GOVERNMENT INTERVENTION IN FINANCIAL MARKETS

by

Kairong Xiao

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2017

Abstract

Governments play an important role in financial markets around the world. This thesis studies theoretical mechanisms and empirical consequences of government actions in financial markets in order to better understand the organization of the financial sector and the inner working of governments. The first essay "Shadow Banks, Deposit Competition, and Monetary Policy" studies the transmission mechanism of monetary policy through the shadow banking system, a group of non-bank financial intermediaries conducting banking business in the economy. This essay shows empirically and theoretically that the shadow banking system partially offsets the impact of monetary policy on the traditional commercial banking system and may lead to unintended consequences in terms of the stability of the financial system. The second essay "Regulation and Market Liquidity" (co-authored with Professor Francesco Trebbi), explores whether the post-crisis financial regulations, including the Dodd-Frank Act and Basel III, have caused liquidity deterioration in the U.S. fixed income market. Against the popular claim that post-crisis regulations hurt liquidity, this essay finds no evidence of liquidity deterioration during periods of regulatory intervention. Instead, liquidity seems to have improved in this period. The third essay, "Factions in Nondemocracies: Theory and Evidence from the Chinese Communist Party" (co-authored with Professor Patrick Francois and Professor Francesco Trebbi), investigates theoretically and empirically the factional arrangements and dynamics within the Chinese Communist Party (CCP), the governing political party of the People's Republic of China. This essay documents a set of new empirical findings showing how factional politics affects the promotion of individual politicians within the CCP hierarchy. This essay proposes a theoretical model to rationalize these findings and conduct a set of counterfactual analyses of possible institutional changes within the CCP.

Lay Summary

This thesis studies the interaction between governments and financial markets. The first essay "Shadow banks, Deposit competition, and Monetary Policy" shows that the shadow banking system, a group of non-bank financial intermediaries conducting banking business in the economy, can partially offset the impact of monetary policy. The second essay "Regulation and Market Liquidity" (co-authored with Professor Francesco Trebbi), shows that the post-crisis financial regulations have not caused liquidity deterioration in the U.S. fixed income market. The third essay, "Factions in Nondemocracies: Theory and Evidence from the Chinese Communist Party" (co-authored with Professor Patrick Francois and Professor Francesco Trebbi), investigates how informal groups of politicians known as factions affect the promotion of individual politicians within the Chinese Communist Party (CCP). In summary, this thesis helps us to better understand the important role of governments in financial markets.

Preface

The research project in Chapter 2 was identified and performed solely by the author. The essay in Chapter 3 is based on unpublished research with Francesco Trebbi (University of British Columbia). The essay in Chapter 4 is based on unpublished research with Francesco Trebbi (University of British Columbia) and Patrick Francois (University of British Columbia). In the coauthored work of Chapters 3 and 4, I worked on the development of the research question, data collection, empirical analysis, structural estimation, and part of the writing of the manuscript. While it is hard to quantify exactly, my personal shares of contribution to Chapters 3 and 4 amount to about 1/2 and 1/3.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Chapter 1

Introduction

This thesis is a collection of three essays at the intersection of finance, industrial organization, and political economy. Although the topics are diverse, they share the common objective of studying the interplay between governments and financial markets. The first essay studies the transmission mechanism of monetary policy through the shadow banking system, a group of non-bank financial intermediaries conducting banking business in the economy. Using the U.S. money aggregate data from 1987 to 2012, I find that the shadow banking system partially offsets the impact of monetary policy on the traditional commercial banking system. I construct a structural model of bank competition and show that this new channel is a result of deposit competition between commercial and shadow banks in a market with heterogeneous depositors. The second essay empirically examines the effects of the post-crisis financial regulations, encompassing the Dodd-Frank Act and Basel III, on market liquidity of the U.S. fixed income market. Against the popular claim that post-crisis regulations hurt liquidity, this study finds no evidence of liquidity deterioration during periods of regulatory intervention. The third essay investigates theoretically and empirically the factional arrangements and dynamics within the Chinese Communist Party (CCP), the governing political party of the People's Republic of China. This study presents a set of new empirical regularities within the CCP and a theoretical framework suited to model factional politics within single-party regimes. Because each essay investigates a different topic, chapters were designed to be self-contained. I thus leave a more exhaustive discussion of the research question and contribution to the introduction specific to each chapter.

Chapter 2

Shadow Banks, Deposit Competition, and Monetary Policy

2.1 Introduction

The U.S. banking system has experienced significant structural changes over the past thirty years. A group of non-bank financial intermediaries, collectively known as the *shadow banking system*, has grown outside of the traditional commercial banking sector. Important components of the shadow banking system include money market funds (MMFs), securitization vehicles, broker-dealers, and mortgage companies. Shadow banks compete with commercial banks in many traditional banking businesses. For example, MMFs compete in the deposit market by creating liquid claims which, in many ways, are similar to commercial bank deposits, yet provide a higher yield. In recent years, more than 30% of deposits have been created by shadow banks.

The rapid growth of shadow banks has raised two main concerns for policy makers.[1] The first concern regards the effectiveness of monetary policy in the presence of a sizable shadow banking sector. Traditionally, commercial banks play an important role in transmitting monetary policy to the real economy. However, a large proportion of deposits are now created outside of the commercial banking sector. How does the deposit competition from shadow banks affect the transmission of monetary policy? The second concern regards the effect of shadow banks on financial stability. Deposit creation involves the risk of bank runs. The risk of bank runs is more severe for shadow banks, because their deposits are not insured by the government. By creating uninsured deposits, shadow banks may have negative impacts on financial stability. Motivated by these concerns, this paper examines how shadow bank deposit competition affects the transmission of monetary policy, and offers an analysis of the implications of such competition for financial stability.

Unlike commercial banks which combine deposit creation and loan origination under one roof, the shadow banking system separates the intermediation process into different entities. MMFs provide depository services for households and businesses, and then pass the proceeds to other shadow banks such as mortgage companies which specialize in loan origination. This paper focuses on MMFs, as these are the main entities which create deposits in the shadow banking system.

I first document a new transmission channel of monetary policy in the shadow banking system. Standard theories of monetary transmission predict that high interest rates are associated with low deposit creation (Bernanke and Blinder, 1988; Drechsler, Schnabl, and Savov, 2016).[2] This prediction has been verified empirically by previous literature in the commercial banking sector (Kashyap and Stein, 1995, 2000; Drechsler,

---

[1]For instance Federal Reserve Board Chair, Janet Yellen, in response to a question by IMF Managing Director Christine Lagarde on Shadow Banking in July 2014, said that "we won't be able to detect them (shadow banking), and if we can, we won't have adequate regulatory tools. That is a huge challenge to which I don't have a great answer".

[2]The shared idea of these two theories is that high interest rate policy increases the opportunity cost of holding liquid deposits, which reduces the amount of bank deposits in economy. The difference between these two theories is how high interest rate policy increases the opportunity cost of holding liquid deposits. Bernanke and Blinder (1988) suggest the reserve requirement of commercial banks as an important channel, while Drechsler, Schnabl, and Savov (2016) show that market power of commercial banks can also play a role.

Schnabl, and Savov, 2016). However, using aggregate U.S. money supply data from 1987 to 2012, I find the *opposite* of commercial banks happens for shadow banks.[3] When the Federal Reserve wants to reduce deposits by raising interest rates, shadow bank deposits expand dramatically, and as a result, dampen the impact of monetary policy. The contrast between shadow and commercial banks can be easily seen in a time-series plot of the deposit growth rates as shown in Figure 2.1. This finding contradicts conventional wisdom that high interest rates are contractionary for deposit creation. It suggests that the monetary transmission channel in the shadow banking sector is different from the traditional channels in the commercial banking sector. Moreover, my results show that monetary policy not only affects the *total amount* of bank deposits, but also the *relative shares* between the shadow and commercial banking sectors. Because shadow bank deposits are outside of government safety nets such as the deposit insurance and the discount window, shifts in the relative shares of deposits have important implications for financial stability. To the best of my knowledge, the present study is one of the first to document this counter-intuitive results of shadow bank deposit creation.

In order to understand the underlying mechanism, I develop a structural model of bank competition following the industrial organization (IO) literature on oligopoly markets.[4] I show that the expansion of shadow bank deposits during periods of monetary tightening is a result of deposit competition between commercial and shadow banks. In my model, banks are differentiated by their respective degrees of transaction convenience and yields. Commercial banks provide superior transaction services such as branch networks and payment systems, while shadow banks compete on yields due to the lack of bank charters to offer transaction services similar to those offered by commercial banks. Banks compete for a continuum of depositors with different preferences over transaction convenience and yields. Commercial banks attract a group of transaction-oriented depositors who value transaction services but are insensitive to yields. Typical examples of transaction-oriented depositors include small and unsophisticated depositors who choose banks mainly based on geographical proximity rather than the potential yields. In contrast, shadow banks attract a group of yield-oriented depositors such as wealthy individuals and corporate treasurers. These yield-oriented depositors are not primarily concerned with transaction convenience, but instead are very sensitive to yields.

Depending on their depositor clientele, deposit rates of different banks exhibit different sensitivities to changes of market interest rates. When the Federal Reserve increases interest rates, commercial banks are reluctant to increase their deposit rates. This is because their main clientele, the transaction-oriented depositors, view cash as the main alternative for transactions. Because cash bears no interests, commercial banks are able to keep paying low interest rates without losing many of their transaction-oriented depositors. As a consequence, commercial banks can earn higher spreads between lending rates and deposit rates in periods of high interest rates. In contrast, shadow banks have to raise their deposit rates together with the market interest rates, because otherwise their yield-oriented clientele will switch to other high-yielding liquid assets such as short-term bonds. Because shadow bank deposit rates are more sensitive to market interest rates than commercial bank deposit rates, high interest rate policy usually widens the difference in deposit rates between shadow and commercial banks, inducing some marginal depositors of commercial banks to switch over to shadow banks. This explains why shadow banks expand their deposit creation when the Federal Reserve tightens monetary policy.

In order to assess the quantitative importance of the shadow bank channel of monetary policy, I estimate my model using institution-level data on U.S. commercial banks and MMFs. The estimation shows that

---

[3]In this paper, I use "MZM" (money zero maturity) as the measure of money supply in the economy. This measure is a modification of M2 after the usefulness of previous measures became comprised in the 1990s. This measure includes currency, traveler's checks of non-bank issuers, demand deposits, other checkable deposits, saving deposits, retail and institutional MMF shares. Choosing a specific definition of money aggregate, however, is not important, because my question is about each component of the money aggregates, rather than the sum.

[4]See Berry (1994), Berry, Levinsohn, and Pakes (1995) and Nevo (2001).

commercial bank deposits provide significantly higher convenience than shadow bank deposits. The estimation also shows that depositors exhibit significant differences in their preference over convenience and yields. Different types of depositors self-select into different types of banks. The heterogeneity in depositor clientele quantitatively explains the different responses to monetary policy of the two banking sectors.

I use my model to study the implications of shadow banking for monetary policy. I simulate a counterfactual economy without shadow banks using estimated parameters. Comparing the real data with the counterfactual economy, I find that the presence of shadow banks reduces the sensitivity of the aggregate money supply to the Fed Funds rates by 40 percent. I further use my model to study the implications of shadow banking for financial stability. I show that shadow banks may be more prone to bank runs because a small shock to asset values may trigger large redemption from their yield-oriented clientele. Using the runs on MMFs in 2008 as a case study, I find that MMFs with higher demand elasticity before the financial crisis subsequently suffered more severe runs. Finally, my results suggest a cautious stance towards a recent policy proposal which suggests using monetary tightening as a tool for promoting financial stability (Stein, 2012; Borio and Zhu, 2012; Ajello et al. 2016). I show that this policy proposal may unintentionally drive deposits from the insured commercial banking sector into the uninsured shadow banking sector, and in doing so, heighten the risk of bank runs.

This paper contributes to three strands of literature. The first strand studies the monetary transmission mechanism in the banking system. Traditionally, this literature has focused on commercial banks (Bernanke and Blinder, 1988; Kashyap and Stein, 1995, 2000; Drechsler, Schnabl, and Savov, 2016). This paper brings shadow banks into the forefront of the theoretical and empirical analysis of monetary policy. I document a new transmission channel of monetary policy in the shadow banking system, which is contrary to the conventional wisdom in the commercial banking sector. This new channel partially offsets the traditional channels in commercial banks and dampens the impact of monetary policy.

I further provide quantitative estimates of monetary transmission mechanisms for both commercial and shadow banks. Previous literature has made reduced form arguments about how monetary policy is transmitted in the commercial banking system. Bernanke and Blinder (1988) and Kashyap and Stein (1995, 2000) argue that monetary policy works through the opportunity cost of bank reserves, while more recent work such as Scharfstein and Sunderam (2016) and Drechsler, Schnabl, and Savov (2016) acknowledges that the market power of the banking sector may also play a role in the transmission of monetary policy. Since the aforementioned literature relies on reduced-form models, their approach cannot quantify the impact of each channel. The present study complements the previous literature by providing a structural IO model to quantify these different channels. My estimates suggest that the market power channel has been playing a dominant role in commercial banks since the 1990's. This structural framework is also used by Egan, Hortaçsu, and Matvos (2015) to study bank run risks in the commercial banking sector. In contrast to their analyses, my paper introduces shadow banks and heterogeneity in depositors' preferences with the aim of demonstrating the heterogeneous transmission of monetary policy in the modern banking system.

The second strand of literature to which this paper contributes concerns the interaction between monetary policy and macro-prudential policies. Prior to the 2008-09 financial crisis, the consensus among policy makers was that monetary authority should focus on price stability and employment (Smets, 2013). However, this consensus has been challenged by an alternative view that took shape after the financial crisis, arguing that monetary policy should also be used to promote financial stability (Stein, 2012; Borio and Zhu, 2012; Ajello et al. 2016). Proponents of this view contend that by tightening monetary policy, the central bank can curb the creation of money-like liabilities within the banking system and reduce the appetite of investors for risk. My findings contribute to this debate by showing that monetary tightening may lead to the unintended consequence of driving deposits to the shadow banking system. Since shadow banks are not protected by deposit insurance, such a policy may actually increase systemic risk. My paper supports the view that "monetary policy is too blunt a tool to address possible financial imbalances" as argued by Bernanke (2011) and Yellen (2014).

The third strand of literature studies shadow banking, and particularly, the sources of fragility in the shadow banking sector. Previous research finds that the lack of deposit insurance (Gorton and Metrick, 2012), high leverage (Adrian and Shin, 2010; Moreira and Savov, 2016), and information opacity (Dang, Gorton, and Holmström, 2016) create fragility in the shadow banking system. My paper contributes to this line of research by showing that the yield-sensitive clientele of shadow banks can also be a source of fragility. My paper also adds to a third group of papers on MMFs (Kacperczyk and Schnabl, 2013; McCabe et al., 2013; Schmidt, Timmermann and Wermers, 2016; Parlatore, 2016). I show that monetary policy has a strong impact on the deposit flows of MMFs by changing the competition environment of the deposit market.

The remainder of this paper is organized as follows. Section 2.2 presents several new stylized facts on deposit creation of the shadow banking system. Section 2.3 presents a structural model of bank competition to rationalize the empirical findings. Section 2.4 presents the estimation procedure and results. Section 2.5 discusses policy implications and Section 2.6 concludes.

2.2 Deposit Creation by Shadow Banks

In this section, I provide a brief description of the institutional background of the shadow banking system. I then present several stylized facts to motivate my study.

Institutional Background

The shadow banking system is a collection of financial intermediaries which conduct maturity, credit, and liquidity transformation outside the traditional commercial banking system.[5] Examples of shadow banks include securitization vehicles, asset-backed commercial paper (ABCP) conduits, MMFs, investment banks, and mortgage companies. Like commercial banks, shadow banks transform long-term illiquid assets into short-term money-like claims. Since households and business have a preference for liquidity, issuing money-like claims allows shadow banks to lower their financing costs.

Figure 2.2 provides a simplified representation of the U.S. banking system.[6] The upper branch represents the commercial banking sector, while the lower represents the shadow banking sector. Unlike commercial banks, which combine deposit creation and loan origination under one roof, the shadow banking system separates the intermediation process into different entities. MMFs constitute the first stage of the shadow banking intermediation process. MMFs take deposits from households and businesses and then pass the proceeds to other shadow banks such as securitization vehicles, mortgage conduits, broker dealers, and mortgage companies, which specialize in loan origination. In this process MMFs create money-like liabilities, MMF shares, which resembles commercial bank deposits.

MMF shares are widely (though not necessarily accurately) regarded as being safe as bank deposits, yet providing a higher yield. Similar to commercial bank deposits, MMFs provide intraday liquidity, and some of them even allow depositors to write checks on their deposits. Unlike other shadow banking liabilities, which are generally held within the shadow banking system, MMF shares are directly held by households and businesses. Due to their similarity with commercial bank deposits, MMF shares are included in the official money supply statistics, while other types of shadow banking liabilities are generally not. The amount of MMF shares also provides a good proxy of the quantity of funds flowing into the shadow banking sector.

---

[5]Former Federal Reserve Chair Ben Bernanke provided a definition of the shadow banking system in April 2012: "Shadow banking, as usually defined, comprises a diverse set of institutions and markets that, collectively, carry out traditional banking functions–but do so outside, or in ways only loosely linked to, the traditional system of regulated depository institutions."

[6]A more detailed description of shadow banking intermediation process can be found in Pozsar et al. (2010).

On the asset side, MMFs hold various money market instruments. The asset holdings of MMFs can be grouped into three major categories. According to the iMoneyNet data, the majority 50 percent are invested in short-term debts of other shadow banks such as repurchase agreement (repos), asset backed commercial papers (ABCPs), commercial papers (CPs) and floating rate notes (FRNs).[7] 20 percent are invested in Treasury and agency securities. Lastly, 18 percent of the shadow bank deposits go back to the commercial banking sector in the form of large denomination commercial bank obligations.

Over the past thirty years, the shadow banking sector has become increasingly important in the economy. Based on the aggregate money supply statistics from the Federal Reserve, the share of shadow bank deposits has increased from around 15 percent in the 1980s to around 40 percent in 2007, while the share of commercial bank deposits is on a downward trend.

Data Source

The first main database used in this paper is iMoneyNet. This data provides monthly share class level data for the U.S. MMFs dating back to 1985. After cross check with the aggregate money supply statistics from the Federal Reserve Board, I find that this database covers essentially all the MMFs which are included in the official statistics after 1987. The data contains rich information on fund characteristics such as deposit amounts, charged expense ratio, yields, check-writing privilege, and fund sponsors. The data also provides information on fund operating cost such as incurred management fee, share service fee, 12b1 fee and other fees. Portfolio holding information becomes available since 1998, which includes average portfolio maturity, and portfolio weight by asset class. As data on shadow banks are generally very scarce, this data set provides a rare opportunity to look into the inner working of the shadow banking system.

The second main data is the Consolidated Report of Condition and Income, generally referred as the Call report. This data provides quarterly bank-level data for every U.S. insured commercial bank, including detailed accounting information such as deposit amounts, interest income, salary expense, and fixed asset expenses. I complement the Call report with FDIC Summary of Deposits, which provides branch-level information on deposit amounts in annual frequency since 1994. Following the literature, deposit rates are imputed from bank financial statements by dividing deposit interest expense over total amount of deposits (Dick, 2008; Hannan and Prager, 2004). In the following analysis, I focus on "liquid deposits" which are defined as the sum of checking and savings deposits[8].

In addition to the above two main data sources, I also use the Survey of Consumer Finance (SCF) 2013 to obtain depositor-level deposit holdings and demographic information. Lastly, aggregate time series of the amount of cash held by households and the Fed Funds rates are retrieved from Federal Reserve Economic Data (FRED). Aggregate time series of the amount of Treasury bills held by households is retrieved from the Financial Accounts of the United States.

Shadow Bank Channel

In what follows, I document a new transmission channel of monetary policy using aggregate money supply data from 1987 to 2012 from the Federal Reserve. I break down the aggregate money supply into cash, commercial bank deposits, and shadow bank deposits. Commercial bank deposits include demand and saving deposits. Shadow banking deposits include retail and institutional MMF shares. Figure 2.1 plots the Fed Funds rates and the annual deposit growth rates of each banking sector over time. Conventional

---

[7]Some of large industrial corporations also issue commercial papers to obtain short term financing. These commercial papers are mainly used to finance their captive finance companies, which are also considered as shadow banks. For example, one of the largest issuers of commercial paper, General Motors Acceptance Corporation (GMAC), is a captive finance company that provides financing for the customers of its parent company, General Motors.

[8]Previous literature has shown that the pricing and quantities of "liquid deposits" are quite different from "illiquid deposits" such as small time saving deposits (Driscoll and Judson, 2002; Drechsler et al., 2009).

monetary transmission channels predict that high Fed Funds rates have tightening effects on the money supply (Bernanke and Blinder, 1988; Kashyap and Stein 1995, 2000; and Drechsler, Schnabl, and Savov 2016). Indeed as shown in the top panel of Figure 2.1, high Fed Funds rates are associated with low growth rates of commercial bank deposits. However, the opposite happens in the shadow banking system. As shown in the bottom panel of Figure 2.1, high Fed Funds rates are associated with high growth rates of shadow bank deposits. This finding implies that monetary policy seems to have a different transmission channel in the shadow banking system. High interest rate policy, which is intended to reduce money supply in the economy, surprisingly increases deposit creation by shadow banks.

