x]$ and $[h]$ in policies x, h and state space variables ψ, n, with respect to which I would like to obtain comparative statics results. Note that the only unknown is $C(.)$, which I will address later. I rename equations as follows: $[x] \to q^1(x, h; \psi, n)$ and $[h] \to q^2(x, h; \psi, n)$. Then, assuming that the policy functions $x(\psi, n), h(\psi, n)$ are differentiable, I can apply the Implicit Function Theorem to obtain $x_\psi(\psi, n), h_\psi(\psi, n), x_n(\psi, n), h_n(\psi, n)$. I begin with the following system of equations:

$$-\begin{bmatrix} g_x^1(x,h;\psi,n) & g_h^1(x,h;\psi,n) \\ g_x^2(x,h;\psi,n) & g_h^2(x,h;\psi,n) \end{bmatrix} \begin{bmatrix} x_i(\psi,n) \\ h_i(\psi,n) \end{bmatrix} = \begin{bmatrix} g_i^1(x,h;\psi,n) \\ g_i^2(x,h;\psi,n) \end{bmatrix} \quad i \in \{\psi, n\}$$

Let Δ be the determinant of the Jacobian matrix. Then the system can be solved as follows for

Call report variable name	Variable label	Call report variable name	Variable label
RCFD2170	Total Assets	RCFD0010	Cash
RCFD2948	Total Liabilities	RIAD4460	Dividend component A
RCFD3210	Total Equity Capital	RIAD4470	Dividend component B
RIAD4000	Total operating Income	RCFD3230	Common Stock
RIAD4107	Interest Income	RIAD4073	Total Interest expense
RIAD4079	Non Interest Income	RIAD4074	Net Interest Income
RIAD4340	Net income	RIAD4093	Total Non Interest Income
RIAD4130	Total operating expense	RSSD9950	Date of Incorporation

Table 11: Call report data

$i \in \{\psi, n\}$:

$$\begin{bmatrix} x_i(\psi,n) \\ h_i(\psi,n) \end{bmatrix} = - \begin{bmatrix} g_x^1(x,h;\psi,n) & g_h^1(x,h;\psi,n) \\ g_x^2(x,h;\psi,n) & g_h^2(x,h;\psi,n) \end{bmatrix}^{-1} \begin{bmatrix} g_i^1(x,h;\psi,n) \\ g_i^2(x,h;\psi,n) \end{bmatrix}$$

$$= -\frac{1}{\Delta} \begin{bmatrix} g_h^2(x,h;\psi,n) & -g_h^1(x,h;\psi,n) \\ -g_x^2(x,h;\psi,n) & g_x^1(x,h;\psi,n) \end{bmatrix} \begin{bmatrix} g_i^1(x,h;\psi,n) \\ g_i^2(x,h;\psi,n) \end{bmatrix}$$

Now, since $x(\psi,n)$ and $h(\psi,n)$ are the policy functions, they represent optimal decisions at the given point of the state space, and therefore the Jacobian matrix associated with the first order conditions must have a non-positive determinant, $\Delta \leq 0$. The individual terms of the RHS can be simplified using the FOCs. I also assume that $C'(.) > 0$, which essentially implies that the value function is some increasing function of ψ. Then, using the conditional distribution of ψ' given ψ, and by plugging in the calibrated value of the known parameters R, β, R_k I can sign the partial derivatives of the policy functions.

A.2 Data analysis

I source aggregate time series data from the Federal Reserve Bank of St. Louis Economic Research Database (FRED). I obtain Individual bank data from the Federal Reserve Bank of Chicago Commercial Bank datasets. Financial asset ownership data is obtained from Survey of Consumer Finances. Finally, bank entry and exit data is sourced from the Historical Statistics on Banking database at the Federal Deposit Insurance Corporation. In constructing target moments of banking industry dynamics, I focus on the variables from Call Reports given in Table 11.

Leverage is computed as the ratio of Total Assets and Total Equity Capital (net worth). Mean dividend payout ratio is computed as the ratio of sum of two dividend components and net worth. To compute size ratio of entrants and incumbents, the date of incorporation of a bank is used to identify entrants, and net worth is used as the measure of size, as is the case through out this paper. For exit rate, total number of failures in a given year is divided by the total number of banks at the start of the year. Assisted / unassisted mergers or bailouts are not counted as exits to be consistent with the interpretation of failures in the model. The distribution of bank net worth is based on inflation adjusted total equity capital. I use the total operating expense profile in data to motivate the assumption that banks are subject to convex operating costs in the model.