To assess the magnitude of the new transmission channel, I regress deposit growth rates of each banking sector on the Fed Funds rates, controlling for a list of macroeconomic variables such as GDP growth rates, inflation, VIX, and TED spread:

$$Deposit\ Growth\ Rates_t = \alpha + \beta Fed\ Funds\ Rates_t + \gamma X_t + \varepsilon_t \tag{2.1}$$

Table 2.2 presents the results. Consistent with the graphical observation, monetary policy has opposite effects on these two sectors: a 1 percent increase in the Fed Funds rates is associated with a 1.4 percent decrease in the growth rates of commercial bank deposits, but a 3.9 percent increase in the growth rates of shadow bank deposits. The estimates are both statistically and economically significant.

Column 4 and 8 show the results for the total money supply. The coefficients of the Fed Funds rates are insignificant different from zero. This result shows that deposit creation by shadow banks partially offsets the reduction of commercial bank deposits and attenuates the impact of monetary tightening on aggregate money supply. As shadow banks create more deposits, they obtain more loanable funds for lending. In Section 2.5.2, I further show that shadow bank lending also increases as the Fed tightens monetary policy.

To summarize, the above results show that shadow banks may dampen the impact of monetary policy by creating more money-like liabilities when the Fed wants to reduce money supply. Furthermore, this channel implies that monetary policy not only affects the total amount of money supply, but also the relative shares of money supply between the shadow and commercial banking sector. Since shadow banks do not have access to government safety nets such as deposit insurance and discount window, such shifts in the composition of money supply have important implications for financial stability.

2.3 A Structural Model of Bank Competition

2.3.1 Intuition

The previous section documents that monetary policy has very different impacts on the amount of deposits created by commercial and shadow banking sector. In this section, I develop a structural IO model to rationalize the above empirical findings. There are two key ingredients of the model. First, banks are differentiated in the dimensions of convenience and yields. Commercial bank deposits offer a lot transaction services such as branch networks, ATMs, and payment system. In contrast, shadow banks compete on yields because they do not have bank charters to provide those transaction services.[9] In addition to heterogeneous banks, the second key ingredient is that depositors exhibit heterogeneous preference over convenience and yields. There are a group of "transaction-oriented" depositors who care a lot about transaction convenience, but are not sensitive to yields. For example, "mom and pop" depositors choose banks mainly based on geographical proximity rather than deposit rates paid by banks. There are also a group of "yield-oriented depositors" who are very sensitive to yields but are relative insensitive to convenience. For example, large corporations

---

[9]In addition, the deposit insurance on commercial bank deposits also increases their convenience relative to shadow banks. The deposit insurance of the commercial bank deposits is less relevant for for very large depositors because the FDIC only insures commercial bank deposits up to a certain amount.

and wealthy individuals usually have large amounts of deposits. A small difference in yields can make a big difference in the dollar value of income. Moreover, these depositors are often more sophisticated than "mom and pop" depositors. Therefore, they are better-equipped to find the highest yielding options in the market and monitor the risks associated with them.

These two groups of depositors are likely to self-select into different types of banks. Commercial banks are likely to attract more transaction-oriented depositors because of the superior transaction services offered by them, while shadow banks attract more yield-oriented depositors because of the high deposit rates. Consistent with this idea, using the Survey of Consumer Finances (SCF) 2013, I find that depositors who are rich or more sophisticated (proxied by college education) are more likely to choose shadow banks. The result is reported in Table 2.3.

Facing different clientele, deposit rates of different banks have different sensitivities to monetary policy. When the Federal Reserve increases interest rates, commercial banks are reluctant to increase their deposit rates. This is because their main depositor clientele, transaction-oriented depositors, view cash as the main alternative transaction medium. Since cash bears no interests, as long as commercial banks pay some interests, the transaction-oriented depositors will stay in commercial banks. This allows commercial banks to keep deposit rates relatively low and earn higher spreads between the rising lending rates and the stagnant deposit rates. In contrast, shadow banks have to raise their deposit rates together with the market interest rates. Otherwise, their yield-oriented clientele will switch to other higher yielding liquid assets such as short-term bonds. As a result, when the Fed raises interest rates, the difference in deposit rates between shadow and commercial banks widens, which induces some of the marginal depositors of commercial banks to switch to shadow banks. This could potentially explain why shadow banks expand their deposit creation when the Federal Reserve raises interest rates.

This explanation seems to be consistent with the data. Figure 2.3 plots the average deposit rates of commercial banks and MMFs over time. I find that when the Fed raises interest rates, shadow banks pass through more rate hikes to depositors than commercial banks do. The changes in relative deposit rates are economically significant. For example, in the 2004 tightening cycle, the difference in deposit rates increased from less than 0.5 percent to nearly 3 percent. Since transaction convenience and safeness of bank deposits are relatively stable over monetary cycles, such big changes in relative yields may significantly affect depositors' choice between these two banking sectors.

In what follows, I introduce the setting of the structural empirical model. The structural empirical model allows me to uncover the nature of competition in the deposit market using data of deposit rates and quantities. I will be able to estimate the extent of product differentiation between the two types of banks, and see how they relate to alternatives such as cash and bonds. I will also be able to estimate demand functions for each commercial bank and MMF, which shed light on how banks set deposit rates in response to changes of monetary policy. More importantly, the structural model can quantify the magnitude of proposed channel. This is crucial given that there are alternative explanations which give the same qualitative results. Lastly, the structural approach allows counterfactual simulations which are useful for examining policy implications.

2.3.2 Model Setting

Having shown the basic intuition that different clientele can lead to different responses to monetary policy, I now proceed with offering a full structural model to quantify the magnitude of this channel. The model uses the discrete choice framework of oligopoly competition developed by Berry, Levinsohn, and Pakes (1995). This framework, by defining consumer preference over product characteristics as oppose to actual products, can endogenously generate a demand system for a large set of differentiated products with only a small number of preference parameters. In the context of this paper, the demand for commercial and shadow bank deposits will be derived endogenously as function of their product characteristics, instead of being exogenously specified. The discrete choice framework has been successfully applied to many industries

such as the automobile, cereal, and airline industries. It has been a workhorse model in the quantitative IO literature over the past 20 years. Early applications of this framework to the commercial banking industry includes Adams, Brevoors, and Kiser (2007) and Ho and Ishii (2010). My paper contributes to this line of literature by introducing competition from shadow banks, a sector which has become increasingly important in the modern banking system. In addition, I use this framework to study the impact of monetary policy on the banking system, an aspect which has not yet been explored in this literature.

I first introduce the basic setup of the framework. It is useful to show how the structural model connects unobservable quantities such as utility and marginal costs to observable quantities such as market shares and deposit rates in the data. I then add two features that are important to the deposit market. Lastly, I estimate the model with the data.

2.3.3 Depositors

There are I depositors. Each of them is endowed with one dollar. Depositors make a discrete choice among options including Treasury bills, cash, commercial bank or shadow bank deposits. Each individual bank is a distinctive option as each provides a differentiated product. Each option is indexed by j, and the choice set is $\{0, 1, ..., J\}$. Each depositor can choose the option which gives him/her the highest utility. The assumptions that each depositor has only one dollar and can choose only one option are not as restrictive as they may appear. We can imagine that depositors make multiple discrete choices for each dollar that they have, and the probability of choosing each of the options can be interpreted as portfolio weights. The utility for depositor i to choose product j is given by

$$\max_{j \in \{0,1,...,J\}} u_{i,j} = \alpha r_j + \beta' x_j + \xi_j + \varepsilon_{i,j} \tag{2.2}$$

x_j is a vector of product characteristics of bank j. Examples of product characteristics of a commercial bank include branch density, number of employees, and the age of the bank. Examples of product characteristics of a MMF include a rating dummy indicating whether the MMF is rated by three major rating agencies, a bank fund dummy indicating whether the MMF is affiliated with a bank holding company, and a check-writing dummy indicating whether the MMF allows depositors to write checks. r_j is the deposit rate. ξ_j is an unobservable demand shock. $\varepsilon_{i,j}$ is a mean-zero idiosyncratic shock to utility, which follows the extreme value distribution with a probability density function $f(\varepsilon) = exp\{-exp(-\varepsilon)\}$. This distribution assumption is standard in structural IO literature. It basically allows closed-form solution of the choice probabilities. Finally, α and β are sensitivities to deposit rates and product characteristics. $\beta' x_j$ captures the convenience of holding deposits from bank j. I normalize the convenience of Treasury bills to 0. Notice that the linear form of utility does not mean that depositors have to be risk neutral. The aversion to risk can be incorporated as disutility to risk, similar to the mean-variance utility formation. For example, Egan, Hortaçsu, and Matvos (2015) use CDS spreads of banks as a proxy of risk. Since CDS spreads are not available for MMFs, in this study the riskiness of deposits depends on whether the deposits are insured by the deposit insurance, which further depends on whether the deposits are issued by a commercial bank or a shadow bank.

I define δ_j as the mean utility of product j across all depositors.

$$\delta_j = E[u_{i,j}] = \alpha r_j + \beta' x_j + \xi_j \tag{2.3}$$

Under the assumption that the idiosyncratic utility shock follows the extreme value distribution, the expected probability that product j is the best choice is given by the following formula:

$$s_j = E\left[\mathbb{1}_{\{u_{i,j} \geq u_{i,l} \forall l\}}\right] = \frac{\exp(\delta_j)}{\sum_{l=1}^{J} \exp(\delta_l)} \tag{2.4}$$

Notice that the expected probability that product j is the best choice for a depositor is also the market share of the product. Intuitively, the higher mean utility that a product generates, the greater market share that it has. In the baseline model, the market share is a simple logit function of the mean utility. Therefore, this model is often referred as "the logit model of demand" in the literature. Later, I will introduce features that are important to fit the deposit market.

2.3.4   Banks

A bank is represented by a vector of product characteristics. The differences in product characteristics between commercial and shadow banks are mainly driven by regulatory constraint. For example, shadow banks are not allowed to operate branches or to provide checking accounts because these activities require bank charters. I assume that product characteristics are exogenous, which leads to exogenous demand function $s_j(r_j)$. Facing the demand, the decision of bank j is to choose a deposit rate r_j to maximize profits

$$\max_{r_j} (f - r_j - c_j) s_j(r_j) \tag{2.5}$$

where f is the Fed Funds rates, r_j is the deposit rate of bank j, c_j is the cost of providing depository services. $s_j(r_j)$ is the market share of bank j. I assume that marginal lending rates of all the banks equal to the Fed Funds rates. This assumption is quite close to reality, since the inter-bank market is usually quite efficient to equalize the marginal lending rates of different banks.[10]

Banks' optimal pricing decision is given by the following FOC:

$$\text{FOC:}\ f - r_j = \left(\frac{\partial \log(s_j)}{\partial r_j} \right)^{-1} + c_j \tag{2.6}$$

On the left hand side, the spread between the Fed Funds rates and deposit rates, commonly refereed as deposit spread, represents the price that banks charge for their depository services. On the right hand side, the first term $\left(\frac{\partial \log(s_j)}{\partial r_j} \right)^{-1}$ is the markup that a bank can charge on its depository service over the cost of providing it. It is inversely related to the demand elasticity. If the demand is inelastic, then the bank can charge a higher markup. In contrast, if the demand is elastic, then the markup is likely to be low.

I specify the marginal cost as a linear function of cost shifters

$$c_j = \gamma' w_j + \omega_j \tag{2.7}$$

where w_j is a vector of observable supply shifters. Examples of supply shifters of a commercial bank include salary paid to employees and fixed asset expenses. Examples of supply shifters of a MMF include management costs and other operating costs. γ is the sensitivity of marginal cost to these cost shifters. ω_j is an idiosyncratic supply shock.

2.3.5   Equilibrium

The pure-strategy Bertrand-Nash equilibrium is a set of deposit rates, r^*, chosen by banks, and a set of product, j^*, chosen by depositors such that each bank maximizes its profits, each depositor maximizes their utility, and the deposit market clears.

The main difference between commercial and shadow banks arises from their demand functions. I do not make any a priori assumptions on how these demand functions should be different. Instead, the demand

---

[10]For instance, if bank A has higher lending rates than bank B, bank A will borrow from bank B and keep lending until the two lending rates converge.

functions and their differences will be determined endogenously in the equilibrium by the exogenous product characteristics and preference parameters which will be estimated from the data. In addition, the cost to produce shadow and commercial bank deposits may also be different, but as the estimation will show later, the demand side difference is the key reason why banks set deposit rates differently. In summary, to fully characterize the equilibrium, I need to know a set of primitive parameters, α, β, γ, which governs how depositors value different products and how much it costs to produce them.

Ideally, if I observe mean utility, δ_j, and marginal costs, c_j, I can pin down these parameters by estimating the following two equations.

$$\delta_j = \alpha r_j + \beta' x_j + \xi_j \tag{2.8}$$

$$c_j = \gamma' w_j + \omega_j \tag{2.9}$$

The first equation is the "mean utility equation" which describes how deposit rates and product characteristics are valued by depositors, and the second is the "marginal cost equation" which describes how observable cost shifters affect the marginal cost of providing depository services. The challenge here, however, is that neither mean utility, δ_j, nor marginal costs, c_j, are observable. Here is how the structural model can help. From the optimal decisions of depositors, I can link unobservable utility to observable market share. Using equation 2.4, I can solve unobservable mean utility as a closed-form function of observable market shares

$$\underbrace{\delta_j}_{\text{Unobservable}} = \underbrace{\log(s_j) - \log(s_0)}_{\text{Observable}} = \alpha r_j + \beta' x_j + \xi_j \tag{2.10}$$

From the optimal decisions of the bank (equation 2.6), I can solve unobservable marginal costs as the difference between deposit spreads and markups. Markups can be further derived from the market share equation 2.4 as a function of observable market shares and yield sensitivity, α, which can be estimated from the mean utility equation.

$$\underbrace{c_j}_{\text{Unobservable}} = f - r_j - \underbrace{\left(\frac{\partial \log(s_j)}{\partial r_j}\right)^{-1}}_{\text{Unobservable}} = f - r_j - \underbrace{(\alpha(1 - s_j))^{-1}}_{\text{Observable}} = \gamma' w_j + \omega_j \tag{2.11}$$

2.3.6 Depositor Heterogeneity and Adjustment Cost

What I have shown above is a basic setup of the discrete choice framework. Now I will add two features that are important to the deposit market. The first one is depositor heterogeneity. In the basic setup, depositors have homogeneous tastes over yields and convenience. In reality, depositors may exhibit strong heterogeneity, as evident in Table 2.3. As argued in Section 2.3.1, different clientele can lead to different exposure to monetary policy for the banks. As a result, it is important to incorporate this empirical feature in the model.

Second, there is considerable stickiness in the adjustment process of deposits. As shown in Figure 2.1 and 2.3, when shadow banks offer higher deposit rates than commercial banks, deposits do not switch into shadow banks immediately. Instead, deposits flow into shadow banks gradually. To capture this feature, I introduce a switching cost for depositors when they change their choices.

Formally, the depositors' problem is modeled as the following maximization problem

$$\max_{j \in \{0, 1, \dots, J\}} u_{i,j} = (\alpha + \sigma v_i) r_j + \beta' x_j + \xi_j - \rho \mathbb{1}_{\{k \neq j\}} + \varepsilon_{i,j} \tag{2.12}$$

σv_i captures the heterogeneous response to deposit rates, where v_i follows a standard normal distribution and σ is the magnitude of dispersion. $-\rho \mathbb{1}_{\{k \neq j\}}$ captures the adjust cost, where k is choice of last period and j is the choice of current period. This term means that if the current choice j is different from the previous choice k, the depositor incurs a cost of ρ.

Notice that such formulation of adjustment cost assumes that depositors are myopic: they simply make trade-off between the contemporaneous utility and adjustment costs instead of trying to forecast future path of interest rates. Ideally, it would be more realistic to have forward-looking depositors. However, such dynamic model with heterogeneous agents would be very difficult to estimate. Therefore, this paper takes a more reduced form approach to model adjustment cost.

With the adjustment cost, the past choice k becomes a state variable for the current decision. Define $P_{i,t}^{k \to j}$ as the transition probability from product k to j for depositor i at time t. The transition probability is given by the following formula:

$$P_{i,t}^{k \to j} = E\left[u_{i,j} \geq u_{i,l} \forall l | k\right] = \frac{\exp\left(\delta_j + r_j \sigma v_i - \rho \mathbb{1}_{\{k \neq j\}}\right)}{\sum_{l=1}^{J} \exp\left(\delta_l + r_l \sigma v_i - \rho \mathbb{1}_{\{k \neq l\}}\right)} \tag{2.13}$$

Define $s_{i,j,t}$ as the expected choice probability for depositor i to choose product j. The current expected choice probability is the sum of the products between transition probability, $P_{i,t}^{k \to j}$, and the probability distribution of last period, $s_{i,k,t-1}$.

$$s_{i,j,t} = \sum_k s_{i,k,t-1} P_{i,t}^{k \to j} \tag{2.14}$$

The aggregate market share of product j is the obtained by summing over different depositor types

$$s_{j,t} = \sum_i \pi_i s_{i,j,t} \tag{2.15}$$

where π_i is the frequency of type i depositors.

Note that after introducing adjustment costs, the banks' problem become dynamic.

$$\max_{r_{j,t}} E \sum_{t=0}^{\infty} \exp(-rt)\left(f_t - r_{j,t} - c_{j,t}\right) s_{j,t} \tag{2.16}$$

where r is the discount rate for the profits.

With adjustment costs for depositors, the optimal deposit spread is determined by long-run elasticity instead of short-run elasticity.

$$f_t - r_{j,t} = \left(E\left[\sum_{\tau=0}^{\infty} \exp(-r\tau) \frac{\partial s_{j,t+\tau}}{\partial r_{j,t}} \frac{1}{s_{j,t}}\right]\right)^{-1} + c_{j,t} \tag{2.17}$$

[11]

One complication introduced by these two new features is that I can no longer solve mean utility, δ_j, as a closed form function of market shares. Instead, I need to solve it numerically from the market share equation 2.15 using the fixed-point algorithm introduced by Berry, Levinsohn, and Pakes (1995).

2.4 Structural Estimation

In this section, I take the model to the data. The goal here is to pin down the primitive structural parameters and quantify effects of depositor heterogeneity on banks' response to monetary policy. This will set the stage for the counterfactual analysis that ensues.

---

[11]The above deposit rates are solved in a stationary equilibrium where the state variables remain constant over time so that $f_{t+\tau} - r_{j,t+\tau} - c_{j,t+\tau} = f_{t+\tau} - r_j - c_j$.

2.4.1 Identification

The set of primitive parameters are $\alpha, \beta, \sigma, \rho, \gamma$. Some of the parameters enter linearly (α, β, γ) in the two structural equations below, and the rest enter non-linearly (σ, ρ) (the time subscript is suppressed here for simplicity).

$$\delta_j(\sigma, \rho) = \alpha r_j + \beta' x_j + \xi_j \tag{2.18}$$

$$c_j = \gamma' w_j + \omega_j \tag{2.19}$$

Two alternative procedures can be implemented to estimate the above structural equations. I can estimate them simultaneously using a joint-equation GMM, or I can estimate them sequentially. I choose the sequential approach, since in a joint estimation the misspecification of the marginal cost equation may contaminate the estimation of preference parameters. More concretely, I first estimate the mean utility equation (2.18). Given the estimated demand side parameters, I calculate the marginal costs implied by equation (2.17). Then I estimate the marginal cost equation (2.19).

Since σ and ρ enter the mean utility equation non-linearly, the estimation of the mean utility equation requires more discussion. I use the Nested Fixed Point (NFP) algorithm as detailed in Nevo (2000). The algorithm first searches over the non-linear parameter space of σ and ρ. Second, for a combination of σ and ρ, it solves $\delta_j(\sigma, \rho)$ through fixed-point algorithm using equation (2.15). Third, I find a set of linear parameters α, β which minimize the moment condition which I will describe later. The above three steps are repeated until the optimal set of parameters $\alpha, \beta, \sigma, \rho$ is found.

A key challenge in identifying the demand parameters is that deposit rates are correlated with unobservable demand shocks ξ_j. As a result, yield sensitivity α will be biased in an OLS regression of mean utility, δ_j, on deposit rates, r_j. The solution is to find shocks that affect deposit rates, but are exogenous to unobservable demand shocks, ξ_j. I follow the literature to use a set of cost shocks, z_j, as instrument variables. Examples of instrument variables include salary, rent, and other operating costs. The intuition is that these shocks shift the supply curve so that the demand curve can be traced out.

The moment condition of the mean utility equation is given by the orthogonality condition between the unobservable demand shocks, ξ_j, and the product characteristics, x_j, and cost shifters, z_j:

$$E\left[\xi_j(x_j, z_j)\right] = 0 \tag{2.20}$$

The moment condition of the cost equation is given by the orthogonality condition between the idiosyncratic supply shock, ω_j, and observable cost shifters, w_j:

$$E\left[\omega_j w_j\right] = 0 \tag{2.21}$$

2.4.2 Data for Structural Estimation

The data used for the structural estimation are a panel of commercial banks and MMFs from 1994 to 2012. Following the literature, a market is defined as a MSA-year combination. Since commercial banks attract deposits mainly through local branches, the choice set of depositors of a MSA includes commercial banks which have local branches in the MSA. In contrast, MMFs generally compete in a national market through distribution channels of brokerage firms and over the internet. Therefore, local depositors can access to all the MMFs in the market. In addition, depositors can also choose cash or Treasury bonds.