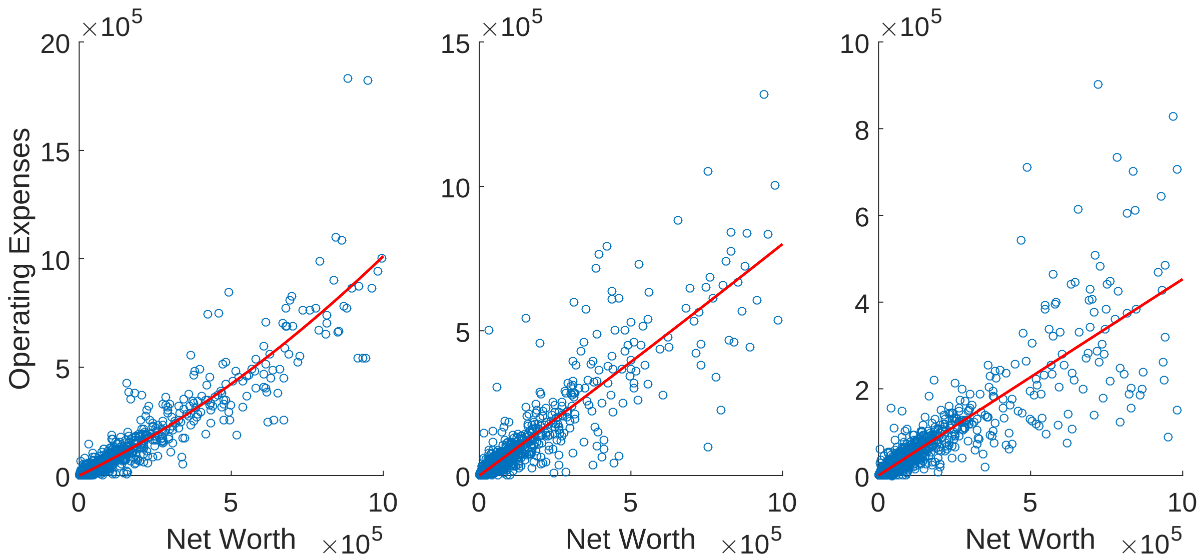


Figure 28: Operating cost and net worth scatter for 1995Q4, 2000Q4, 2005Q4

The result from Kolmogorov-Smirnov (KS) two sample test is given in Table 12. The null hypothesis is that the empirical distribution functions are drawn from the same distribution. I consider various time intervals in the calibration period, and run the test on the distribution of inflation adjusted log net worth of banks as of the beginning (Year 1) and end (Year 2) of the interval. I cannot reject the null, even if the distributions are 11 years apart.:

Year 1	Year 2	Distance	P-value
1995	2005	10	0.3863
1997	2005	8	0.1550
1995	2003	8	0.8727
2000	2005	5	0.0623
1997	2002	5	0.4812
1995	2000	5	0.4526
2003	2005	2	0.6052
1995	1997	2	0.9568

Table 12: KS two sample test

A.3 Additional Figures

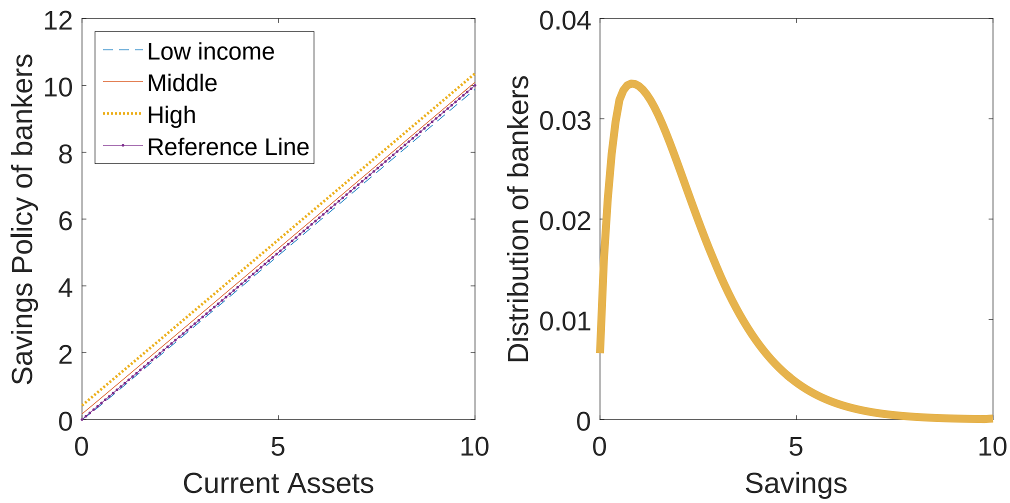


Figure 29: Solution to the bankers' problem

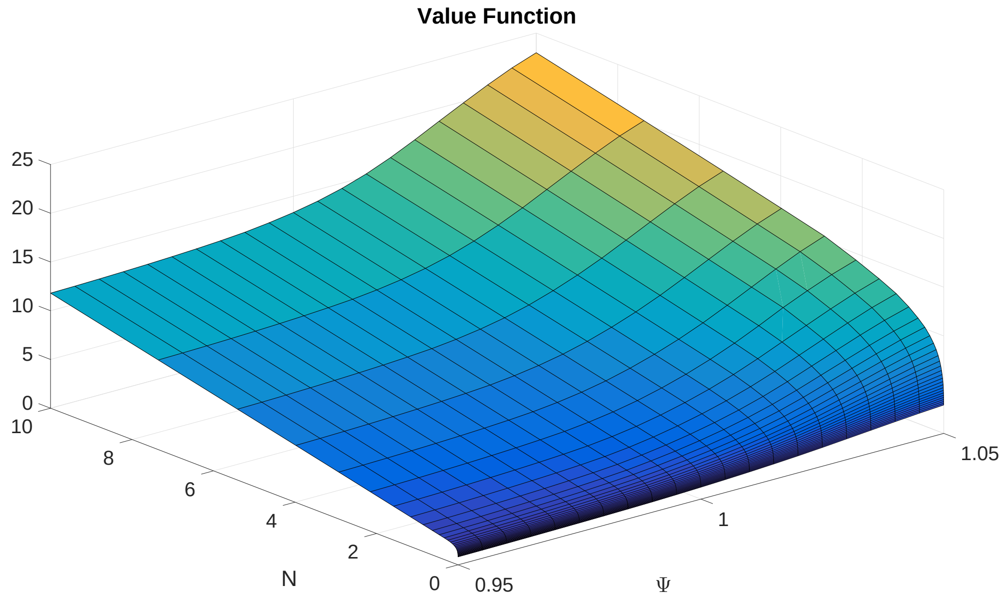


Figure 30: Bank value function

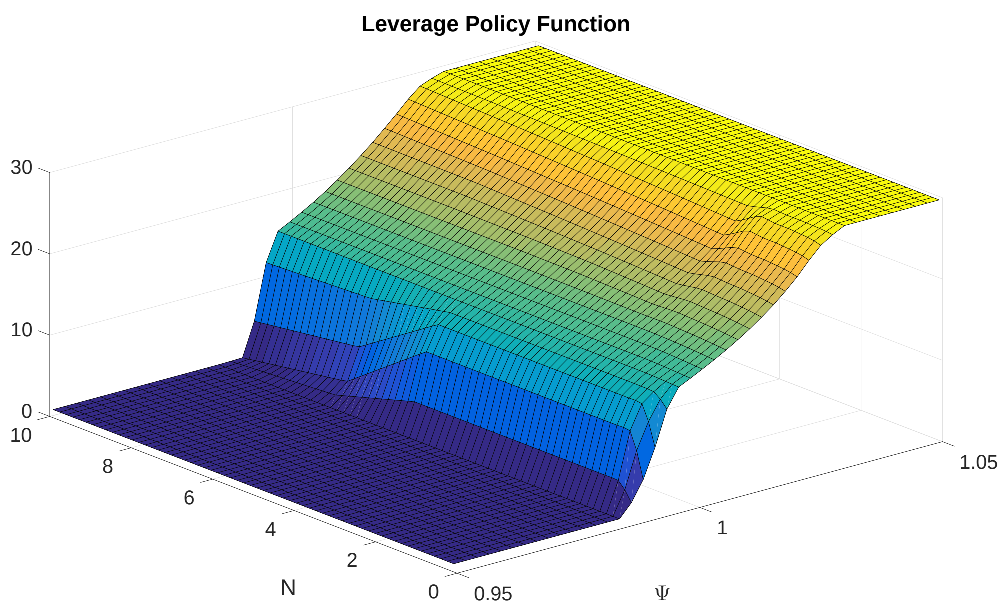


Figure 31: Bank leverage policy function

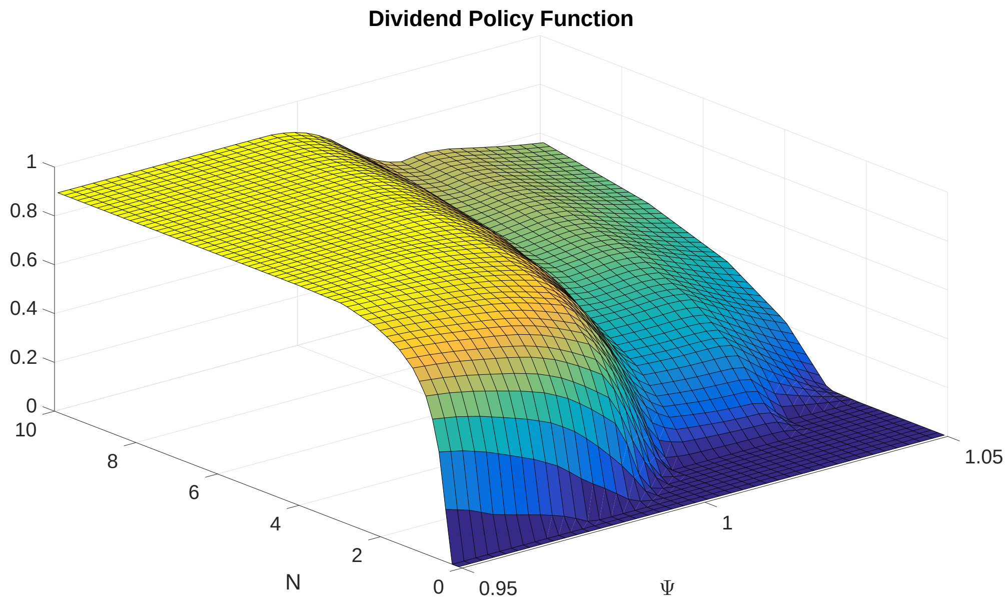


Figure 32: Bank dividend policy function

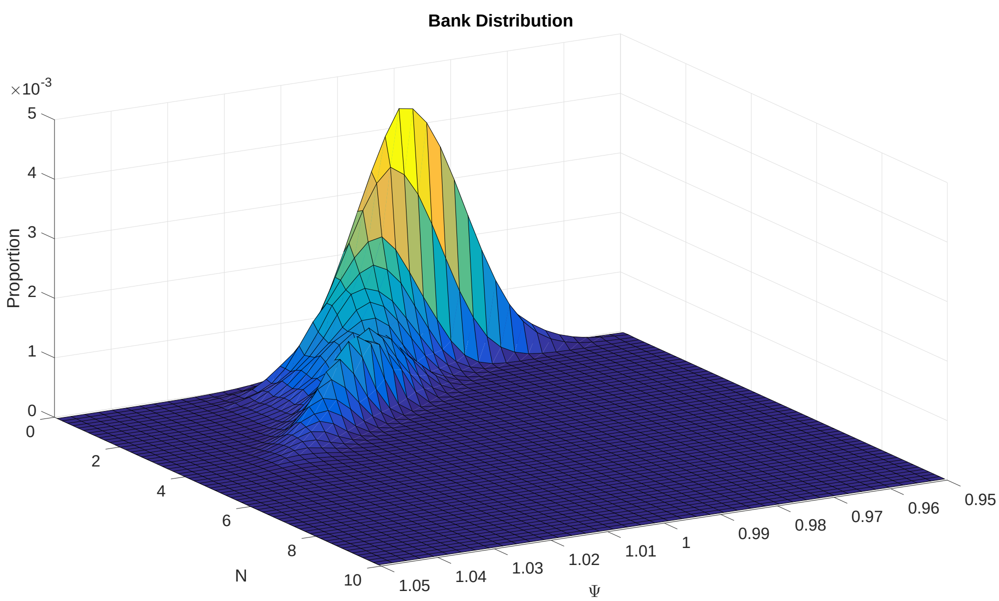


Figure 33: Bank distribution

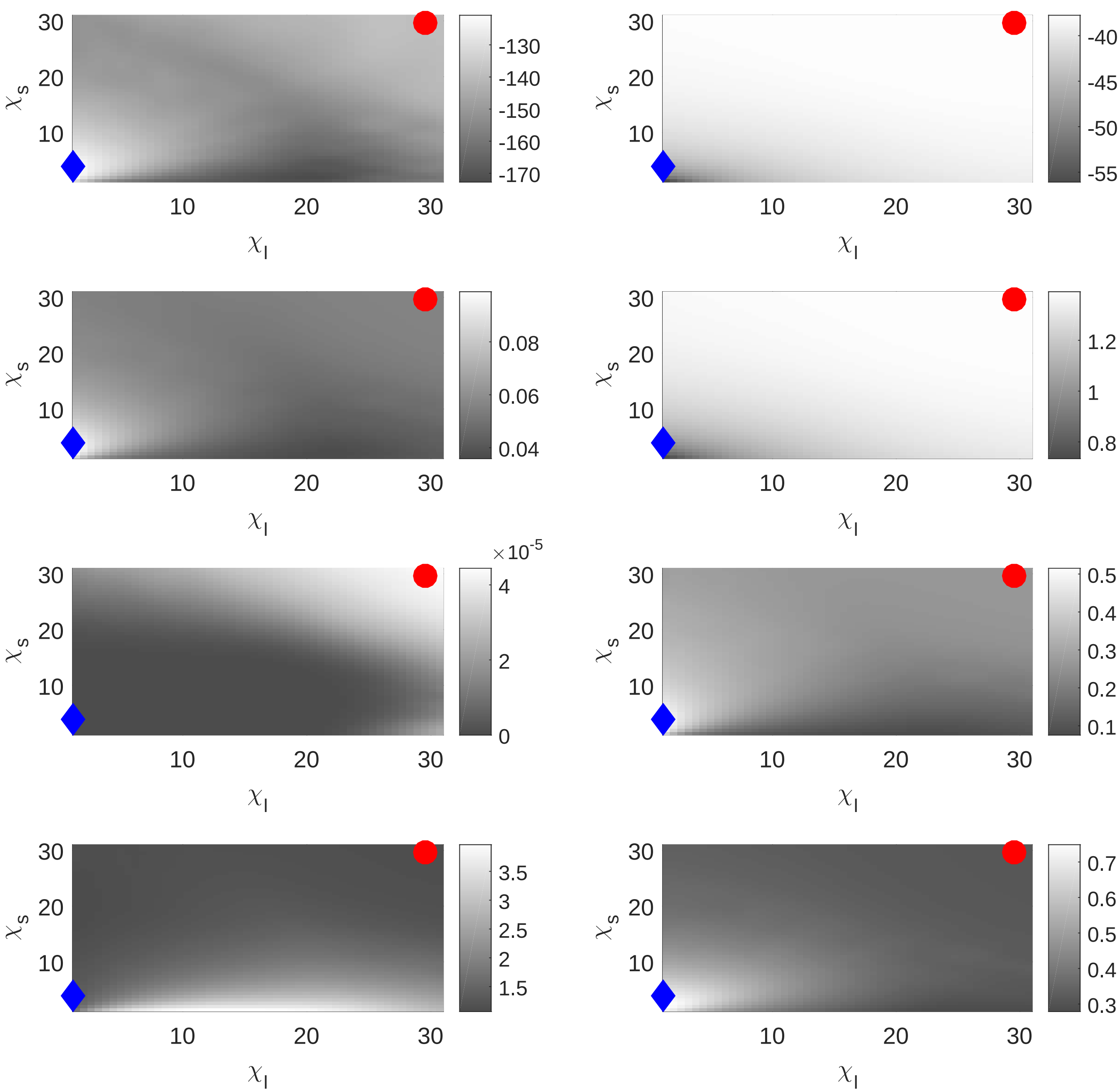
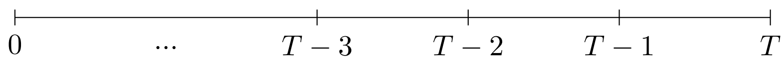


Figure 34: Response of macroeconomic aggregates to change in regulation regime. X_l is the limiting bank regulation, while X_s is the smallest bank regulation. Red is the benchmark regulation, while blue is the optimal regulation.

B Appendix to Chapter 3

B.1 Proof

In this appendix, we prove the proposition that discretionary policy leads to increased policy uncertainty relative to a rule-based regime in a finite horizon economy, the time scale for which is as follows:



We begin by way of backward induction. At date-(T-1), the bank does not care about capital requirement (CR) at date-T since there is no CR in the last period. At date-(T-2), in order to obtain estimates of the CR at date-(T-1), the bank needs to solve the regulator's problem at date-(T-1) using priors η_L and η_H. However, the case of a bank at date-(T-2) in a T-period model is identical to case of a bank at date-0 in the three-period model. For the three-period model, we have shown that contingent on future state of the economy, the variance of CR is higher under the discretionary regime as compared to the rule-based regime. We use this result to conclude that at date-(T-2), banks face greater uncertainty with respect to CR at date-(T-1) in the discretionary regime compared to the rule-based regime. Next, we consider the case of a bank at date-(T-3) and show that this result still holds.