I calculate market shares of a commercial bank by summing up deposits of local branches of the bank in the MSA. For MMFs, cash, and Treasury bonds, no MSA-level information on quantities is available. I impute MSA-level deposit amounts assuming that deposit amounts are proportional to local personal income

level.[12] The total market size is the sum of cash, commercial bank deposits, MMF shares and Treasury bonds in a MSA.[13] Following the literature, I combine tiny banks and MMFs (market share less than 0.2 percent) with Treasury bonds as the outside option.

Product characteristics are chosen based on the belief that they are important and recognizable to depositors' choice. Product characteristics of commercial banks include deposit rates, branch density in the local market, average number of employees in a branch, bank age, and single-market dummy (whether a bank operates in a single market or multiple markets).[14] Product characteristics of MMFs include deposit rates, rating dummy (whether a fund is rated by three major rating agencies), bank fund dummy (whether a fund is affiliated with a commercial bank), check-writing dummy (whether a fund allows depositors to write a check), and fund age. I include product fixed effects to absorb unobservable time-invariant product characteristics. Notice that bank fixed effects also absorb observable time-invariant product characteristics. To retrieve the taste coefficients on these product characteristics, I follow the minimum-distance procedure proposed by Chamberlain (1982) to estimate coefficients of time-invariant product characteristics. Lastly, I include time fixed effects to absorb aggregate demand shocks, and MSA fixed effects to absorb cross-market differences in demand.

The marginal cost equation includes product characteristics and cost shifters. The set of product characteristics is the same as the demand function. The cost shifters of MMFs include management costs and other operating costs. The cost shifters of commercial banks include salary expenses and expenses of fixed assets.[15] Lastly, I include bank fixed effects to absorb time-invariant bank-specific cost shocks, time fixed effects to absorb aggregate shocks to marginal costs, and MSA fixed effects to absorb cross-market differences in the cost of providing depository services.

As discussed in 2.4.1, I need a set of instruments to identify the yield sensitivity, α. Following previous literature, I use the cost shifters and their second-order polynomials as instruments for the mean utility function. I use Chamberlain's (1987) optimal instruments in the second stage of estimation to increase the estimator's efficiency and stability (Reynaert and Verboven, 2014). The optimal instruments are defined as the conditional expectation of the derivatives of the residuals with respect to the parameter vector. The details of constructing the optimal instruments can be found in Reynaert and Verboven (2014).

Table 2.1 provides summary statistics of the sample used for the structural estimation. A commercial bank typically has larger market shares than a MMF: the average market share is 3.5 percent for a commercial bank and is 0.44 percent for a MMF. A commercial bank also tends to offer lower deposit rates than shadow banks: the average deposit rates are 1.79 percent for commercial banks and 3.04 percent for MMFs. A commercial bank on average has 22.5 branches per million population in a MSA, and each branch has 18.45 employees. 52 percent of MMFs are rated by a least one of three rating agencies, 45 percent are affiliated with bank holding companies, 34 percent allow depositors to write checks.

---

[12]More specifically, I first compute the percentage of a MSA personal income in the national total personal income. Then I calculate the historical average of this percentage over the sample period to calculate the weight of the MSA. Lastly, I impute the MSA-level deposits according to the weight. Alternatively, I estimate the model using national-level data. This alternative approach generates similar results.

[13]Treasury bills are more appropriate for the model setting. However, the information of the aggregate Treasury bills outstanding is not always available in the sample period.

[14]For commercial banks which operate in multiple markets, I only have bank-level rather than branch-level information on deposit rates, so there is no cross-market variation for these multi-market banks. Nevertheless, this may not be a major issue since previous empirical studies have shown multi-market banks usually use uniform pricing across local markets within a state (Radecki, 1998).

[15]This set of cost shifters of commercial banks are also used in previous literature such as Dick (2008) and Ho and Ishii (2011).

2.4.3 Parameter Estimates

This section presents the results of the structural estimation. First, it discusses the demand parameters. Column 1 of Table 2.4 reports the estimates of demand parameters of the full model. The estimated yield sensitivity are positive and significant. Most importantly, there is statistically significant dispersion in depositors' sensitivity to yields. Later, I will explore the economic implications of such dispersion. Depositors also face a significant adjustment cost: the pecuniary cost of changing choices for an average depositor amounts to 8.81 percent. Depositors prefer banks with higher branch density and more employees per branch. Depositors also prefer single-market banks and younger banks. For MMFs, depositors prefer old funds and funds sponsored by bank holding companies. Interestingly, funds with check-writing privilege and credit rating do not gain favor among depositors on average.

Column 2 of Table 2.4 shows a simple logit model in which depositors are homogeneous and face no adjustment costs. One can see that depositor heterogeneity and adjustment costs are important to get the correct sign for the yield sensitivity: without these two features, the model implies that depositors counter-intuitively derive negative utility from deposit rates.

Regarding the supply side parameters, column 1 of Table 2.5 presents the estimated cost coefficients of the full model. The cost coefficients for MMFs are precisely estimated: a 1 percent increase in management costs is associated with 1.069 percent increase in marginal costs, and a 1 percent increase in other operating cost is associated with 1.151 percent increase in marginal costs. These magnitude is very close to the theoretical value, which is 1. For commercial banks, greater branch density and more employees per branch are associated with lower marginal costs, implying that there is an increasing return to scale. Higher salary expense is associated with higher marginal costs, which is intuitive. The estimated coefficient of expense of fixed assets, however, has counter-intuitive sign. This could be due to the fact that expense of fixed assets is quite sticky over time. It is difficult to identify its impacts when bank fixed effects are included in the regression. The reserve cost is the product between the Fed Funds rates and the amount of reserves normalized by total deposits. This variable captures the effect of monetary policy on the cost of providing deposits for commercial banks. Theoretically, a 1 percent increase in the reserve cost should lead to a 1 percent increase in the marginal cost. The estimated coefficient of the reserve cost is 1.765, which is close to the theoretical value.[16] Although the positive coefficient of reserve cost shows that monetary tightening indeed increases the marginal costs of commercial banks relative to shadow banks, the magnitude of this channel is likely to be small. The summary statistics in Table 2.1 show that commercial banks on average hold 1.1 cents of reserve per dollar of deposits. This tiny amount of reserve will have limited quantitative impacts on marginal costs of commercial banks. Lastly, the interaction between the Fed Funds rates and cash dummy has a coefficient of 1, implying that the cost of holding cash increases one for one with the Fed Funds rates.

Column 2 of Table 2.5 presents the estimated cost coefficients of the logit model. In this regression, the cost function has exactly the same specification as column 1. The only difference is that the dependent variable, marginal cost, is calculated from a logit model of demand. Comparing with the full model, the logit model implies counter-intuitive signs and magnitudes of these cost shifters. For example, higher management costs are associated with lower marginal costs. These results imply that the logit model does not generate as good estimates of the markups as the full model. This again confirms the importance to incorporate depositor heterogeneity and adjustment costs in the model.

2.4.4 Model Fit

Figure 2.4 compares model predicted market shares with the data. The full model successfully generates the counter-cyclical deposit growth rates for commercial banks, and pro-cyclical deposit growth rates for

---

[16]The reason why it is greater than 1 could be that it picks up other transmission channels not specified by the supply equation.

shadow banks. The magnitude matches the data closely.

What drives the different responses to monetary policy for commercial and shadow banks? In the following subsection, I will use the structural model to understand the underlying mechanism. The proposed mechanism consists of the following four steps:

1. Banks offer differential products. Shadow banks provide lower convenience than commercial banks.

2. Different products attract different depositor clientele. Shadow banks attract more yield-oriented depositors than commercial banks.

3. Facing different depositor clientele, banks have different interest rate pass-through. Shadow bank deposit rates are more sensitive to market interest rates than commercial bank deposit rates.

4. Different types of depositors respond to changes in relative prices differently. Yield-oriented depositors move in and out shadow banks, while transaction-oriented depositors stay with commercial banks.

In the following analysis, I examine the above four steps one by one.

2.4.5 Differential Products

The underlying assumption of the proposed channel is that commercial bank deposits offer greater convenience than MMFs. In this subsection, I verify this assumption. I construct a composite measure of convenience by $x'_j \hat{\beta}$. x_j is the vector of bank characteristics which is related to transaction services. Example of these characteristics includes the density of branches, number of staffs in a branch, check-writing privileges and so on. I also include bank fixed effects to absorb residual time-invariant difference in transaction services. $\hat{\beta}$ is the vector of estimated sensitivity to these characteristics. This composite measure shows how much the average depositor values each product if deposit rates are zero.

In Figure 2.5, I plot the distribution of the estimated convenience. Each observation is a MSA median for each sector. Consistently with the assumption that commercial bank deposits are safer and offer more transaction services, commercial banks have higher estimated convenience than MMFs. In addition, cash has the highest convenience, which is also intuitive. The figure also shows the relative market position of commercial and shadow bank deposits in relation to cash. Commercial bank deposits are a closer substitute to cash than shadow bank deposits. This could potentially allow commercial banks to attract a group of transaction-oriented depositors, a conjecture to be examined in the next section.

2.4.6 Depositor Clientele

With commercial and shadow banks offering differentiated products, I expect different types of depositors to self-select into different types of banks. The estimates show that this is indeed the case. Figure 2.6 plots average demand elasticities of commercial banks and shadow banks. The detailed summary statistics of demand elasticities can be found in Table 2.6. The median own-rate elasticity of commercial banks is 1.0668, which has the same magnitude as previous literature. The median own-rate elasticity of MMFs is 1.8112, which is almost twice as large as that of commercial banks. This is consistent with the idea that the clientele of MMFs is more yield-sensitive than commercial banks.

Next, I examine the cross-rate demand elasticity. The cross-rate elasticity measures the percent change of market shares due to changes in deposit rates of a competitor. Table 2.7 presents the median and standard deviation of cross-rate elasticity. The entry of the i-th row and j-th column shows the percent change of the market share of a product in category i (cash, CB, MMF) with one percent change of the deposit rates of a rival product in category j (cash, CB, MMF). Comparing across columns, price changes of a product with high convenience seem to have greater effects on other products. Comparing across rows, MMFs in general have the highest cross-rate elasticities, which is also likely a result of their yield-sensitive clientele.

2.4.7 Banks' Responses to Monetary Policy

I have shown above that shadow and commercial banks face different clientele. Now the question is whether the difference in clientele can lead to different responses to monetary policy. Figure 2.7 plots the average difference in deposit rates between the two banking sectors in the data and predicted by the model. The model generates similar pro-cyclical patterns as in the data: as the Fed Funds rates go up, deposit rates of shadow banks become higher than that of commercial banks.

Two possible mechanisms that may explain the above pattern for deposit rates. The first possibility is that monetary policy has differential impacts on the demand side of the banks. As the Fed Funds rates increase, commercial banks are able to charge higher markups on the transaction-oriented depositors, while the markups charged by shadow banks remain stable because their depositors are more yield-sensitive. The second possibility is that monetary policy has differential impacts on the costs of providing depository services. To elaborate, commercial banks are required to hold reserves while MMFs do not. High Fed Funds rates drive up the cost of holding reserves. Commercial banks do not increase their deposit rates because they have higher costs to cover. To separate the two channels, I decompose deposit spreads into two components: marginal costs and markups. The top panel of Figure 2.8 shows the average difference in markups and marginal costs between commercial and shadow banking sectors over time. It is clear that the markup term fully drives variations in deposit spreads. In contrast, the difference in marginal costs is almost flat over monetary cycles. This figure shows that the differential pricing pattern is mainly driven by the demand side, rather than the supply side. It is hardly surprising, given the summary statistics in Table 2.1 showing that commercial banks on average only hold 1.1 cents of reserves for every dollar of deposits. A 1 percent increase in the Fed Funds rates only leads to a 0.011 percent increase in marginal costs.

What is the key demand parameter that generates the different cyclical patterns of markups? In the bottom panel of Figure 2.8, I re-estimate the markups and marginal costs assuming that the dispersion in yield sensitivity, σ, is zero. In this case, the markup term becomes flat as well. This result shows that depositor heterogeneity is the key feature to explain the different pricing patterns in the data. The intuition is that when σ goes to zero, depositors become essentially homogeneous. This means that commercial and shadow banks will have the same clientele and monetary policy will have similar impacts on their demand elasticity.

The decomposition of deposit spreads into markups and marginal costs also sheds lights on the monetary transmission mechanisms in the commercial banking sector. Traditionally, the banking system is modeled as a perfectly competitive industry. There is no role for market power because markups are always zero. It is until recently that several papers such as Scharfstein and Sunderam (2016) and Drechsler, Schnabl, and Savov (2016) start to point out that market power of the banking sector may play a role in transmitting monetary policy. Since the aforementioned papers rely on reduced-form models, they cannot quantify the importance of the market power channel relative to the traditional bank reserve channel. This paper complements the above studies by providing a structural model to quantify the impact of market power and reserve cost. As it is clear from the top panel of Figure 2.8, the markups are main sources of variations in the deposit spreads of commercial banks. This evidence suggests that the market power channel has been playing a dominant role in the transmission of monetary policy since the 1990's. In contrast, the reserve channel has become less relevant due to technological and regulatory changes (Teles and Zhu, 2005). In addition to quantifying the magnitude of market power channel, my model also provides a deeper understanding on the determinants of market power. It shows that market positioning affects the pricing power of banks. This helps us understanding why monetary policy may have heterogeneous impacts on different types of banks.

2.4.8 Choice of Depositors

Lastly, I examine the choices of different types of depositors over monetary cycles. I classify depositors with above-median yield sensitivity as yield-oriented depositors, and depositors with below-median yield sensitivity as transaction-oriented depositors. Figure 2.9 plots their probability to choose a commercial bank or a MMF over time.

The first observation is that yield-sensitive depositors are on average more likely to choose MMFs, while transaction-oriented depositors are more likely to choose commercial banks. The second observation is that the choice probability of yield-oriented depositors varies significantly over monetary cycles: they are more likely to choose commercial banks when the Fed Funds rates are low, and switch to MMFs when the Fed Funds rates go up. In contrast, transaction-oriented depositors are more likely to choose commercial banks all the time. This is consistent with the intuition that yield-oriented depositors are constantly looking for higher yields, while transaction-oriented depositors stay in commercial banks because of the transaction convenience and safety of commercial bank deposits.

2.4.9 Alternative Explanations

Thus far, I have shown that the model produces coherent evidence that depositor heterogeneity can quantitatively explain the different responses to monetary policy by commercial and shadow banks. One may argue that there are many other institutional differences across banking sectors could also possibly explain their different responses to monetary policy. One intuitive candidate is the reserve requirement. When commercial banks take deposits, they are required to keep a fraction of the deposits as reserves instead of lending them out. Historically, bank reserves do not bear interests.[17] Therefore, holding reserves imposes a cost for commercial banks, and the cost of holding reserves is increasing in the Fed Funds rates. In contrast, shadow banks are not subject to reserve requirements. As a result, monetary policy may have differential impacts across banking sectors through the cost of providing depository services. The bank reserve channel features underlying mechanism of several papers such as Kashyap and Stein (1995), Stein (2012), Sunderam (2015) and Nagel (2016).

The reserve based explanation does not seem to quantitatively explain the magnitude of pricing difference documented in this paper. To do a back-of-envelope calculation, I assume that 10 percent reserve requirement applies to all commercial bank deposits[18]. In the 2004 tightening cycle, the Fed Funds rates increase by 4.25 percent, which leads to an increase in the marginal cost by 0.425 percent through the reserve channel. This number is still far from explaining the 2.5 percent increase in the data. The structural model provides more concrete evidence. In the top panel of Figure 2.8, we hardly see any differential impacts of monetary policy on the marginal costs of commercial and shadow banks, despite of their different reserve requirement. This result is not surprising given extensive research has suggested that reserve requirement has become less relevant for the current banking system due to technological innovations and regulatory reforms.[19]

The second potential explanation is based on asset-side differences between commercial banks and MMFs. The asset duration of MMFs are much shorter than commercial banks for both economic and

---

[17]After October 2008, the Fed started to pay interest on reserves.

[18]In practice, saving deposits face much less reserve requirement (1 percent), which further reduce the magnitude of this channel.

[19]One example of technological innovations is the sweep technology, which allows banks to easily transfer funds from transaction accounts to saving accounts to avoid the reserve requirement (Teles and Zhu, 2005). As a result, the amount of bank reserve in the economy has become very small before the recent unconventional monetary policy: as of December 31, 2007, the aggregate reserve balance is only 48 billion, which accounts for less than 0.4 percent of 6,720 billion commercial bank deposits. It is hard to imagine such a small opportunity cost could quantitatively explain the substantial deposit spreads observed in the data. After the start of unconventional monetary policy in 2008, the reserve balance grew dramatically. However, in this period, the Fed started to pay interest on reserves, which essentially eliminated this reserve channel.

regulatory reasons. Economically, the shadow banking system breaks down the intermediation process in several steps. MMFs only conduct a small amount of maturity transformation: the average maturity of MMF assets is around 40 days based on the iMoneyNet data, while commercial banks usually have much longer asset maturity. In terms of regulation, Rule 2a-7 of the Investment Company Act of 1940 restricts the highest maturity of any debt held by MMFs to be under 13 months, and the portfolio must maintain a weighted average maturity (WAM) of 60 days or less. Therefore, a change in interest rates may lead to different impacts on the value of the assets due to different asset duration. However, this channel is only relevant for the periods shortly after interest rate changes. It cannot explain the persistent differences in deposit rates between MMFs and commercial banks long after the change of the Fed Funds rates.

The third potential explanation relies on the risk of holding shadow bank deposits. The risk of holding shadow bank deposits is likely to be lower when the economy is doing well. This is usually the time when the Fed adopts high interest rate policy to cool down the economy. Therefore, the inflows to the shadow banking sector may be driven by lower risks of shadow banks, rather than by monetary policy. Although this explanation can potentially explain the inflows to the shadow banking sector, it cannot explain why shadow banks usually pay higher deposit rates in periods of high interest rates. If anything, this explanation would predict that shadow banks should pay lower deposit rates because their risks are lower during these periods.

2.5 Policy Implications

Using the structural model, I conduct a set of counterfactual exercises to study several questions relevant to shadow banking. How does shadow banking change the effectiveness of monetary policy? What are the implications of shadow banking for financial stability? How does shadow banking affect depositor surplus?

2.5.1 Shadow Banks and Effectiveness of Monetary Policy

There is a long-lasting concern that financial innovation may undermine monetary control of the central bank. Such concern has intensified in recent years as the shadow banking sector has grown outside the traditional commercial banking sector. Has the rise of the shadow banking system affected the effectiveness of monetary policy? To answer this question, I calculate the aggregate money supply in the counterfactual economy without MMFs. Figure 2.10 shows the growth rates of aggregate money supply in the counterfactual economy. In a simple regression of the aggregate money growth rates on the Fed Funds rates controlling for a time trend, the coefficient of the Fed Funds rates decreases by around 40 percent in the actual economy comparing to the counterfactual one. This means that the presence of shadow banks reduces the responsiveness of aggregate money supply to monetary policy.

The counterfactual analysis offers insights on the monetary transmission mechanisms in an economy with both commercial banks and shadow banks. In an economy without shadow banks, commercial banks are more reluctant to pass through the rate increase to depositors. Depositors flow out of the banking system, leading to a reduction in money supply and credit supply.

Shadow banks provide buffers for depositors. When depositors are unsatisfied with the low rates paid by commercial banks in periods of monetary tightening, they do not have to switch from deposits to bonds. They can switch within the banking system from commercial banks to MMFs. Having more loanable funds, MMFs pass the proceeds to other shadow banks which specialize in lending. An increase in shadow bank lending can substitute the decline in commercial bank lending, which dampens the impact of monetary tightening.

2.5.2 Monetary Policy and Credit Supply of Shadow Banks

So far, my empirical analysis has been focusing on the money supply. This section examines the credit supply of shadow banks. As discussed in Section 2.2, while shadow bank deposit creation is conducted by MMFs, loan origination is conducted by different shadow banking entities such as funding corporations, finance companies, ABCP issuers, captive financial institution and broker-dealers.[20] These loan-origination shadow banks do not issue deposits directly. Instead, they obtain funding from MMFs through issuing money market instruments. There are four major categories of money market instruments issued by these loan-origination shadow banks: commercial papers (CPs), asset-backed commercial papers (ABCPs), repurchase agreements (repos), and floating rates notes (FRNs). I regress annual changes of MMF lending through these four money market instruments on the Fed Funds rates, controlling for macroeconomic variables, fund characteristics and fund fixed effects:

$$\Delta MMF\ Lending_{i,t} = \alpha + \beta Fed\ Funds\ Rates_t + \gamma X_{i,t} + \varepsilon_{i,t} \tag{2.22}$$

Column 1 to 4 of Table 2.8 show that MMFs significantly increase their lending to the loan-origination shadow banks as the Fed Fund rates increase. The economic magnitude is significant, too: a 1 percent increase in the Fed Fund rates is associated with a 0.17-0.45 percent increase in lending from MMFs to other shadow banks.

In addition to the four types of money market instruments discussed above, MMFs also hold large denomination commercial bank obligations (CBs), which are issued by commercial banks to obtain short-term funding. Column 6 of Table 2.8 shows that MMFs also increase the holding of large denomination bank obligations when the Fed raise interest rates. This result reveals an interesting interaction between the shadow and commercial banking system. As the Fed tightens monetary policy, commercial banks borrow more from MMFs to compensate their loss of the core deposits. Such arrangement is profitable for both types of banks: it effectively conducts price discrimination on transaction-oriented depositors. However, it has a downside: through this lending relationship, bank runs on the MMF industry may spread to commercial banks. This result comprises another unintended consequence of monetary tightening on financial stability.