In order for a bank at date-(T-3) to obtain estimates of CR at date-(T-2), it needs to solve the planner's problem at date-(T-2) using priors η_L and η_H. To solve this problem, we adopt the Ramsey approach – first we solve the competitive equilibrium (CE) at date-(T-2) and use the solution from the CE as implementability conditions in the planner's problem at date-(T-2). The CE consists of the bank's and the household's problems. Without loss of generality, suppose the state at date-(T-2) is L. The bank's problem then is as follows, given state variables and current regulation x_{T-2}:

$$V_{T-2}(\psi_{T-2}, N_{T-2}; O_{T-2}) = \max_{S_{T-2}, E_{T-2}, D_{T-2}} E_{T-2} + \beta \mathbb{E}_{T-2} V(\psi_{T-1}, N_{T-1}; O_{T-1})$$

$$s.t. \;\; E_{T-2} + S_{T-2} = N_{T-2} + D_{T-2}$$

$$(E_{T-2} + S_{T-2}) x_{T-2} \leq N_{T-2}$$

$$V_{T-2}(\psi_{T-2}, N_{T-2}; O_{T-2}) \geq \theta_{T-2} Q_{T-2} S_{T-2}$$

$$N_{T-1} = (Z_{T-1} + 1 - \delta)\psi_{T-1} S_{T-2} - R_{T-2} D_{T-2} -$$

$$\chi\Big(\big(R_{T-2} D_{T-2} - (Z_{T-1} + 1 - \delta)\psi_{T-1} S_{T-2}(1 - x_{T-1})\big)_+ \Big)$$

The solution to this problem requires the bank to form an estimate of the distribution of x_{T-1}; however, from the previous stage of induction, we know what this distribution is, and we can obtain a solution $(S_{T-2}, E_{T-2}, D_{T-2})$ to the bank's problem at date-(T-2) as a function of policy x_{T-2} and market prices. Similarly, the household problem can be solved at date-(T-2) to obtain C_{T-2} as a function of market prices. Combined with the market clearing condition, we get a set of *implementability conditions*. The regulator's problem at date-(T-2) that the bank solves at date-(T-3) is as follows:

$$\max_{x_L} u(C_{T-2}) + \beta \mathbb{E} u(C_{T-1}) + \beta^2 \mathbb{E} u(C_T)$$

Note that the expectation operators above are with respect to priors η_L and η_H. The solution to this problem yields x_L at date-(T-2) as a function of these priors, implying positive variance of $x_L(\eta_L, \eta_H)$. Therefore at date-(T-3), the variance of capital requirements at date-(T-2), contingent on the state of the economy, is higher in the discretionary regime as compared to the rule based regime, where the regulation is exactly known. To show that policy uncertainty exists at dates before (T-3), we use backward induction in a similar way. In particular, we use the result about date-(T-2) policy uncertainty for a bank at date-(T-3) while solving for a bank's regulation estimating problem at date-(T-4), which includes solving the bank's and regulator's problems at date-(T-3).

References

Acemoglu, D. and M. K. Jensen (2015). Robust Comparative Statics in Large Dynamic Economies. *Journal of Political Economy 123*(3), 587 – 640.

Acharya, V. V. (2011). The Dodd-Frank act and basel 3 intentions, unintended consequences, transition risks, and lessons for india.

Admati, A. and M. Hellwig (2014). *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It.* Princeton University Press.

Adrian, T. and N. Boyarchenko (2012). Intermediary leverage cycles and financial stability. *Becker Friedman Institute for Research in Economics Working Paper* (2012-010).

Agarwal, I. and T. Goel (2016). Countercyclical capital regulation: Rules versus discretion.

Agarwal, S., J. Barrett, C. Cun, and M. De Nardi (2010). The asset-backed securities markets, the crisis, and talf. *Economic Perspectives 34*(4).

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 659–684.

Albuquerque, R. and H. A. Hopenhayn (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies 71*(2), 285–315.

Allen, F., E. Carletti, and D. Gale (2009). Interbank market liquidity and central bank intervention. *Journal of Monetary Economics 56*(5), 639–652.