With an increase in funding supply from MMFs, the loan-origination shadow banks should be able to expand their credit supply. I examine five types of shadow banks which rely on MMFs to obtain financing: funding corporations, finance companies, ABCP issuers, captive financial institution and broker-dealers. I regress aggregate asset growth rates of these five types of shadow banks on the Fed Funds rates and various macroeconomic controls:

$$Shadow\ Bank\ Asset\ Growth_t = \alpha + \beta Fed\ Funds\ Rates_t + \gamma X_t + \varepsilon_t \tag{2.23}$$

Table 2.9 presents the results. When the Fed Funds rates are high, the assets of these shadow banks also grow faster. The composition shift in the aggregate credit supply may also increase the systemic risk,

---

[20]Finance companies are financial entities that sell commercial paper and use the proceeds to extend credit to borrowers which usually tend to be riskier than that of commercial banks (Carey, Post, and Sharpe, 1998). In the mortgage market, these shadow lenders such as Quicken Loans, PHH and loanDepot.com accounted for 53 per cent of government-backed mortgages originated in April, 2015. Funding corporations are subsidiaries of foreign banks and non-bank financial firms that raise funds from the commercial paper market and pass the proceeds to foreign parent companies abroad or to foreign banking offices in the U.S.. ABCP issuers are structured investment vehicles which purchase and hold financial assets from a variety of asset sellers and finance their portfolio by selling asset-backed commercial paper to MMFs or other "safe asset" investors like retirement funds. Captive finance company is a subsidiary whose purpose is to provide financing to customers buying the parent company's product through issuing commercial papers. Examples include the captive finance of the Big Three car manufacturers: General Motors Acceptance Corporation (GMAC), Chrysler Financial and Ford Motor Credit Company. Broker-dealers include both non-bank firms and subsidiaries of commercial banks that engages in the business of trading securities for its own account or on behalf of its customers. Broker-dealers heavily rely on repo to obtain funds from MMFs and then lend to their customers through reverse repo. A prominent example of broker-dealers is Lehman Brothers which went bankrupt during the 2008-09 financial crisis.

because shadow banks usually lend to the riskier segment of borrower (Carey, Post, and Sharpe, 1998). The positive relation between shadow bank asset growth rates and the Fed Funds rates is also documented by a contemporaneous paper by Nelson, Pinter, and Theodoridis (2015). The main difference between their work and the present study is that they attribute the expansion of shadow bank assets to negative shocks of high interest rate policy on equity values of commercial banks, while my paper argues that the expansion of shadow bank assets is driven by the increase in shadow bank deposit creation. Empirically, their assumption that high interest rate policy reduces equity values of commercial banks seems to be inconsistent with the data, as stock prices of commercial banks usually increase when the Fed raises interest rates. However, the increase in stock prices of commercial banks during periods of high interest rates is consistent with my model, which shows that high interest rate policy widens the spread between lending rates and deposit rates, which boosts the profitability of banks.

2.5.3 Depositor Clientele and Risks of Bank Run

The shadow banking system played a central role in the 2008-09 financial crisis. Why is the shadow banking system so fragile? Previous literature has pointed out factors such as the lack of deposit insurance (Gorton and Metrick, 2012), high leverage (Adrian and Shin, 2010; Moreira and Savov, 2016), and information opacity (Dang, Gorton, and Holmström, 2016). In this section, I explore whether the yield-sensitive clientele could be a source of fragility for the shadow banking system.

The runs on MMFs in September 2008 provide a unique laboratory to study this question. On September 16th, 2008, the Reserve Primary Fund, one of the oldest MMF, broke the buck when its net asset value fell below $1. This event triggered widespread runs on the whole prime MMF industry. By October 7th , 2008, deposits of prime MMFs fell by $498 billion (24 percent).[21] Interestingly, there was significant heterogeneity in the severity of runs across funds: some funds suffered enormous amount of withdrawal, while others were less affected.

I examine whether the severity of the runs is related to the clientele of the funds. I use demand elasticity of the largest 30 prime funds to predict the severity of the runs in the following month. The demand elasticity is estimated using my structural model as of September 9th, 2008, one week before the runs. Figure 2.11 shows the result. Funds with higher demand elasticity experienced greater redemption subsequently. Why did funds with higher demand elasticity suffer more severe runs? A fund with higher demand elasticity may face greater redemption for a given loss. Because of the lack of deposit insurance, the initial redemption may force the fund to liquidate its assets in fire sale, imposing negative externality on the remaining depositors and causing further redemption.[22] This means that the yield-sensitive clientele served by shadow banks could be a source of fragility in the shadow banking sector.

The above results also contribute to a recent debate on the merit of using monetary policy as a macro-prudential policy tool. After the 2008-2009 financial crisis, many argue that monetary tightening should be used to promote financial stability, because high interest rates can slow down deposit creation in the banking system and reduce risk appetite of investors (Stein, 2012; Borio and Zhu, 2012; Williams, 2014; Smets, 2016). My findings suggest a cautious stance towards this policy proposal. I show that while monetary tightening can reduce deposit creation in the commercial banking sector, it may unintentionally drive more yield-sensitive depositors to the shadow banking sector as shown in Figure 2.9. Since shadow bank deposits are not insured by the government, a higher concentration of yield-sensitive depositors may amply the risk

---

[21]MMFs are usually categorized into three types based on investment strategy: prime, Treasury and Tax-exempt. Prime MMFs can invest in private debts, while Treasury and Tax-exempt only invest in government securities. During the 2008 crisis, prime funds were the ones that exposed to losses of other shadow banks and suffered the runs. Therefore, my analysis focuses on prime MMFs.

[22]From this perspective, a recent rule change adopted by the SEC which requires all prime MMFs to float their share prices may help to reduce the risk of bank runs, because it reduces the negative externality of early redemption on remaining depositors.

of bank runs. Therefore, this policy may increase the fragility of the system, rather than reducing it.[23]

In summary, my paper supports the view that "monetary policy is too blunt a tool to address possible financial imbalances" as expressed by Bernanke (2011) and Yellen (2014). The modern banking system is extremely complex. It is very difficult to fine-tune monetary policy to address financial stability concerns, because different institutions may have very different responses to monetary policy. Instead, we should use macro-prudential regulations, which can be targeted to specific institutions.

2.5.4 Implication of Shadow Banking for Depositor Surplus

Commercial banks have considerable market power in local depository markets. The entry of shadow banks may increase rate competition in the deposit market and potentially bring significant gain in depositor surplus. To access the impact of shadow banking on depositor surplus, I use the estimated structural model to simulate a counterfactual economy with no MMFs. I solve deposit rates and market shares of commercial banks in this counterfactual economy and calculate depositor surplus according to the new set of choices and prices. In absence of MMFs, commercial banks pay slightly lower deposit rates (9 basis points) but gain much greater market shares (3.86 percent). I follow Nevo (2001) to estimate the gain for depositor surplus from the entry of MMFs. I first compute the expected utility for each type of depositor i from its optimal choice.

$$E\left[\max_{j\in\{0,1,...J\}} u_{i,k,j} \mid k\right] = \ln\left(\sum_{j=0}^{J} \exp\left(\delta_j + r_j\sigma v_i - \rho \mathbb{1}_{\{k\neq j\}}\right)\right) \tag{2.24}$$

Then, I divide expected utility by the yield sensitivity to calculate the equivalent utility in the unit of deposit rates. Lastly, I sum over past choices and depositor types to calculate the aggregate surplus.

$$\text{Depositor Surplus}_t = \sum_i \pi_i \sum_k s_{i,k,t-1} \frac{1}{\alpha_i} E\left[\max_{j\in\{0,1,...J\}} u_{i,k,j} \mid k\right] \tag{2.25}$$

I compare the surplus in the counterfactual economy with the actual economy. The entry of shadow banks on average generates 0.36 cents of a dollar per year in the sample period. This amounts to 50 billions increase in depositor surplus with an aggregate money supply of 14 trillions at the end of 2015. The change in depositor surplus has the same magnitude as national branching deregulation in the 1990s estimated by Dick (2008), which is estimated to be 0.50 cents of a dollar. I further examine the time-series variation of the change in depositor surplus, which is plotted in Figure 2.12. The change in depositor surplus is larger when the Fed Funds rates are high, which is consistent with the previous result that commercial banks enjoy greater market power during these periods.

2.6 Conclusion

This paper documents a new monetary transmission mechanism: the shadow bank channel. I find that shadow bank money supply expands when the Fed raises interest rates. This is at odds with the conventional wisdom in the commercial banking sector that monetary tightening reduces deposit creation. I show that the different clientele served by shadow banks can explain their different responses to monetary policy by using a structural model of bank competition. Fitting my model to institution-level commercial bank and money market fund data shows that this channel reduces the impact of monetary policy on money supply by 40 percent. The macro-prudential implications of shadow banking are also explored.

---

[23]On the contrary, using ultra-low interest rates to squeeze shadow banks may not be a good policy either. As interest rates go down, MMFs lose deposits to commercial banks. Since MMFs need to maintain a minimum scale to cover fixed operating costs, they may take excessive risk to increase their yields. This conjecture is supported by empirical evidence in Di Maggio and Kacperczyk (2016).

This paper highlights the complexity of the current banking system. The rise of the shadow banking sector may have fundamentally changed the structure of the U.S. banking system and the transmission mechanisms of monetary policy. This paper provides a quantitative framework to analyze these issues. In future research I hope to extend my analysis to the asset side of the shadow banking system. This will allows me to understand the potential interaction between monetary policy and risk-taking behaviors of shadow banks.

2.7 Tables and Figures

Commercial Banks
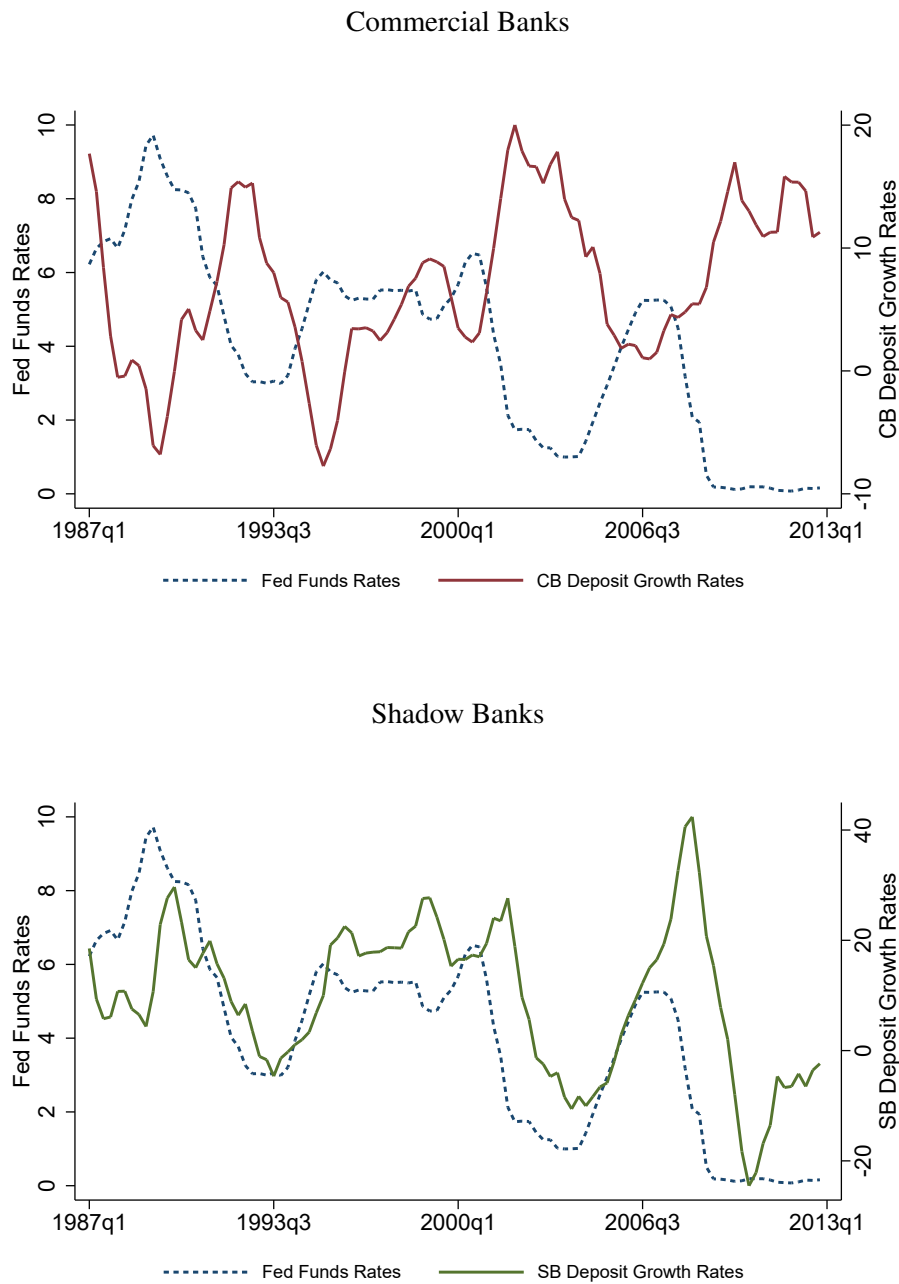
Shadow Banks

Figure 2.1: Deposit Growth Rates and the Fed Funds Rates

This figure shows the annual growth rates of the U.S. commercial and shadow bank deposits from 1987 to 2012. The data are quarterly. Commercial bank deposits are the sum of checking and saving deposits. Shadow bank deposits include all the U.S. retail and institutional MMF shares. The data are obtained from FRED.
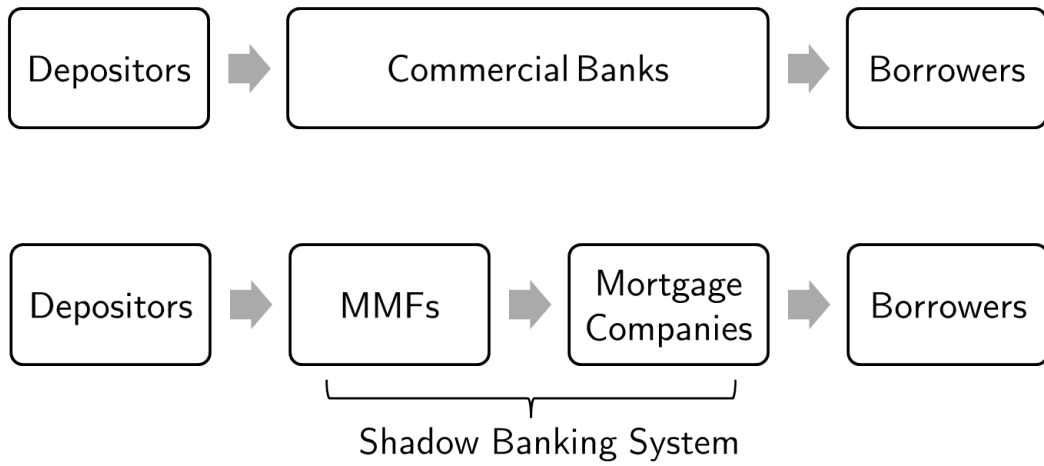
Figure 2.2: The U.S. Banking System
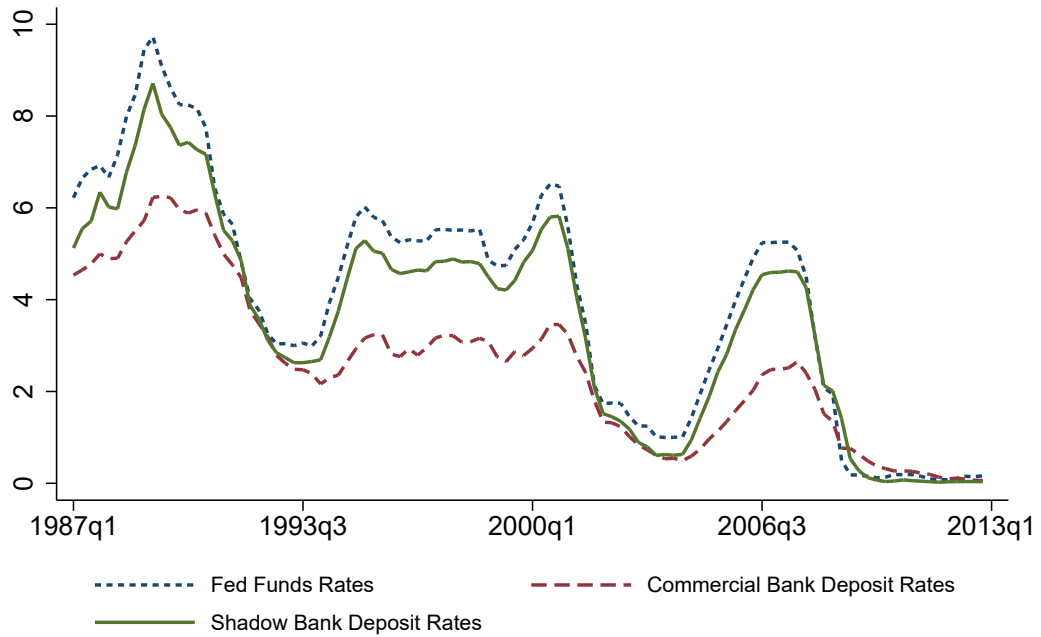
Figure 2.3: Deposit Rates and the Fed Funds Rates

This figure shows the average deposit rates of the U.S. commercial banks and MMFs from 1987 to 2012. The data are quarterly. Commercial bank deposit rates are obtained from the Call report. MMF yields are obtained from iMoneyNet.

Commercial Banks
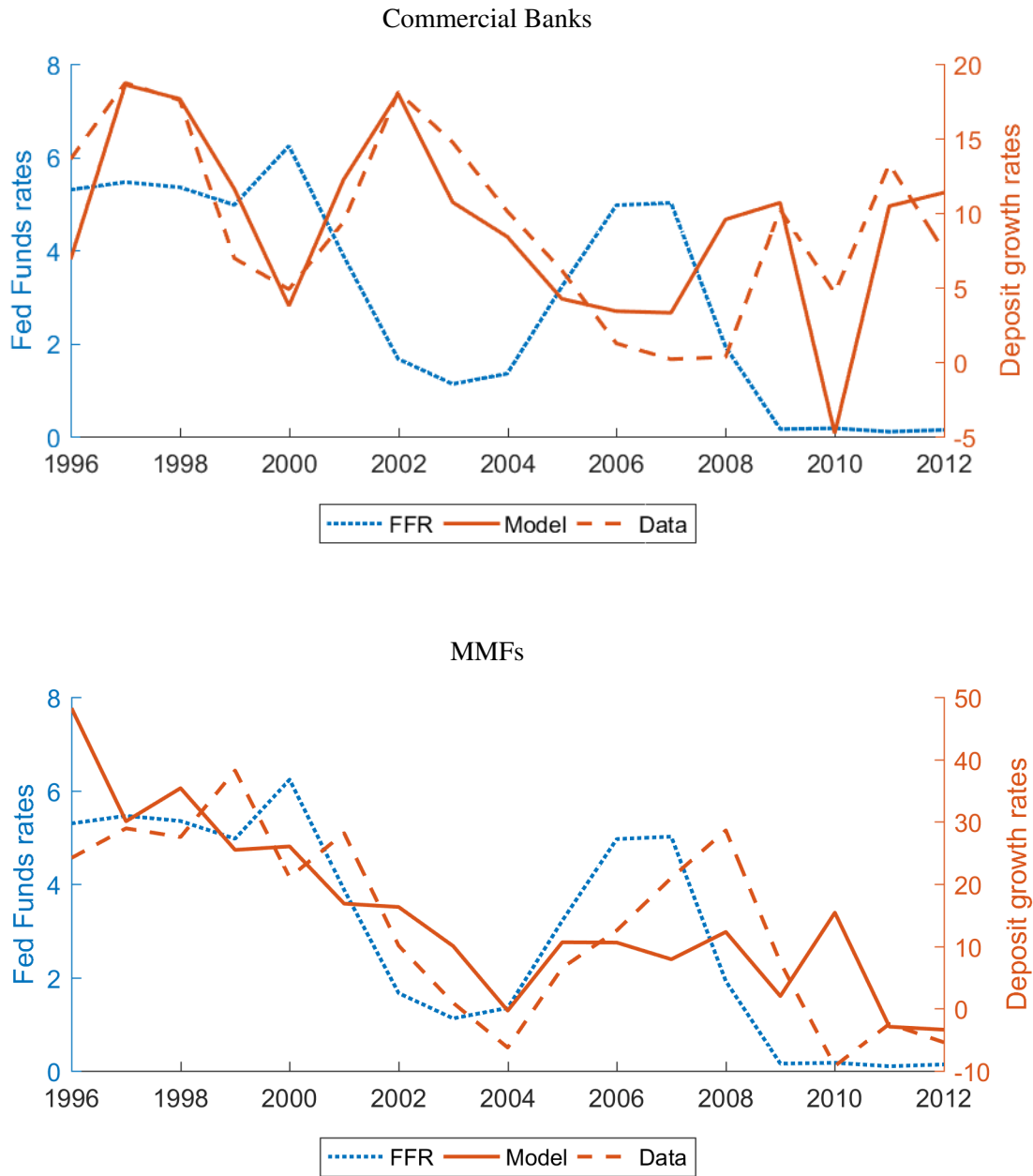
MMFs

Figure 2.4: Model Predicted Deposit Growth Rates

This figure shows the average deposit growth rates of commercial and MMFs predicted by the structural model and in the data. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.
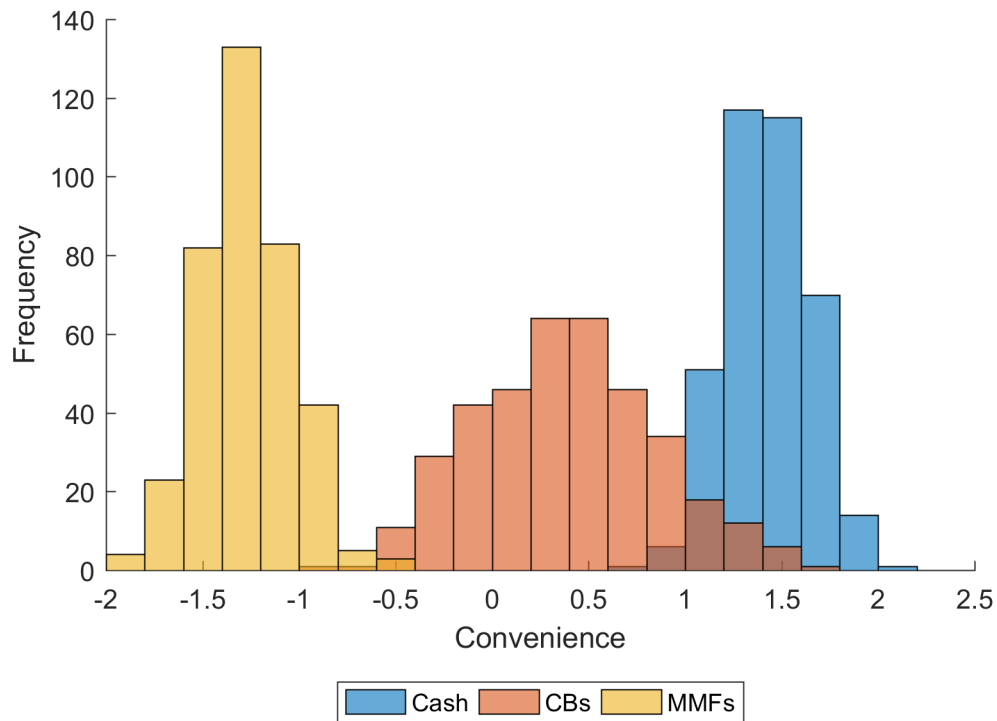
Figure 2.5: Distribution of Estimated Convenience
This figure shows the histogram of the estimated convenience for cash, commercial banks and MMFs. The convenience is defined as the inner product between the vector of characteristics, x, and corresponding sensitivities, β. Each observation is a MSA-sector median. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.
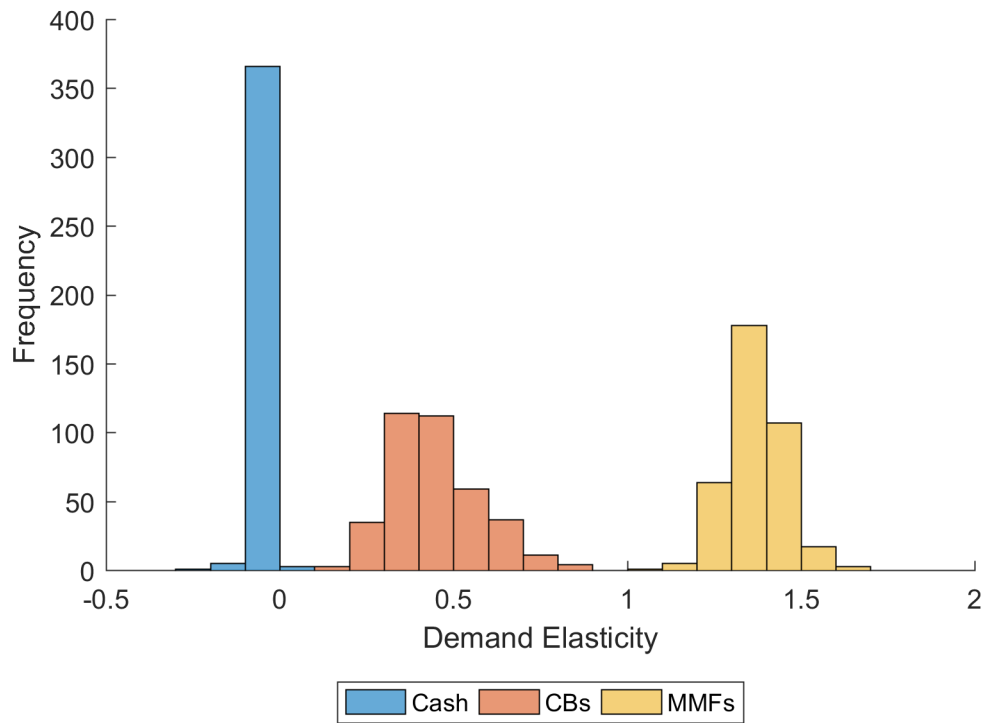
Figure 2.6: Distribution of Estimated Demand Elasticity
This figure shows the histogram of the estimated demand elasticity for cash, commercial banks and MMFs. Each observation is a MSA-sector median. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.
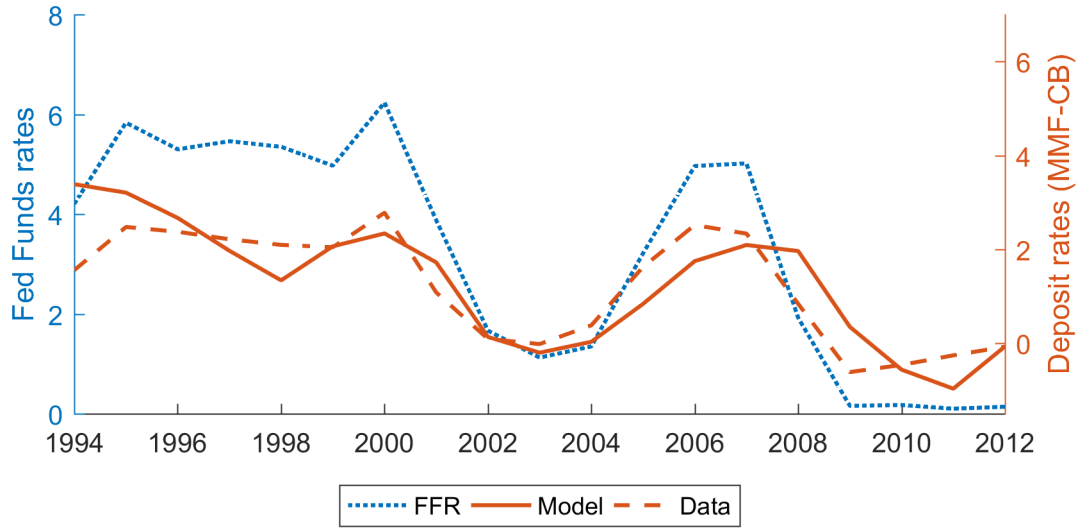
Figure 2.7: Difference in Deposit Rates (MMF-CB)

This figure shows the difference in average deposit rates between MMFs and commercial banks predicted by the structural model and in the data. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.

Heterogeneous Depositors ($\sigma > 0$)

Homogeneous Depositors ($\sigma = 0$)
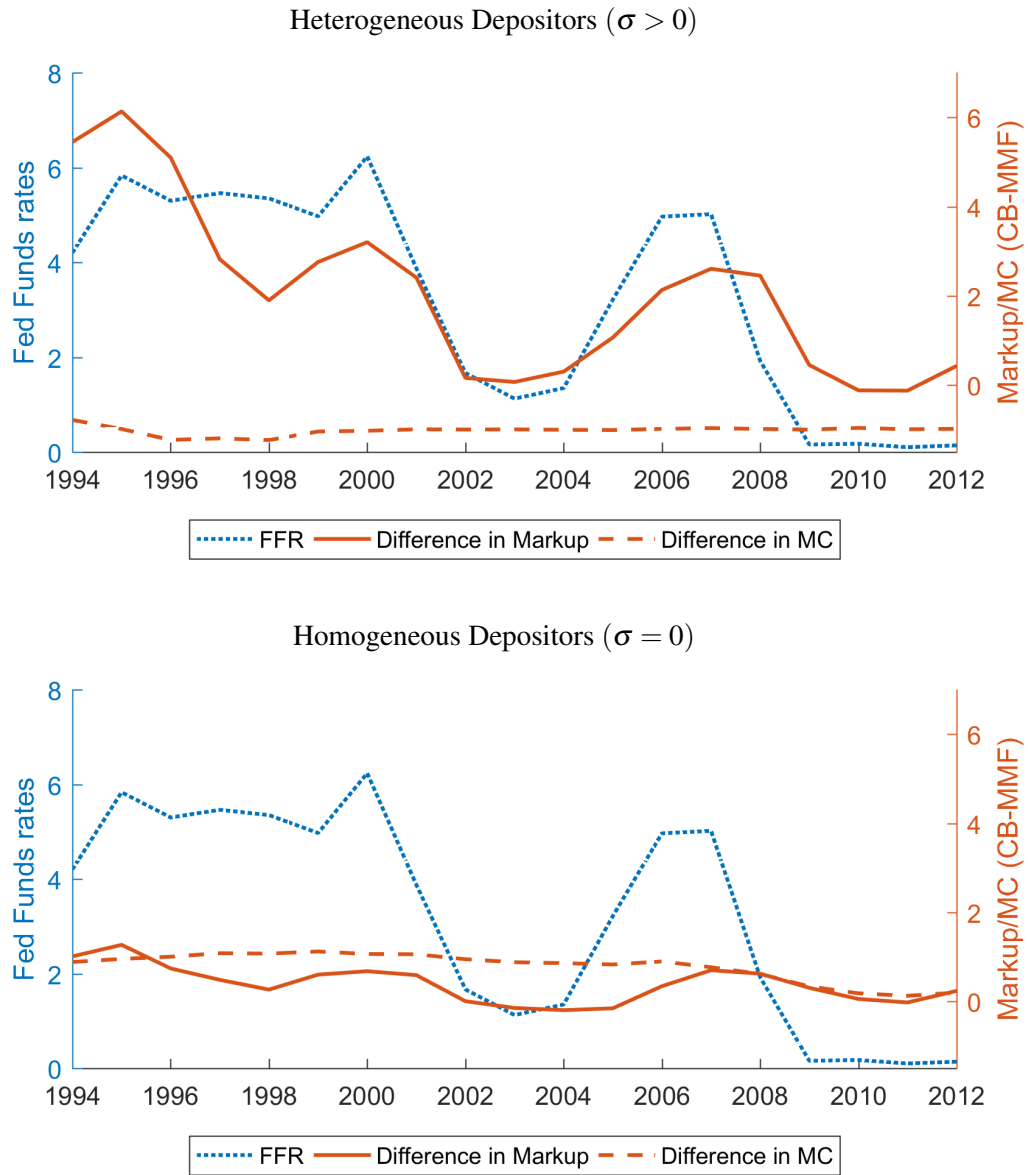
Figure 2.8: Difference in Markup and Marginal Cost (CB-MMF)

This figure shows the difference in average markups and marginal costs between commercial and MMFs predicted by the structural model and in the data. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.

Yield-oriented Depositors

Transaction-oriented Depositors
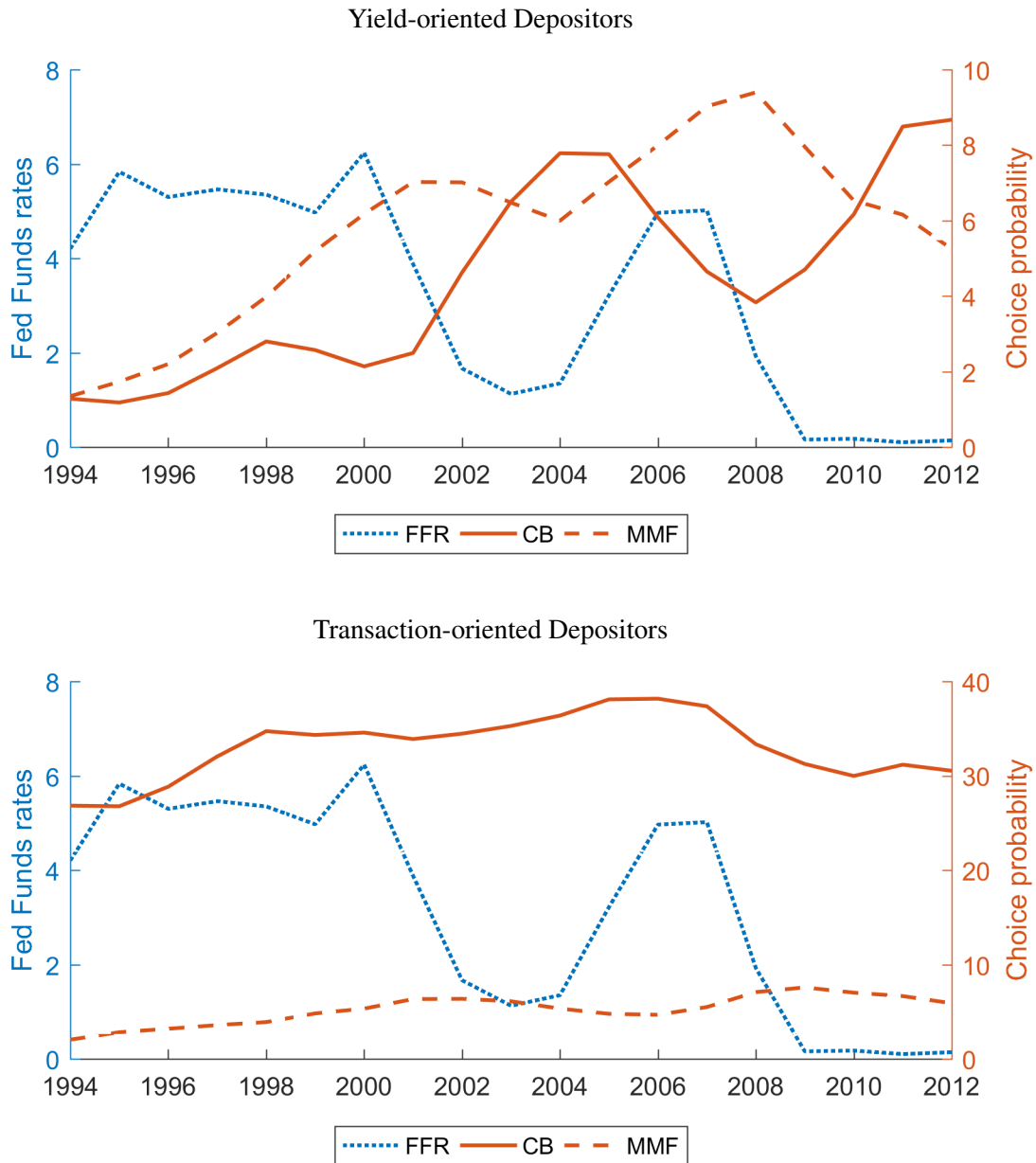
Figure 2.9: Choice Probability of Depositors by Type

This figure shows the estimated probability for yield-oriented and transaction-oriented depositors to choose commercial banks or MMFs over time. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.
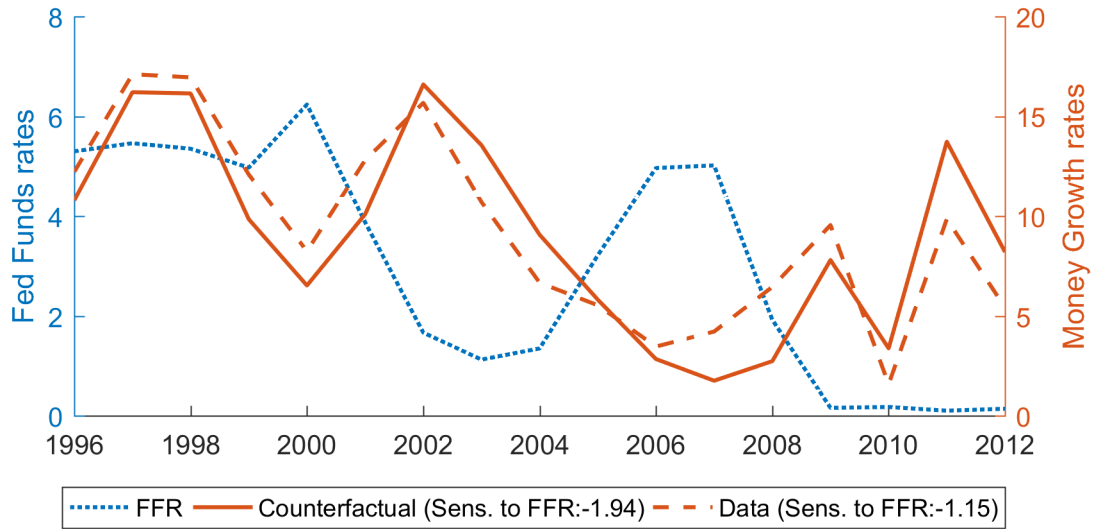
Figure 2.10: Counterfactual Aggregate Money Growth Rates

This figure shows the observed and counterfactual aggregate money growth rates. The counterfactual simulation is conducted by assuming that there are no MMFs in the economy. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.
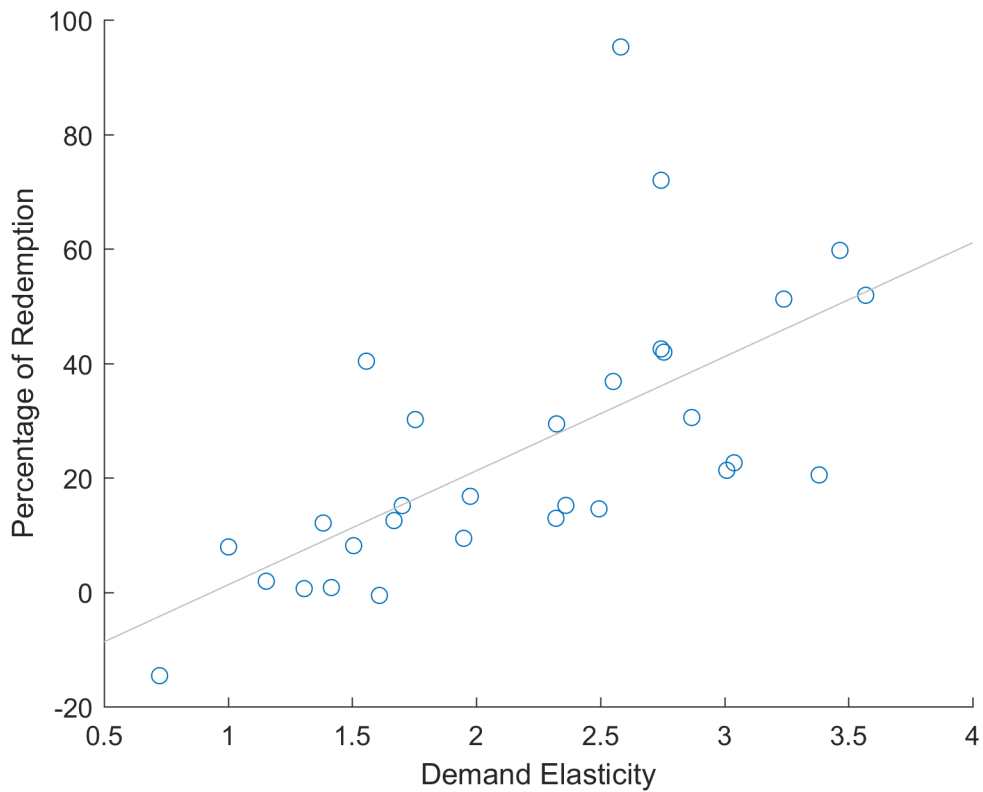
Figure 2.11: Demand Elasticity and the 2008 Runs on MMFs

This figure shows the scatter plot of demand elasticity against the percentage of redemption during the 2008 runs on MMFs. The demand elasticity is estimated using data before the default of Lehman Brothers on September 15, 2008. The percentage of redemption is calculated over four weeks following the start of the run (September 9th, 2008 - October 7th, 2008). The sample includes 30 largest prime MMFs in the U.S. as of September 9th, 2008.
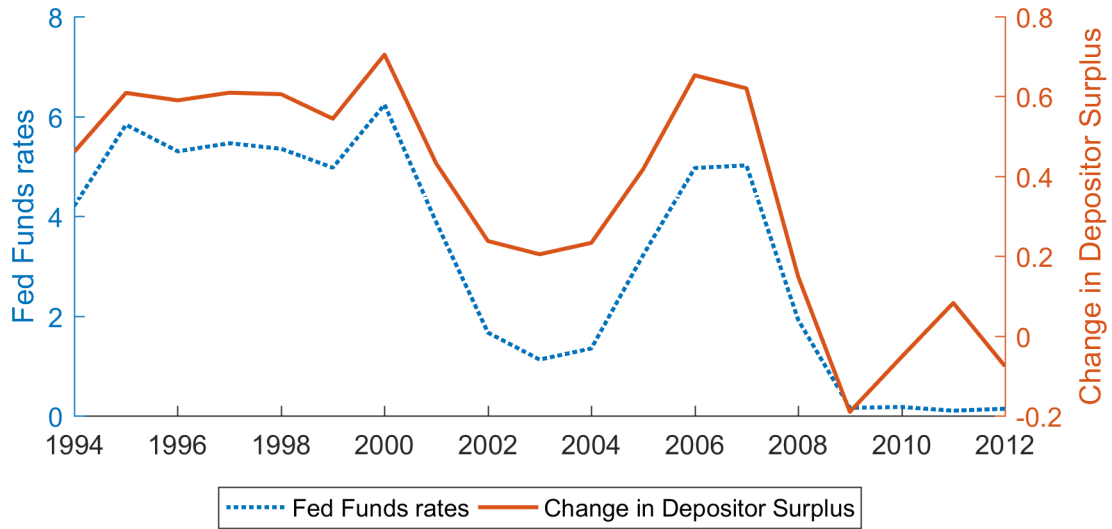
Figure 2.12: Change in Depositor Surplus

This figure shows change in depositor surplus due to the competition from MMFs. The change in surplus is calculated by comparing the actual economy with a counterfactual economy where there are no MMFs. The model is estimated using institution-level data on U.S. commercial banks and MMFs from 1994 to 2012.