Allen, F. and D. Gale (2000). Financial contagion. *Journal of political economy 108*(1), 1–33.

Babus, A. (2007). The formation of financial networks. *Working Paper, Tinbergen Institute Discussion Paper*.

Barro, R. J. and D. B. Gordon (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of monetary economics 12*(1), 101–121.

Basel:III (2011). Basel 3: A global regulatory framework for more resilient banks and banking systems. *Bank of International Settlements.*

Begenau, J. (2014). Capital requirement, risk choice, and liquidity provision in a business cycle model. *Risk Choice, and Liquidity Provision in a Business Cycle Model (October 24, 2014)*.

Benes, J. and M. Kumhof (2011). Risky bank lending and optimal capital adequacy regulation. Technical report.

Bernanke, B. S. (1983). Non-monetary effects of the financial crisis in the propagation of the great depression.

Bianchi, J. and E. G. Mendoza (2015). Optimal time-consistent macroprudential policy. Technical report, National Bureau of Economic Research.

Blake, A. P. and F. Zampolli (2011). Optimal policy in markov-switching rational expectations models. *Journal of Economic Dynamics and Control 35*(10), 1626–1651.

Bonfim, D. and N. Monteiro (2013). The implementation of the countercyclical capital buffer: rules versus discretion. *Banco de Portugal Financial Stability Report*.

Born, B. and J. Pfeifer (2014). Policy risk and the business cycle. *Journal of Monetary Economics 68*, 68 – 85.

Boyd, J. H. and G. De Nicolo (2005). The theory of bank risk taking and competition revisited. *The Journal of finance 60*(3), 1329–1343.

Caballero, R. J. and A. Simsek (2009). Fire sales in a model of complexity. Technical report, National Bureau of Economic Research.

Carroll, C. D. and M. S. Kimball (2007). Precautionary saving and precautionary wealth. *Palgrave Dictionary of Economics and Finance*.

Choe, H., R. W. Masulis, and V. Nanda (1993). Common stock offerings across the business cycle: Theory and evidence. *Journal of Empirical finance 1*(1), 3–31.

Christensen, I., C. Meh, and K. Moran (2011). Bank leverage regulation and macroeconomic dynamic. *CIRANO-Scientific Publications 2011s-76*.

Christiano, L. and D. Ikeda (2013). Leverage restrictions in a business cycle model. Technical report, National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics 1*, 65–148.

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.

Cifuentes, R., G. Ferrucci, and H. S. Shin (2005). Liquidity risk and contagion. *Journal of the European Economic Association 3*(2/3), 556–566.

Clementi, G. L. and H. A. Hopenhayn (2006). A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics*, 229–265.

Clementi, G. L. and B. Palazzo (2015). Entry, exit, firm dynamics, and aggregate fluctuations. Technical report, National Bureau of Economic Research.

Cooley, T. F. and V. Quadrini (2001). Financial markets and firm dynamics. *American Economic Review*, 1286–1310.

Corbae, D. and P. D'erasmo (2013). A quantitative model of banking industry dynamics.

Corbae, D. and P. D'erasmo (2014). Capital requirements in a quantitative model of banking industry dynamics.

Covas, F. and J. C. Driscoll (2014). Bank liquidity and capital regulation in general equilibrium.

Covas, F. and W. J. D. Haan (2011). The cyclical behavior of debt and equity finance. *The American Economic Review 101*(2), 877–899.

Dasgupta, A. (2004). Financial contagion through capital connections: A model of the origin and spread of bank panics. *Journal of the European Economic Association 2*(6), 1049–1084.

DeAngelo, H. and R. M. Stulz (2013). Why high leverage is optimal for banks. Technical report, National Bureau of Economic Research.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), pp. 401–419.

Domanski, D. and P. Turner (2011). The great liquidity freeze: What does it mean for international banking?

Drehmann, M., C. E. Borio, L. Gambacorta, G. Jimenez, and C. Trucharte (2010). Countercyclical capital buffers: exploring options. *Bank for International Settlements*.

Drehmann, M., C. E. Borio, and K. Tsatsaronis (2011). Anchoring countercyclical capital buffers: the role of credit aggregates.

Fernandez-Villaverde, Jesid Guerriuintana, P., K. Kuester, and J. Rubio-Ramirez (2015). Fiscal volatility shocks and economic activity. *American Economic Review 105*(11), 3352–84.

FOMC (2009). Federal open market committee meeting minutes. *Federal Reserve Board*.