Table 2.1: Summary Statistics

variable	mean	sd	p10	p25	p50	p75	p90
Cash							
Amount	1485.360	4284.194	150.692	224.929	402.519	1004.472	3007.820
Market share	11.345	1.474	9.653	10.432	11.270	12.228	13.180
Commercial Banks							
Amount	420.480	1561.996	49.098	70.789	115.569	244.662	735.627
Market share	3.512	1.707	1.752	2.370	3.200	4.294	5.533
Deposit rates	1.787	1.096	0.287	0.787	1.867	2.744	3.223
HHI	0.214	0.104	0.124	0.151	0.190	0.244	0.330
No. of banks	12.579	5.616	6.000	9.000	12.000	15.000	21.000
Branch density	22.589	11.324	10.273	14.546	20.699	28.602	37.074
No. of employee per branch	18.447	4.812	13.536	15.175	17.449	20.452	25.063
Single-market	0.181	0.165	0.000	0.000	0.154	0.286	0.417
Age	90.868	20.659	65.636	75.800	89.900	104.313	118.727
Expense of fixed assets	0.111	0.022	0.089	0.096	0.107	0.121	0.138
Salary	0.515	0.070	0.434	0.466	0.509	0.554	0.604
Reserve	1.100	1.500	0.149	0.260	0.477	1.051	3.365
MMFs							
Amount	65.919	224.330	4.930	8.721	15.693	40.365	127.887
Market share	0.445	0.110	0.288	0.385	0.451	0.517	0.586
Deposit rates	3.042	2.107	0.047	0.798	3.804	5.046	5.503
HHI	0.072	0.026	0.046	0.053	0.066	0.083	0.109
No. of banks	24.272	7.820	15.000	18.000	23.000	29.000	35.000
Rating dummy	0.516	0.068	0.438	0.467	0.520	0.565	0.600
Bank fund dummy	0.453	0.076	0.357	0.400	0.455	0.519	0.552
Check-writing dummy	0.343	0.063	0.261	0.296	0.333	0.391	0.438
Age	27.628	3.166	23.629	24.500	27.348	30.333	31.750
Management cost	0.182	0.052	0.123	0.166	0.191	0.217	0.228
Other cost	0.141	0.045	0.114	0.121	0.137	0.158	0.208

Table 2.2: Effect of Monetary Policy on Aggregate Deposit Growth Rates

	Dependent variable: Deposit growth rates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CB	MMF	Cash	Total	CB	MMF	Cash	Total
Fed Funds rates	-1.896***	3.890***	-0.0204	-0.0594	-1.407***	3.929***	0.340	0.177
	[0.307]	[0.766]	[0.181]	[0.326]	[0.263]	[0.875]	[0.256]	[0.369]
GDP growth					0.253	-1.289*	-0.156	-0.171
					[0.265]	[0.691]	[0.285]	[0.333]
Inflation rates					0.609	-1.124	-0.337	0.0252
					[0.412]	[1.520]	[0.419]	[0.548]
VIX					0.412***	0.161	0.106	0.282**
					[0.101]	[0.263]	[0.0866]	[0.128]
TED					-8.175***	14.47**	-2.914*	0.231
					[1.735]	[6.503]	[1.607]	[2.725]
N	92	92	92	92	92	92	92	92
adj. R-sq	0.517	0.405	-0.011	-0.010	0.662	0.636	0.072	0.194

This table presents time series regressions of aggregate deposit growth rates on the Fed Funds rates. The data frequency is quarterly. The sample period is from 1990 to 2012. Standard errors in brackets are computed with Newey-West standard errors with 4 lags. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Table 2.3: Demographic Determinants of Shadow Bank Deposit Holding

Dependent variable:	Shadow bank deposit holding		
	(1)	(2)	(3)
	Shadow dummy	Shadow share	Shadow amount
Income	0.036***	0.020***	0.818***
	[0.002]	[0.001]	[0.030]
College	0.040***	0.014***	1.392***
	[0.004]	[0.003]	[0.096]
Unemployed	0.005	0.002	-0.210
	[0.008]	[0.005]	[0.219]
Age	-0.011***	-0.006***	0.578***
	[0.002]	[0.001]	[0.045]
Age2	0.001	0.004***	0.120***
	[0.002]	[0.001]	[0.039]
Home owner	-0.010**	-0.007**	0.394***
	[0.004]	[0.003]	[0.097]
Car owner	-0.010**	-0.007***	0.233**
	[0.004]	[0.003]	[0.107]
Female	0.023***	0.018***	-0.458***
	[0.005]	[0.004]	[0.156]
Married	0.018***	0.009***	-0.048
	[0.005]	[0.003]	[0.131]
N	27764	27764	2860
adj. R-sq	0.047	0.025	0.593

This table presents cross-sectional regressions of shadow bank deposit holding on demographical variables. The sample includes 27,764 households in the Survey of Consumer Finance (2013). Shadow bank deposits are defined as deposits which are not insured by the government. Shadow dummy equals 1 if a household has shadow bank deposits, 0 otherwise. Shadow share is the share of shadow bank deposits in the total deposits of a household. Shadow amount (log) is the log dollar amount of shadow bank deposits held by a household. The independent variables are the demographics of the head of the household. Robust standard errors are in brackets. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Table 2.4: Demand Parameter Estimation

Dependent variable:	Mean utility	
	(1) Full	(2) Logit
Yield sensitivity	0.764***	-0.323***
	[0.019]	[0.0039]
Branch density	0.026***	0.045***
	[0.00053]	[0.00018]
No. of employees	0.007***	0.012***
	[0.0013]	[0.00041]
Single-market dummy	0.055***	0.011*
	[0.011]	[0.0056]
Age (CB)	-0.076***	-0.081***
	[0.011]	[0.005]
Age (MMF)	0.586***	0.352***
	[0.037]	[0.019]
Rating	-0.020***	0.009**
	[0.007]	[0.0037]
Bank fund	0.000	-0.022***
	[0.0071]	[0.0037]
Check writing	-0.014*	-0.015***
	[0.0076]	[0.0042]
Yield sensitivity dispersion	1.053***	
	[0.025]	
Adjustment cost	6.681***	
	[0.035]	
Bank F.E.	Y	Y
City F.E.	Y	Y
Time F.E.	Y	Y
N	269687	269687
adj. R-sq	0.496	0.759

This table presents the estimates of demand parameters of the structural model. The first column is the full model, and the second is the logit model in which there is no depositor heterogeneity and adjustment cost. The sample is a panel of U.S. commercial banks and MMFs from 1994 to 2012. Robust standard errors are in brackets. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Table 2.5: Supply Parameter Estimation

Dependent variable:	Marginal costs	
	(1)	(2)
	Full	Logit
Management costs	1.069***	-0.085***
	[0.037]	[0.026]
Other cost	1.151***	0.078***
	[0.027]	[0.019]
Branch density	-0.011***	0.003***
	[0.000]	[0.000]
No. of employees	-0.010***	0.008***
	[0.001]	[0.000]
Expense of fixed assets	-1.217***	6.819***
	[0.154]	[0.106]
Salary	0.475***	0.584***
	[0.040]	[0.027]
Reserve cost	1.765***	11.449***
	[0.135]	[0.093]
FFR*Cash dummy	1.000***	1.000***
	[0.006]	[0.004]
	269687	269687
adj. R-sq	0.504	0.610

This table presents the estimates of supply parameters of the structural model. The first column is the full model, and the second is the logit model in which there is no depositor heterogeneity and adjustment cost. The sample is a panel of U.S. commercial banks and MMFs from 1994 to 2012. Robust standard errors are in brackets. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Table 2.6: Own-rate Elasticity

	Cash	CB	MMF
Cash	0.1486		
	[0.571]		
CB		1.0668	
		[1.0208]	
MMF			1.8112
			[0.4647]

This table presents the median and standard deviation (in brackets) of own-rates elasticity of cash, commercial banks, and MMFs estimated from the full model. Each entry gives the percent change of the market share of product i with one percent change of its own deposit rates.

Table 2.7: Cross-rate Elasticity

	Cash	CB	MMF
Cash		0.0018	-0.0002
		[0.0204]	[0.0007]
CB	0.0073	-0.0087	-0.0021
	[0.1264]	[0.0247]	[0.0032]
MMF	-0.0062	-0.0119	-0.0048
	[0.0456]	[0.0156]	[0.0052]

This table presents the median and standard deviation (in brackets) of cross-rates elasticity of cash, commercial banks, and MMFs estimated from the full model. The entry of the i-th row and j-th column shows the percent change of the market share of a product in category i (cash, CB, MMF) with one percent change of the deposit rates of a different product in category j (cash, CB, MMF).

Table 2.8: Monetary Policy and MMF Lending

	Dependent variable: Change in lending/total lending					
	(1) CPs	(2) ABCPs	(3) Repos	(4) FRNS	(5) Treasury	(6) CB
Fed Funds rates	0.670***	0.0848***	0.457***	0.270***	0.171***	0.217***
	[0.0176]	[0.00468]	[0.0184]	[0.0128]	[0.0243]	[0.00887]
GDP growth	-0.102***	0.00534	-0.144***	0.111***	-0.534***	-0.0968***
	[0.0246]	[0.00648]	[0.0256]	[0.0178]	[0.0361]	[0.0122]
Inflation rates	0.139***	-0.000660	0.355***	-0.0913***	-0.0337	0.195***
	[0.0411]	[0.0113]	[0.0448]	[0.0307]	[0.0606]	[0.0209]
VIX	0.0315***	-0.00506***	-0.0117*	0.0186***	0.126***	0.00990***
	[0.00670]	[0.00180]	[0.00706]	[0.00487]	[0.00949]	[0.00337]
TED	-0.268**	-0.145***	-0.0882	-0.0463	1.568***	-0.0861
	[0.126]	[0.0332]	[0.132]	[0.0915]	[0.180]	[0.0619]
Fund Characteristics	Y	Y	Y	Y	Y	Y
Fund F.E.	Y	Y	Y	Y	Y	Y
N	41006	41006	41006	41006	41006	41006
adj. R-sq	0.102	0.093	0.098	0.069	0.094	0.076

This table presents regressions of MMF Lending on Fed Funds rates. The dependent variable is the annual change in a specific type of lending normalized by the lagged total lending (lagged one year). Fund characteristics include fund size (log), fund age, management costs, and other costs. The sample includes 1,148 MMFs in the period of 1998 to 2012. The data frequency is quarterly. Standard errors in brackets are clustered by time. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Table 2.9: Monetary Policy and Asset Growth of Shadow Banks

| | Dependent variable: Asset growth rates | | | | | |
	(1) Funding Corporations	(2) Finance Companies	(3) ABCP Issuers	(4) Captive Financial Institutions	(5) Broker-dealers	(6) Shadow Bank Total
Fed Funds rates	2.765***	1.439**	4.528***	0.972*	0.747	1.773***
	[0.617]	[0.649]	[1.184]	[0.500]	[0.839]	[0.555]
GDP growth	3.068***	1.813***	0.849	0.841	1.791	1.645***
	[0.693]	[0.574]	[0.851]	[0.638]	[1.089]	[0.369]
Inflation rates	-2.853***	0.647	-0.138	-4.271***	1.667	-1.001*
	[0.951]	[0.720]	[1.078]	[0.784]	[1.608]	[0.583]
VIX	0.206	0.418***	-0.203	-0.0540	-0.528*	-0.137
	[0.147]	[0.151]	[0.336]	[0.109]	[0.267]	[0.144]
TED	16.98***	-5.206*	-4.021	11.61***	-5.235	2.261
	[2.861]	[2.625]	[6.480]	[2.036]	[4.939]	[2.984]
N	92	92	92	92	92	92
adj. R-sq	0.641	0.386	0.495	0.484	0.449	0.561

This table presents time series regressions of the aggregate asset growth rates of shadow banks on the Fed Funds rates. The dependent variable is the annual growth rates of the shadow bank assets. The data frequency is quarterly. The sample period is from 1990 to 2012. Standard errors in brackets are computed with Newey-West standard error with 4 lags. ***, **, * represent 1%, 5%, and 10% significance, respectively.

Chapter 3

Regulation and Market Liquidity

3.1 Introduction

The aftermath of the 2008-09 financial crisis has witnessed one of the most active periods of regulatory intervention in U.S. financial history since the New Deal (Barr, 2012). A centerpiece of this sweeping reaction to the near collapse of the financial system, the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank), was signed into law in July 2010. With Dodd-Frank, hundreds of regulatory rulemaking requirements have been subsequently met, affecting virtually every dimension of modern financial activity, from derivatives trading to housing finance to capital requirements for depository institutions. In the backdrop of this intervention, a lack of rigorous assessment of the complex costs and benefits of the new rules has been highlighted (Cochrane, 2014). While Law scholars have been active in the regulatory debate at the qualitative level, quantitative work in Economics and Finance has been occasional and surprisingly sparse.

Pertinently to this debate, this paper investigates the crucial claim that U.S. post-crisis financial regulatory over-reach might have adversely affected the provision of market liquidity of a vast class of financial assets, structurally decreasing liquidity levels and increasing liquidity risk in fixed-income markets across the board.

Such claim is linked, but not uniquely, to a specific set of provisions embedded within recent legislation, the so-called Volcker Rule, statutorily delineated in Section 619 Title VI of the 2010 Dodd-Frank Act and finalized by multiple regulatory agencies in January 2014. According to this provision, any banking entity is prohibited from engaging in proprietary trading or from acquiring or retaining an ownership interest in, sponsoring or having certain relationships with a hedge fund or private equity fund, subject to certain exemptions. Although this is in no way the only dimension of Dodd-Frank along which serious welfare losses or liquidity shortages could have been potentially triggered, it emerged as one of the most hotly debated, with roughly $17,000$ public comments filed during the process of federal regulatory rulemaking (Bertrand, Bombardini and Trebbi, 2015). Specifically, some commentators[24] have highlighted how by placing undue artificial limits on securities inventory and retained risk and directly affecting inter-dealer

---

[24]For instance regulators write in the final version of the Volcker Rule (p.5578 Federal Register / Vol. 79, No. 21 / Friday, January 31, 2014 / Rules and Regulations) "*As discussed above, several commenters stated that the proposed rule would impact a banking entity's ability to engage in market making related activity. Many of these commenters represented that, as a result, the proposed exemption would likely result in reduced liquidity[...]*" and the Federal Register explicitly mentions on the matter of reduced liquidity comments received from "*AllianceBernstein; Rep. Bachus et al. (Dec. 2011); EMTA; NASP; Wellington; Japanese Bankers Ass'n.; Sen. Hagan; Prof. Duffie; Investure; Standish Mellon; IR&M; MetLife; Lord Abbett; Commissioner Barnier; Quebec; IIF; Sumitomo Trust; Liberty Global; NYSE Euronext; CIEBA; EFAMA; SIFMA et al. (Prop. Trading) (Feb. 2012); Credit Suisse (Seidel); JPMC; Morgan Stanley; Barclays; Goldman (Prop. Trading); BoA; Citigroup (Feb. 2012); STANY; ICE; BlackRock; SIFMA (Asset Mgmt.) (Feb. 2012); BDA (Feb. 2012); Putnam; Fixed Income Forum/Credit Roundtable; Western Asset Mgmt.; ACLI (Feb. 2012); IAA; CME Group; Wells Fargo (Prop. Trading); Abbott Labs et al. (Feb.14, 2012); Abbott Labs et al. (Feb. 21, 2012); T. Rowe Price; Australian Bankers Ass'n. (Feb. 2012); FEI; AFMA; Sen. Carper et al.; PUC Texas; ERCOT; IHS; Columbia Mgmt.; SSgA (Feb. 2012); PNC et al.; Eaton Vance; Fidelity; ICI (Feb. 2012); British Bankers' Ass'n.; Comm. on Capital Markets Regulation; Union Asset; Sen. Casey; Oliver Wyman (Dec. 2011); Oliver Wyman (Feb. 2012) (providing estimated impacts on asset valuation, borrowing costs, and transaction costs in the corporate bond market based on hypothetical liquidity reduction scenarios); Thakor Study. The Agencies respond to comments regarding the potential market impact of the rule in Part IV.A.3.b.3., infra.*"

Available at http://www.gpo.gov/fdsys/pkg/FR-2014-01-31/pdf/2013-31511.pdf

trading, the Volcker Rule could have severely limited market liquidity[25]. When recently the Congressional debate shifted on the merits of regulatory relief, one of the provisions considered for rolling back within Dodd-Frank included the prohibition of proprietary trading on the part of insured banking entities and their affiliates below certain thresholds[26].

A balanced view of the potential adverse welfare consequences of such provision is summarized in Duffie (2012): "*The Agencies' proposed implementation of the Volcker Rule would reduce the quality and capacity of market making services that banks provide to U.S. investors. Investors and issues of securities would find it more costly to borrow, raise capital, invest, hedge risks, and obtain liquidity for their existing positions. Eventually, non-bank providers of market-marking services would fill some or all of the lost market making capacity, but with an unpredictable and potentially adverse impact on the safety and soundness of the financial system. These near-term and long-run impacts should be considered carefully in the Agencies' cost-benefit analysis of their final proposed rule. Regulatory capital and liquidity requirements for market making are a more cost effective method of treating the associated systemic risks.*" Duffie (2012) further remarks on the needs for an appropriate assessment of the cost and benefits of the rule, an assessment that the empirical analysis we perform systematically complements. Thakor (2012) raises similar issues.

Another focal point of post-crisis regulatory reform has been the Basel III framework, which was produced in 2010 by the Basel Committee on Banking Supervision at the Bank for International Settlements. The Basel III final rule adopted by the U.S. federal banking regulators also implements some provisions from the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 (Dodd-Frank Act; P.L. 111-203), which also addressed capital reserve requirements for banks (Getter, 2014). Basel III demands higher capital and liquidity buffers for banks, and imposes leverage restrictions on systemically important financial institutions. Despite the fact that higher levels of bank capital may reduce the probability of another financial crisis, critics claim that these regulations might have unduly constrained banks' ability to deploy capital to market-making, and forced banks to charge clients more to use their balance sheet when they facilitate trades or provide financing[27].

This paper formally assesses the effect of the U.S. post-crisis regulatory intervention, encompassing the Dodd-Frank Act and Basel III, on market liquidity of a large portion of the U.S. fixed-income market.

Our biggest empirical challenge is the unknown timing of regulatory impact. As vividly shown in Figure 3.1, post-crisis financial regulation is characterized by protracted rulemaking processes and complicated anticipatory responses and lagging reactions of market participants. For example, the Volcker Rule took almost four years to finalize, with the deadline being postponed several times. During the four years of rulemaking, different banks wound down their proprietary trading desks at different times[28]. Conventional micro-econometric methods which compare liquidity before and after a treatment date are difficult to apply in this setting because it is unclear when regulation should have effects. The result of these methods could be sensitive to the assumption of the date around which the comparison is conducted[29].

---

[25]For example, on May 20, 2015 The Wall Street Journal in an article titled "Why Liquidity-Starved Markets Fear the Worst" reports "[..] *a large part of the explanation lies in changes to regulation aimed at addressing weaknesses exposed by the financial crisis. Banks must now hold vastly more capital, particularly against their trading books. The ring-fencing of proprietary trading in the U.S. and retail banking in the U.K. has also squeezed liquidity.* " Similar reasoning is implied by Alan Greenspan on the Financial Times on August 17, 2015, who writes "*Lawmakers and regulators, given elevated capital buffers, need to be far less concerned about the quality of the banks' loan and securities portfolios since any losses would be absorbed by shareholders, not taxpayers. This would enable the Dodd-Frank Act on financial regulation of 2010 to be shelved, ending its potential to distort the markets — a potential seen in the recent decline in market liquidity and flexibility.*"

[26]See S.1484 - Financial Regulatory Improvement Act of 2015, Title I: Regulatory Relief and Protection of Consumer Access To Credit. The bill is sponsored by Senate - Banking, Housing, and Urban Affairs Chairman Richard Shelby (R-AL).

[27]See "Global Macro Research: A Look at Liquidity", Goldman Sachs, August, 2015.

[28]The section "A Brief History of the Volcker Rule" in online appendix provides a detailed account of the rulemaking process of the Volcker Rule.

[29]For example, if liquidity deterioration occurred before the regulation is implemented, a test comparing the liquidity around the date of implementation may find no liquidity reduction.