FRB (2016). Regulatory capital rules: The federal reserve board's framework for implementing the u.s. basel iii countercyclical capital buffer. Technical report.

Freixas, X., B. M. Parigi, and J.-C. Rochet (2000). Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of money, credit and banking*, 611–638.

Gai, P., A. Haldane, and S. Kapadia (2011). Complexity, concentration and contagion. *Journal of Monetary Economics 58*(5), 453–470.

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*, 547.

Giavazzi, F. and M. McMahon (2012). Policy uncertainty and household savings. *Review of Economics and Statistics 94*(2), 517–531.

Handley, K. and N. Limao (2012). Trade and investment under policy uncertainty: theory and firm evidence. Technical report, National Bureau of Economic Research.

Hanson, S., A. Kashyap, and J. Stein (2010). A macroprudential approach to financial regulation. *Chicago Booth Research Paper* (10-29).

Hassett, K. A. and G. E. Metcalf (1999). Investment with uncertain tax policy: Does random tax policy discourage investment. *The Economic Journal 109*(457), 372–393.

Heer, B. and A. Maussner (2009). *Dynamic general equilibrium modeling: computational methods and applications*. Springer Science & Business Media.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, 1127–1150.

Hughes, J. P. and L. J. Mester (2013). Who said large banks do not experience scale economies? evidence from a risk-return-driven cost function. *Journal of Financial Intermediation 22*(4), 559–585.

IMF (2012). World economic outlook.

Jayaratne, J. and P. E. Strahan (1997). The benefits of branching deregulation. *Economic Policy Review 3*(4).

Julio, B. and Y. Yook (2012). Political uncertainty and corporate investment cycles. *The Journal of Finance 67*(1), 45–83.

Kalemli-Ozcan, S., B. Sorensen, and S. Yesiltas (2012). Leverage across firms, banks, and countries. *Journal of international Economics 88*(2), 284–298.

Kashyap, A. K. and J. C. Stein (2000). What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, 407–428.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy 105*(2), pp. 211–248.

Kowalik, M. (2011). Countercyclical capital regulation: should bank regulators use rules or discretion? *Economic Review-Federal Reserve Bank of Kansas City*, 63.

Krusell, P. and A. A. Smith, Jr (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Kydland, F. E. and E. C. Prescott (1977). Rules rather than discretion: The inconsistency of optimal plans. *The journal of political Economy*, 473–491.

Lang, L., E. Ofek, and R. Stulz (1996). Leverage, investment, and firm growth. *Journal of financial Economics 40*(1), 3–29.

Luttrell, D., T. Atkinson, and H. Rosenblum (2013). Assessing the costs and consequences of the 2007–09 financial crisis and its aftermath. *Economic Letter 8*.

Miao, J. (2005). Optimal capital structure and industry dynamics. *The Journal of finance 60*(6), 2621–2659.

Morris, S. (2000). Contagion. *The Review of Economic Studies 67*(1), 57–78.

Müller, J. (2006). Interbank credit lines as a channel of contagion. *Journal of Financial Services Research 29*(1), 37–60.

Nguyen, T. T. (2014). Bank capital requirements: A quantitative analysis. *Available at SSRN 2356043*.

Pastor, L. and P. Veronesi (2012). Uncertainty about government policy and stock prices. *The Journal of Finance 67*(4), 1219–1264.

Pástor, L. and P. Veronesi (2013). Political uncertainty and risk premia. *Journal of Financial Economics 110*(3), 520–545.

Pindyck, R. S. (1988). Irreversible investment, capacity choice, and the value of the firm. *The American Economic Review*, 969–985.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control 33*(3), 649–665.

Repullo, R. and J. Saurina Salas (2011). The countercyclical capital buffer of basel iii: A critical assessment.

Repullo, R. and J. Suarez (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies 26*(2), 452–490.

Rodrik, D. (1991). Policy uncertainty and private investment in developing countries. *Journal of Development Economics 36*(2), 229–242.

Shin, H. S. (2009). Securitisation and financial stability*. *The Economic Journal 119*(536), 309–332.

Stokey, N. L. (1989). *Recursive methods in economic dynamics.* Harvard University Press.

Stulz, R. (1986). Interest rates and monetary policy uncertainty. *Journal of Monetary Economics 17*(3), 331–347.

Van den Heuvel, S. J. (2008). The welfare cost of bank capital requirements. *Journal of Monetary Economics 55*(2), 298–320.

Zhu, H. (2007). Capital regulation and banks' financial decisions. *BIS Working Paper*.