To address this challenge, we employ recent econometric approaches based on large factor models (Stock and Watson, 2011; Chen, Dolado and Gonzalo, 2014) to identify structural breaks in both levels and dynamic latent factors for a large set of liquidity proxies in fixed-income markets. Our empirical approach is attractive on several dimensions. First, our tests do not require a priori knowledge of the exact timing of the breaks. Second, we can capture not only sudden breaks in levels, but also breaks in slow-moving trends. Finally, the tests display excellent power properties and appear robust to confounding factors in a battery of Monte Carlo simulations.

We explore the market for U.S. corporate bonds, a heterogeneous asset class directly affected by the Volcker Rule and Basel III capital regulation. Exploiting the segmented nature of corporate bond market, we construct a large panel of liquidity measures by bond issue size, credit rating, and lead underwriter's identity. Given that original underwriters typically tend to make markets on the specific securities underwritten, this allows us to potentially identify bank-specific liquidity breaks and more nuanced disaggregated dynamics. We also study U.S. Treasuries, an asset class which is exempted from the Volcker Rule, but is still affected by the stringent capital regulation of Basel III. Several commentators have ascribed recent episodes of trading disruption (e.g. the flash crash of October 15, 2014) to liquidity depletion.

Against the popular claim that post-crisis regulation systematically hurt liquidity, we find no evidence of liquidity deterioration during periods of regulatory interventions. While our methodologies do not allow to exactly quantify the causal effect of specific regulatory provisions on market liquidity, our portfolio of tests rebuts the hypothesis of a permanent reduction in liquidity. The empirical evidence robustly shows either no breaks or statistically significant breaks toward higher liquidity in the aftermath of Dodd-Frank and Basel III. Under the shared assumption that Dodd-Frank and Basel III represented in fact massively consequential policy interventions, large negative effects on liquidity have to be rejected. This is a substantial contribution in sharpening the debate on post-crisis financial regulatory intervention in the U.S. and Europe. We also present concordant evidence from microeconometric approaches based on difference-in-differences of matched bonds samples that support these findings. Our work both qualifies frequent informal discussion on the lack of evidence of large deterioration in market liquidity provision, a view shared by a growing group of market participants and policy makers[30], and is relevant to the rigorous assessment of the welfare consequences of the Dodd-Frank Act and Basel III in terms of hindering the market making capacity of large financial institutions, one of the main welfare costs observers have ascribed to the recent regulatory surge.

This paper employs four different estimation strategies. First, we employ standard multiple breakpoint testing (Bai and Perron, 1998, 2003) on the level of liquidity as a first-pass examination on the potential dates around which liquidity depletion may manifest. We find no evidence of liquidity depletion during the period of regulatory intervention (July 2010-December 2014), a period encompassing regulatory events such as the passage of the Dodd-Frank Act and Basel III, the proposal and finalization of the Volcker Rule, and related shutdowns of proprietary trading desk by different banks. On the contrary, statistically significant

---

[30]For example, the semi-annual Monetary Policy Report of the Federal Reserve in July 2015 writes: "*Despite these increased market discussions, a variety of metrics of liquidity in the nominal Treasury market do not indicate notable deteriorations*", and "*similar to the Treasury market, a range of conventional liquidity metrics in corporate bond markets also generally do not point to a significant deterioration of market liquidity in recent years*". See also Dudley (2015) and the New York Fed's Liberty Street Economics blog series, in particular "Has U.S. Corporate Bond Market Liquidity Deteriorated?" by Adrian et al., Liberty Street Economics, October 05, 2015.

This view is also echoed by some market participants. A Wall Street Journal commentary titled "Overlooking the Other Sources of Liquidity" writes that "*fortunately for investors, recent reforms and regulatory pressures have dramatically increased the number of participants who can make prices and provide liquidity across many fixed-income markets. Markets that have opened to competition now enjoy better pricing, efficiency and resiliency*". The global head of credit at Morgan Stanley, Steve Zamsky, said that "in our day-to-day, moment-to-moment business today, marketplace works just fine". The chief investment officer of Oppenheimer Funds, Krishna Memani, the president of Bianco Research, Jim Bianco, and the president of Better Markets, Dennis Kelleher, also voiced scepticism on the "overheated" worries on bond market liquidity.

breaks toward higher liquidity are often detected during this period.

Our second and third methodologies apply recent econometric approaches based on large factor models (Stock and Watson, 2011) to capture breaks in latent factor structures in the large panel of disaggregated liquidity measures. Specifically, our second methodology focuses on single breakpoint testing for large dynamic factor models (Chen, Dolado and Gonzalo, 2014), while our third methodology extends to more a realistic multiple breakpoint case, transposing the intuition of Chen, Dolado and Gonzalo (2014) to Bai and Perron (2003) type tests. These methodologies allow flexible forms of structural breaks (including breaks in trends, in serial correlation, or in factor loadings), and help us to answer the deeper question whether market liquidity would be higher or lower in absence of regulatory intervention. In simulations we show that our methodologies can successfully identify the onset of a gradual liquidity deterioration, even when masked by confounding factors, and accurately estimate the counterfactual path of liquidity using observed data.

We apply these methodologies to a large panel of disaggregate liquidity measures for corporate bonds. Our tests robustly capture breaks in latent liquidity dynamics at the start and at the end of the 2008-09 crisis (and indeed these tests can be employed to precisely time the beginning and end of the liquidity crisis). This reassures us on the tests having sufficient power within this specific empirical application. However, we find no systematic statistical evidence of structural breaks leading to lower liquidity during the period of regulatory intervention (July 2010-December 2014).

As opposed to time-series approaches delineated above, our fourth estimation strategy relies on a standard microeconometric approach in estimating liquidity deterioration around salient regulatory events, namely difference-in-differences matching (Heckman, Ichimura, Todd, 1997; Heckman, Ichimura, Smith, and Todd, 1998; Smith and Todd, 2005). In this part of the analysis we focus on the finalization of the Volcker Rule alone. We construct a dataset of bonds matched by issue size and credit rating, split between treatment and control based on whether the original underwriter is covered or not by Volcker Rule provisions. Matching allows for balancing between covered and non-covered bonds, assuaging concerns of attenuation due to heterogeneity across the two groups of securities.

Consistently across all four estimation strategies, this paper finds no systematic evidence of deterioration in liquidity levels or structural breaks in dynamic latent factors of the U.S. fixed-income market during periods of heightened regulatory interventions. This is in stark contrast to the popular claim that post-crisis would cause severe depletion in market liquidity. Instead, consistent with the view shared by an increasing group of policy makers and market participants, we find breaks toward higher liquidity during these periods, possibly due to entry of non-banking participants and increase in competition between market makers. We also document some changes in the market structure, notably the diminishing dealer inventory and the shift from principal-based trading towards agency-based trading. These evolutions in market structure started before the regulatory intervention, and do not appear to be associated with deterioration in commonly used liquidity measures. To the best of our knowledge, this is one of the very first studies to statistically assess liquidity depletion related to regulatory activity post-2008.

Our work is related to several strands of literature in both economics and finance. The first strand of literature studies the determinants and measurement of market liquidity. A recent comprehensive survey on this literature can be found in Vayanos and Wang (2012). Theoretical works such as Grossman and Stiglitz (1980), Kyle (1985), Roll (1984), Grossman and Miller (1988), Amihud and Mendelson (1986) , Gromb and Vayanos (2002), Duffie, Garleanu, and Pedersen (2005), and Brunnermeier and Pedersen (2009) relate illiquidity to underlying market imperfections such as participation costs, transaction costs, asymmetric information, imperfect competition, funding constraints, and search frictions. Many empirical works have since studied various measures of market liquidity across different asset classes, such as price impact (Amihud measure), price reversal (Roll measure), and bid-ask spreads. It has been shown that these liquidity measures are related to market frictions as suggested by theory, and can explain asset returns in both cross section and time series (see Amihud, Mendelson and Pedersen (2006) for a recent survey). Recent studies of fixed-income market liquidity can be found in Edwards, Harris, and Piwowar (2007), Bao, Pan

and Wang (2011), Feldhütter (2011), Dick-Nielsen, Feldhütter, and Lando (2012), Krishnamurthy (2002), and Hu, Pan and Wang (2012).

A second strand of connected literature studies statistical tests of structural changes[31]. These methodologies have been widely used in the macroeconomic literature to study structural changes in inflation-output relations, labor productivity, and monetary policy regimes[32]. Our paper contributes to this literature by employing a test of multiple breaks with unknown dates in dynamic factor models, transposing the intuition of Chen, Dolado, and Gonzalo (2014) to Bai and Perron (1998) type tests. We show that this type of tests is particularly useful when the timing of regulatory impact is unclear.

A third and important strand of literature pertains to the cost-benefit analysis of financial regulation. By every stretch of imagination, this literature remains considerably underdeveloped relative to potential welfare benefits of rigorous and data-driven regulatory intervention. Such limitations have been lamented not only by financial economists such as Cochrane (2014), but have been central motivation of judicial intervention[33]. Cochrane (2014) discusses at length the complexity of deriving meaningful assessments of regulatory counterfactuals in financial and banking regulation, question also discussed in Posner and Weyl (2013). Relative to the pessimistic assessment in Coates and John (2014) of the infeasibility of meaningful cost-benefit analysis in financial and banking regulation[34], our paper offers a more optimistic counterpoint, at least in terms of ex-post quantitative assessment[35] along the specific dimension of market liquidity depletion. Related our study, Bessembinder et al. (2016) find that trade execution costs of corporate bonds have decreased over time, a finding consistent with ours. However, they interpret the decline in inventory and the shift of dealers' business model as a sign of liquidity deterioration induced by post-crisis regulation, while we find that the shift started before regulatory intervention, and does not seem to be associated with deterioration in other commonly used liquidity measures. In other OTC markets, Loon and Zhang (2016) provide evidence that Dodd-Frank improves the liquidity in the CDS market through several reforms such as public dissemination of transactions and central counterparty (CCP) clearing.

A fourth literature touched by this paper revolves around the post-financial crisis policy responses. McCarthy, Poole, and Rosenthal (2013) debate political distortions in post-crisis responses, an issue also explored in Frieden (2015) and Mian, Sufi, and Trebbi (2014). More explicitly, Mian, Sufi and Trebbi (2010) focus on the legislative response to the financial crisis pre-dating the Dodd-Frank Act, while Kaiser (2013) offers an interesting and detailed discussion of the congressional evolution of the Dodd-Frank Act itself. Finally, the regulatory rulemaking of Dodd-Frank is fully explored from a systematic empirical perspective by Bertrand, Bombardini, and Trebbi (2015).

The remainder of this paper is organized as follows. In Section 3.2 we discuss the main empirical measures, the variables construction, and provide a descriptive analysis of our samples. In Section 3.3 we discuss our econometric model and single breakpoint/multiple breakpoint testing in dynamic factor models. Our main empirical results on U.S. corporate bonds are reported in Section 3.4 and on Treasuries in Section 3.5. Section 3.6 concludes.

---

[31]Important theoretical contributions include Andrews (1983), Andrews and Ploberger (1994), Bai and Perron (1998), Stock and Watson (2002, 2011), and Chen, Dolado, and Gonzalo (2014).

[32]See McConnell and Perez-Quiros (2000), Hansen (2001), and Stock and Watson (2011) for examples of applications.

[33]Coates and John (2014) referring to Business Roundtable et al. v. SEC, 647 F. 3d 1144 (D.C. Cir. 2011), report that "*One panel of the U.S. Court of Appeals for the District of Columbia Circuit, composed entirely of Republican-appointed judges, has held that existing law requires the SEC to quantify the costs and benefits of its proposed rules*".

[34]Specifically speaking about the Volcker Rule, Coates and John (2014, p.73): "*Could the agencies go beyond conceptual CBA and conduct a reliable, precise, quantified CBA/FR? The short answer is no. There is simply no historical data on which anyone could base a reliable estimate of the benefits of preventing banks from engaging in proprietary trading or investing in hedge and private equity funds.*"

[35]See also Cochrane (2014)'s discussion of retrospective analysis of financial regulation.

3.2 Data

3.2.1 U.S. Corporate Bonds Sample Description

The first main data set used for this paper is the Financial Industry Regulatory Authority's (FINRA) TRACE. This data currently provides transaction-level information of approximately 99% of all secondary corporate bond market transactions. Our sample period is from April 1, 2005 to December 31, 2014, covering the 2008-09 financial crisis and post-crisis regulatory interventions. We filter out erroneous trades following Dick-Nielsen, Feldhütter, and Lando (2012).

We merge the cleaned TRACE transactions to bond characteristics provided by Mergent Fixed Income Data. This data provides bond-level information such as issue date, issuance size, coupon rate, maturity date, credit ratings, underwriter identity and roles. Following Dick-Nielsen, Feldhütter, and Lando (2012), we limit the sample to fixed-rate bonds that are not callable, convertible, putable, or have sinking fund provisions. We drop bonds issued more than 10 years ago, since these old bonds present very few transactions. Since our goal is to provide the most comprehensive coverage of U.S. corporate bond market, we keep bonds with semi-annual coupons because they are the most common bonds in the U.S.. The raw TRACE data contains $34,422$ bonds. After applying the above filters, our final sample contains $18,632$ semi-annual coupon bonds[36]. Using the underwriting information from Mergent, we link each bond to its lead underwriters.

We first construct the nine measures for each corporate bond in our sample. Then we calculate the equal weighted average by bond rating group (investment-grade v.s high-yield) and issue size (above \$1 billion v.s. below \$1 billion) for each underwriter, which we refer as disaggregate series[37]. Since smaller underwriters only underwrite a limited number of bonds, this makes the underwriter-level measure of liquidity quite noisy. Therefore, we keep the top 4 biggest underwriters, Bank of America (Merrill Lynch), JPMorgan Chase, Morgan Stanley and Goldman Sachs, and combine the rest into a residual "Others" group. We also construct aggregate liquidity measures for the whole corporate bond market.

3.2.2 Corporate Bonds Liquidity Measures: Construction

Market liquidity is the degree to which investors can execute a given trade size within a given period of time without moving the price against the trade. We use the following nine liquidity measures which are commonly used in the literature to capture different aspects of liquidity (the easiness to trade, the pecuniary cost of trading, etc.). Previous literature has shown that these liquidity measures generally affect asset prices, indicating that investors indeed care about them[38]. All measures below are decreasing in the level of liquidity[39].

1. Amihud measure. Amihud (2002) constructs an illiquidity measure based on the theoretical model of Kyle (1985). We use a slightly modified version of this measure following Dick-Nielsen, Feldhütter, and Lando (2012). The Amihud proxy measures the price impact of a trade per unit traded. For a given bond, define $r_{j,i,t}$ as the return and $Q_{j,i,t}$ as the trade size (in million \$) of the $j-$th trade on day i in month t. The daily Amihud measure is the average of the absolute returns divided by the corresponding trade size within

---

[36]This is different from Dick-Nielsen, Feldhütter, and Lando (2012), who keep the no-coupon bullet bonds. They cover $2,224$ bullet bonds and turn to focus on more liquid segment of the market.

[37]We also experimented with value-weighted averages with similar results to the ones reported below.

[38]Dick-Nielsen, Feldhütter, and Lando (2012) show that higher value of Amihud measure, Roll measure, IRC, Amihud variability, and IRC variability are associated with significantly higher credit spreads of corporate bonds. However, the evidence of turnover and zero-trading days is mixed.

[39]Some measures (e.g. Amihud) require a minimum number of trades to compute. We keep all the observations even if some liquidity measures are missing in certain days because we want to have a comprehensive coverage of the entire bond universe. To be sure, measures such as zero-trading days and turnover can be computed for all bonds.

day i:

$$Amihud_{i,t} = \frac{1}{N_{i,t}} \sum_{j=1}^{N_{i,t}} \frac{|r_{j,i,t}|}{Q_{j,i,t}} \tag{3.1}$$

where $N_{i,t} + 1$ is the number of trades recorded on day i. We exclude retail trades (i.e. trades below \$100,000 in volume), as they are unlikely to have price impact. At least two trades are required on a given day to calculate the measure, and we define a monthly Amihud measure by taking the median of the daily measures within month t.

2. Imputed round-trip cost (IRC). Feldhütter (2012) shows that if a bond that does not trade for days suddenly has two or three trades with the same volume within a short period of time (one day in our definition), then such trades are likely part of a pre-matched arrangement in which a dealer has matched a buyer and a seller. These trades are defined as a set of imputed round-trip trades. The difference between highest and lowest price in a set of imputed round-trip trades is the bid-ask spread collected by the dealer, which is a measure of liquidity of the bond. We follow this approach. Specifically, for a given bond, on each day i we identify sets of imputed round-trip trades indexed by k. A set of imputed round-trip trades involves two or more transactions with the same trading volume. Define $P_{k,i,t}^{\max}$ (resp. $P_{k,i,t}^{\min}$) as the maximum (resp. minimum) price among all the transactions in the k-th set of round-trip trades for that bond on day i in month t. The imputed round-trip cost of k-th set of round-trip trade is defined as:

$$IRC_{k,i,t} = \frac{P_{k,i,t}^{\max} - P_{k,i,t}^{\min}}{P_{k,i,t}^{\min}}. \tag{3.2}$$

We define a monthly IRC measure by taking the mean of the IRC of each set of imputed round-trip trades within month t, weighted by the number of transactions involved in each set of imputed round-trip trades.

3. Roll measure. The intuition of the Roll measure is as follows: the transaction price tends to bounce between the bid and ask price, which causes consecutive trade returns to be negatively correlated. Under certain assumptions as shown in Roll (1984), the Roll measure equals to the bid-ask spread. The Roll measure is defined as two times the square root of the negative covariance between two consecutive daily returns $r_{i,t}, r_{i-1,t}$ in month t. If the covariance is positive, the covariance is replaced with zero.

$$Roll_t = 2\sqrt{-Cov(r_{i,t}, r_{i-1,t})} \tag{3.3}$$

4. Non-block trades. A trade is defined as non-block trade if the trading volume is less than \$5 million for investment-grade bonds, and \$1 million for high-yield bonds. The frequency of non-block trades is defined as the ratio between the number of non-block trades and the total number of trades in month t.

5. Size (negative log). Lower liquidity is usually associated with smaller size of trade. We first take the negative logarithm of the par value for each trade, then compute the monthly median for each security.

6. Turnover (negative). The annualized turnover for month t is defined as the annualized trading volume devided by the amount outstanding. In what follows we take the negative of turnover as proxy of illiquidity, for consistency with the other measures.

7. Zero trading days. We define this measure as the ratio between days with zero trade and the number of trading days in month t.

8 . Variability of Amihud and 9. Variability of IRC. Investors not only care about the current level of liquidity, but also the risk of future liquidity. Therefore, we create the standard deviations of the daily Amihud measure and imputed round-trip costs in a month as measures of liquidity risk.

3.2.3 U.S. Treasuries Sample Description

We use the CRSP Treasury database to construct our liquidity measures for the U.S. Treasury market. The daily data file is used to construct the Roll measure, and the monthly data file is used to construct the

on-the-run premium.

We restrict our analysis to the same period as our corporate bond sample, April 1, 2005 to December 31, 2014. Our sample consists of Treasury bills, notes, and bonds that are noncallable, nonflowering, and with no special tax treatment. We also drop observations with obvious pricing errors such as negative prices. Treasury securities with remaining maturity less than 30 days are also dropped because of potential liquidity problems. After applying the filters, our final sample contains 1,124 bonds. In addition to bond prices, we obtain the total Treasury trading volume from Securities Industry and Financial Markets Association (SIFMA), and the total public debt outstanding from Bloomberg.

The liquidity measures for U.S. Treasuries are the following:

1. Yield curve fitting noise. Hu, Pan, and Wang (2013) proposes a market-wide liquidity measure by exploiting the connection between the amount of arbitrage capital in the market and observed "noise" in U.S. Treasury bonds—the shortage of arbitrage capital allows yields to deviate more freely from the curve, resulting in more noise in prices. They construct the noise measure by first fitting Treasury daily prices into a smooth yield curve, and then calculate the mean squared errors[40].

2. On-the-run premium. On-the-run Treasury bond (latest issue) usually enjoys a price premium over old bonds with similar maturity. We follow Gurkaynak et al. (2007) to construct the liquidity premium as the difference between the yield of this synthetic off-the-run bond and the on-the-run bond.

3. Roll measure and 4. Turnover (negative). Roll measure and Turnover (negative) measure are constructed similarly as in the case of corporate bonds.

3.2.4 Summary Statistics and Descriptives

Table 3.1 reports the summary statistics of the aggregate-level liquidity measures of the U.S. corporate bonds for the period April 2005 to December 2014. For a typical bond, there is no single trade on 74% of business days. The annualized turnover rate is only 29%[41]. In comparison, stocks in NYSE have a turnover ratio of 92% in December 2014[42]. Among all the trades, only 4% are block trades, and the median trade size is $35,000.

To get a quantitative assessment of the illiquidity, one can compare various trading cost measures to credit spreads, the compensation for investors to bear the credit and liquidity risk of corporate bonds. The average credit spread of a U.S. corporate bond over a Treasury bond is 2.20% over our sample period. In comparison, the mean Amihud measure, which is based on the impact of $1 million dollar trade, is 1.29%, as reported in Table 3.1. This amounts to half of the average credit spread earned in a year. The average IRC, which measures the cost charged by dealers in a round-trip trade, is 0.70%. This equals to a third of the average credit spread. The average Roll measure is 1.59%, which implies a bid-ask spread as large as three-fourth of the average credit spread.

Additionally, investors face high uncertainty in trading cost when executing their trades, as shown by a high time series variability of the Amihud and IRC measure. In synthesis, Table 3.1 shows that the U.S. corporate bond market is typically not particularly liquid. In this respect, the a priori concerns of public commentators of the effects of regulatory intervention on market liquidity were well placed.

In Table 3.2 we report the monthly linear correlations for each pair of liquidity proxies, to show consistency across our nine different measures of liquidity. Correlations are typically positive and sizeable, with partial exceptions of the Turnover (negative) measure[43].

---

[40]We obtain the measure from the authors' website at http://www.mit.edu/~junpan/Noise_Measure.xlsx

[41]The average of turnover across bonds is much lower than the aggregate turnover of the market (total trading volume divided by total bond outstanding). This is because most of the total trading volume comes from a small group of large size bonds.

[42]See http://www.nyxdata.com/nysedata/asp/factbook/ for the historical trading volume of NYSE stocks.

[43]In online appendix Table 1, we provide summary statistics of the 180 disaggregate series. In online appendix Figure 1, we plot time series of nine liquidity measures for each underwriter.

3.3 Econometric Model

Our goal is to formally test for structural breaks in the market liquidity of fixed-income assets in the aftermath of the financial crisis. If post-crisis financial regulation indeed generates adverse impacts on market liquidity, we should be able to detect structural breaks towards lower liquidity in the period of regulatory intervention (July 2010-December 2014). We present here the econometric setup that we are going to employ.

As anticipated in Section 3.2 we take both an aggregate-level and a disaggregate-level perspective in our analysis. Let us define the matrix Y of L aggregate liquidity measures observed for T periods. Y is of dimension $(T \times L)$. With the term "aggregate" liquidity measure we mean a measure of liquidity (such as those listed in Subsection (3.2.2)) that aggregates all securities in a market irrespective of identity of the underwriter, issue size, or credit rating. Although intuitive, this approach may mask heterogeneity in the dynamics of different types of securities. Therefore, to identify specific structural breaks that might arise only within particular classes of securities or only for bonds where markets are made by specific underwriters/banks, we will refer to disaggregate liquidity measures as the matrix X of $N > L$ liquidity measures observed for T periods. X is of dimension $(T \times N)$ where each column measures liquidity grouping bonds at the level of

$$(\text{identity of the underwriter} \times \text{issue size} \times \text{credit rating}) \tag{3.4}$$

As a matter of accounting, recall that for our case we have $L = 9$ measures. With 4 major underwriters plus 1 for the residual Others, 2 types of issue sizes (small or large), 2 types of credit rating (high yield and investment grade), we have $N = 180$. Our sample covers $T = 117$ months.

3.3.1 Multiple Breakpoint Tests for Liquidity Levels

Our first methodology studies the question of whether regulatory intervention has produced structural breaks in the *level* of liquidity, in either Y or X. We employ tests for multiple breakpoint estimation (Bai and Perron, 1998, 2003). The underlying assumption of these tests is that the level of liquidity fluctuates around a stable mean in absence of structural changes. If regulation shifts the long-run mean towards a different level, these tests will detect the dates when the changes occur. Although highly stylized, this analysis offers a first-pass examination of the potential dates around which liquidity depletion may have happened. More flexible models allowing for more general types of breaks will be presented below.

3.3.2 Single Breakpoint Testing for Dynamic Factor Models

Our second and third methodologies employ a more innovative approach based on dynamic factor models (Stock and Watson, 2011; Chen, Dolado and Gonzalo, 2014) to capture breaks in the latent factor structure. This approach allows flexible forms of structural breaks, such as breaks in trends, in serial correlation, or in factor loadings. These methodologies are more recent and deserve a more complete discussion. We now introduce the basic notation, econometric setup, and follow the exposition in Chen, Dolado and Gonzalo (2014), to which we refer for a detailed discussion of the proofs and the Monte Carlo evidence of power and size of the tests.

Consider a set of N observed liquidity measures constructed as in Section 3.2 and observed for $t = 1, ..., T$ periods, say, at monthly frequency. The matrix of observed disaggregate variables[44] X of dimension $(T \times N)$ is expressed as function of r unobserved factors F of dimension $(T \times r)$, a matrix Λ of factor loadings of dimension $(N \times r)$, and a matrix of idiosyncratic errors ε of dimension $(T \times N)$. As typical in the literature, we have in period t:

$$X_t = \Lambda F_t' + \varepsilon_t. \tag{3.5}$$

---

[44]For the dynamic factor model analysis let us indicate with an abuse of notation X as the matrix of first differenced and normalized liquidity measures, as indicated by Stock and Watson (2011).

This formulation accommodates flexibly several possible latent structures: r static factors; or $\tilde{r}$ dynamic factors and $p = r/\tilde{r} - 1$ lags; or an arbitrary combination of static and dynamic factors and lags (Stock and Watson, 2011).

Due to their flexibility in accommodating general dynamics across correlated time series, large factor models have enjoyed substantial success in the macroeconomics and finance literature. Stock and Watson (2002) show that the latent factors are consistently estimable by principal component analysis (PCA), an approach we follow here. PCA allows to estimate the r factors of X :

$$\hat{F}_t \equiv \left[\hat{F}_{1t}, \hat{F}_{2t}, ... \hat{F}_{rt}\right] \tag{3.6}$$

by focusing on the first r largest eigenvalues of the matrix XX' in the case $T \leq N$ (or of the matrix $X'X$ in the case $T > N$) and selecting the (appropriately orthogonalized and normalized) corresponding eigenvectors. Following Chen, Dolado and Gonzalo (2014) we also define $\hat{F}_{-1t} \equiv \left[\hat{F}_{2t}, ... \hat{F}_{rt}\right]$.

The number of factors r has to be estimated, as the true number of factors is unknown. Let us indicate with $\hat{r}$ such estimated value over the full sample.

To this goal we employ ten different estimators, some with better finite sample properties than others, with the aim of providing an exhaustive range of $\hat{r}$'s. Eight of the estimators we employ follow the popular information criteria (IC) proposed by Bai and Ng (2002), including their preferred IC_{p1}, IC_{p2}, PC_{p1}, and PC_{p2}. IC estimators, however, can occasionally display in finite samples a somewhat undesirable dependency on a specific parameter necessary to the estimation: the maximum number of admissible factors in the model (typically indicated as *kmax*). This may lead to overestimation of the true number of factors (Ahn and Horenstein, 2014). It is also the reason we additionally employ the recent *ER* (eigenvalue ratio) and *GR* (growth ratio) estimators of Ahn and Horenstein (2014), which do not share this drawback and, by focusing on the ratio of subsequent eigenvalues (or the ratio of their logs), also hinge on the straightforward intuition of principal component analysis screeplots (i.e. a popular graphical representation of the progressive explanatory power of each principal component ranked by size of its eigenvalue). We consider all number of factors between the minimum and the maximum of the estimated $\{IC_{p1}, IC_{p2}, IC_{p3}, PC_{p1}, PC_{p2}, PC_{p3}, AIC_3, BIC_3, ER, GR\}$, allowing for at least $\hat{r} = 2$ unobserved factors (a necessary condition for the statistical tests below).

We now proceed in introducing structural breaks in (3.5) and focus initially on the methodology for testing a single breakpoint, leaving multiple breakpoints to Section 3.3.3. It is relevant first to specify whether one is interested in breaks in the factor loadings Λ or in the factors F. Let us begin by representing a single structural break in all factor loadings at date τ:

$$X_t = \Lambda F_t' + \varepsilon_t \quad t = 1, ..., \tau \tag{3.7}$$

$$X_t = \Gamma F_t' + \varepsilon_t \quad t = \tau + 1, ..., T \tag{3.8}$$

where Γ is the post-break matrix of factor loadings of dimension $(N \times r)$. An important insight of Chen, Dolado and Gonzalo (2014) is that (3.7)-(3.8) can be represented as

$$X_t = \Lambda F_t' + \Delta G_t' + \varepsilon_t \tag{3.9}$$

where $\Delta = \Gamma - \Lambda$ measures the change in the loadings and

$$G_t = 0 \quad t = 1, ..., \tau \tag{3.10}$$

$$G_t = F_t \quad t = \tau + 1, ..., T.$$

The notation so far has focused on a single structural breakpoint for all r factors. At a given breakpoint, Chen, Dolado and Gonzalo (2014) distinguish between two types of breaks: small and large. Consider k_2

small breaks, of the type discussed by Stock and Watson (2002, 2009). These are defined as local-to-zero instabilities in the factor loadings that asymptotically average out without affecting estimation and inference under PCA. These are not the type of breaks we are interested in. In the context of large policy shifts, one is most likely interested in big structural breaks, indicated as $k_1 = r - k_2$. The formal definition is given in Chen, Dolado and Gonzalo (2014), but more importantly it is proven that under k_1 big breaks in (3.9), $\hat{F}_t$ estimated by PCA delivers inconsistent estimates of the space of the original factors F_t. Instead, defining G_t^1 the partition of G_t corresponding to the large breaks only, the full sample PCA delivers consistent estimates of the space of the new factors $\begin{bmatrix} F_t & G_t^1 \end{bmatrix}$. Specifically, over the full sample the number of factors tends to be overestimated by k_1. Chen, Dolado and Gonzalo (2014) prove that a factor model with r unobserved factors and with $0 < k_1 \le r$ big structural breaks in the factor loadings at time τ admits a representation with (asymptotically) $r + k_1$ factors. Particularly, given an IC estimator in Bai and Ng (2002) $\hat{r}$ and under general assumptions, it is shown (Proposition 2, p.34):

$$\lim_{N,T \to \infty} \mathbb{P}\left[\hat{r} = r + k_1\right] = 1. \tag{3.11}$$

An important remark at this point is to notice that if the break date τ were known, one could recover a consistent estimate of r by simply splitting the sample in a "before-breakpoint" and "after-breakpoint" subsamples and performing PCA and Bai and Ng (2002) or Ahn and Horenstein (2014) in either subsample. In either case,

$$\lim_{N,T \to \infty} \mathbb{P}\left[\hat{r}_{before} = r\right] = 1 \tag{3.12}$$

$$\lim_{N,T \to \infty} \mathbb{P}\left[\hat{r}_{after} = r\right] = 1.$$

both $\hat{r}_{before}$ and $\hat{r}_{after}$ typically lower than the full sample estimate $\hat{r}$.

For the sake of generality, we take the exact breakpoint date τ as unknown. Although we explicitly consider the exact date of the finalization of the Volcker Rule in the difference-in differences matching below, the possibility of anticipatory behavior or of delayed response for a policy intervention so sizeable and publicly debated would caution against a 'known breakpoint' approach. Hence, we do not impose such restriction here.

Chen, Dolado and Gonzalo (2014) present a test for the null $H_0 : k_1 = 0$ versus the alternative of at least one big break $H_1 : k_1 > 0$ based on detecting breaks in $\hat{F}_t$ estimated over the full sample by PCA. The implementation is straightforward. Define $\hat{\beta}$ the estimated $(\hat{r} - 1) \times 1$ coefficient vector obtained by regressing $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ and $\hat{S}$ its corresponding Newey-West HAC covariance matrix[45]. One can test for structural breaks in β by focusing for the case of unknown breakpoint $\tau = T\pi$ with $\pi \in \Pi \equiv (\pi_0, 1 - \pi_0)$ and $0 < \pi_0 < 1$ based on Andrews (1993) Sup-Wald statistic or Sup-LM statistic. Specifically, for given τ, and hence $\pi = \tau/T$, define $\hat{\beta}_1(\pi)$ the estimated $(\hat{r} - 1) \times 1$ coefficient vector obtained by regressing $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ for $t = 1, ..., \tau$ and $\hat{\beta}_2(\pi)$ the estimated $(\hat{r} - 1) \times 1$ coefficient vector obtained by regressing $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ for $t = \tau + 1, ..., T$ the Sup-Wald statistic is:

$$\mathscr{L}^*(\Pi) = \sup_{\pi \in \Pi} T\pi(1 - \pi)\left(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)\right)' \hat{S}^{-1}\left(\hat{\beta}_1(\pi) - \hat{\beta}_2(\pi)\right) \tag{3.13}$$

and the Sup-LM statistic is:

$$\mathscr{L}(\Pi) = \sup_{\pi \in \Pi} \frac{1}{\pi(1 - \pi)}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi} \hat{F}_{-1t}\hat{F}_{1t}\right)' \hat{S}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi} \hat{F}_{-1t}\hat{F}_{1t}\right) \tag{3.14}$$

---

[45]Newey and West (1987). $\hat{S}$ is estimated over the full sample.

In the analysis we will maintain a conservative $\pi_0 = 0.3$ which in our case is not overly restrictive as it allows a search for structural breaks between January 2008 and January 2012 covering the full financial crisis, the full legislative debate on Dodd-Frank and large part of the regulatory rulemaking period for the Volcker Rule. We employ the critical values for the (3.13) and (3.14) statistics reported in Andrews (1993).

To conclude this subsection, let us consider the matter of detecting a structural break in the factors themselves as opposed to a break in the factor loadings at τ. There are at least two different formulations for a break in the factors one should consider. First, the formulation discussed in Chen, Dolado and Gonzalo (2014) considers maintaining unvaried the loadings Λ, but changing the variance-covariance matrix of the r original factors:

$$\mathbb{E}\left[F_t F_t'\right] = \Sigma \quad t = 1, ..., \tau \tag{3.15}$$

$$\mathbb{E}\left[F_t F_t'\right] = \Xi \quad t = \tau + 1, ..., T \tag{3.16}$$

where Σ is the factor covariance before the break and Ξ after the break and both are $(r \times r)$. Given that the approach above focused on testing breaks in the $\hat{F}_t$ PCA factors estimated over the full sample, it may not appear surprising that the Sup tests above (based on regressing $\hat{F}_{1t}$ on $\hat{F}_{-1t}$) will be naturally able to pick up breaks of the type (3.15)-(3.16). In fact, the same regression approach described above will reject the null of big breaks in presence of changes in factors.

It is possible however to discriminate between breaks in loadings and breaks in factors by noticing that in the case of breaks in factors:

$$\lim_{N,T \to \infty} \mathbb{P}\left[\hat{r} = r\right] = \lim_{N,T \to \infty} \mathbb{P}\left[\hat{r}_{before} = r\right] = \lim_{N,T \to \infty} \mathbb{P}\left[\hat{r}_{after} = r\right] = 1. \tag{3.17}$$

This implies that in the case of breaks in the factors typically $\hat{r}$ estimated over the whole sample will be identical as when estimated on subsamples either before or after the breakpoint. In the case of breaks in the loadings, instead, $\hat{r}$ estimated over the full sample will be higher than when estimated on subsamples either before or after the breakpoint, as evident from the result in (3.11).

A second formulation for a break is more drastic and entails a break in the number of factors r in (3.5), that is the addition or subtraction of specific factors in the model at date τ. Section 3.3.4 offers an application of this methodology to this formulation and shows how it can be incorporated in this setting.

3.3.3 Multiple Breakpoint Testing for Dynamic Factor Models

Let us now focus on multiple structural breaks M in factor loadings at unknown dates $\tau_1, \tau_2, ..., \tau_M$. This structure partitions the sample period of length T in $M + 1$ intervals:

$$X_t = \Lambda F_t' + \varepsilon_t \quad t = 1, ..., \tau_1 \tag{3.18}$$

$$X_t = \Gamma^1 F_t' + \varepsilon_t \quad t = \tau_1 + 1, ..., \tau_2$$

$$...$$

$$X_t = \Gamma^M F_t' + \varepsilon_t \quad t = \tau_M + 1, ..., T$$

where Γ^m with $m = 1, ..., M$ are the post first break matrices of factor loadings of dimension $(N \times r)$. In the context of multiple breakpoints, standard estimators in the literature include the ones proposed by Bai and Perron (1998, 2003), which we employ in combination to the regression approach delineated in Section 3.3.2. Considering the regression of $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ with the goal of detecting not one, but multiple breakpoints, we implement the recommended approach of Bai and Perron (1998, 2003).

Consider for the interval $t = \tau_m + 1, ..., \tau_{m+1}$ the regression of $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ in this subsample and call the estimated coefficient $\hat{\beta}_m$. Notice that, like $\hat{\beta}_1(\pi)$ and $\hat{\beta}_2(\pi)$ in Section 3.3.2, $\hat{\beta}_m$ depends on the breakpoint

parameters, $\pi_m = \tau_m/T$ and $\pi_{m+1} = \tau_{m+1}/T$. Given M, let us also define $\hat{\beta} = \left(\hat{\beta}'_1, \hat{\beta}'_2, ..., \hat{\beta}'_{M+1}\right)'$. Bai and Perron (1998) first consider the Sup-F type test of the null hypothesis of no structural break ($M = 0$) against the alternative hypothesis that there is a known number of breaks $M = k$:

$$\sup_{(\pi_1,...,\pi_k)} F_T\left(\pi_1, ..., \pi_k; r-1\right) \tag{3.19}$$

$$= \frac{1}{T} \left(\frac{T - (k+1)(r-1)}{k(r-1)}\right) \hat{\beta}' R' \left(R\hat{S}R'\right)^{-1} R\hat{\beta}$$

where R is the matrix such that $\left(R\hat{\beta}\right)' = \left(\hat{\beta}'_1 - \hat{\beta}'_2, ..., \hat{\beta}'_k - \hat{\beta}'_{k+1}\right)$ and $\hat{S}$ is now an estimated HAC variance covariance matrix of $\hat{\beta}$ [46].

As the number of breaks is unknown, a second type of test is more useful: Bai and Perron (1998) consider a test of the null hypothesis of no structural break ($M = 0$) against the alternative hypothesis that there is an unknown number of breaks $M = m$ with m ranging between 1 and $\bar{m}$, which is given[47]. The test is referred to as the double maximum test and two different statistics are employed:

$$UD\max F_T\left(\bar{m}; r-1\right) = \max_{1 \leq m \leq \bar{m}} \sup_{(\pi_1,...,\pi_k)} F_T\left(\pi_1, ..., \pi_k; r-1\right) \tag{3.20}$$

which is unweighted with respect of each break number, and

$$WD\max F_T\left(\bar{m}; r-1, a_1, ..., a_{\bar{m}}\right) = \max_{1 \leq m \leq \bar{m}} a_m \sup_{(\pi_1,...,\pi_k)} F_T\left(\pi_1, ..., \pi_k; r-1\right) \tag{3.21}$$

which is a weighted version, where weights are defined such that the marginal p-values are equal across values of m [48].

The final test proposed by Bai and Perron is a sequential test. One proceeds by testing ℓ breaks against $\ell+1$ breaks. The test is commonly labelled $\sup F_T\left(\ell+1|\ell\right)$ and intuitively is built as follows. Consider the $\ell+1$ intervals generated by the ℓ break points under the null hypothesis. Within each interval a separate test of the type $\sup_{(\pi_1)} F_T\left(\pi_1; r-1\right)$ is run, i.e. a test of the null hypothesis of no break versus the alternative hypothesis of 1 break. The test rejects the null hypothesis in favor of $\ell+1$ breaks if, relatively to the sum of squared residuals obtained under the ℓ breaks model obtained by regressing $\hat{F}_{1t}$ on $\hat{F}_{-1t}$ and aggregated across all intervals, there is one additional break that produces a sum of squared residuals sufficiently smaller under the $\ell+1$ breaks model.

Bai and Perron (2003) recommend to first obtain both the $UD\max$ and $WD\max$ to test whether at least one break is detected in the entire sample, as these tests are more prompt in rejecting the null hypothesis in presence of multiple but contiguous breaks (e.g. which would be the case for instance if there were a break at the beginning of the crisis and one at its end). If at least one break is detected, then the sequential approach should be employed. Specifically one should select $M = m$ such that $\sup F_T\left(\ell+1|\ell\right)$ are insignificant for $\ell \geq m$. We follow this approach here.

---

[46]In the tests we perform we apply a short trimming of 10%. The Bai and Perron requires a minimal admissible distance expressed as fraction of T among any pair of breakpoints τ_m and τ_{m+1} and we set it to 10% of the sample length, in order to allow for relatively close multiple breaks. In all the test we also allow the distribution of ε_t to vary across different intervals.

[47]In the tests we perform we allow for a maximum of $\bar{m} = 5$ total breakpoints (which, as shown below, will prove to be sufficiently high and is also the value suggested in Bai and Perron, 2003).

[48]Specifically $a_1 = 1$ and $a_m = c(r-1, \alpha, 1)/c(r-1, \alpha, m)$, for $m > 1$, where α is the significance level of the test and $c(r-1, \alpha, m)$ is the asymptotic critical value of the corresponding Sup-F test for m breaks, which is reported by Bai and Perron (1998, 2003).

3.3.4 Breaks in Trends and a Simulation Example

We first provide a simple example to illustrate the flexibility of the dynamic factor model to capture breaks in trends, which are a realistic type of structural break in our setting. Suppose the illiquidity measure, l_t, is jointly driven by supply of liquidity, s_t, and demand for liquidity, d_t. Suppose that post-crisis regulations lead to a upward trend with a constant drift γ in illiquidity from $\tau + 1$:

$$l_t = -\alpha s_t + \beta d_t + e_t \quad t = 1, ..., \tau \tag{3.22}$$

$$l_t = -\alpha s_t + \beta d_t + \gamma(t - \tau) + e_t \quad t = \tau + 1, ..., T \tag{3.23}$$

Taking the first difference of the above equation system gives:

$$x_t = -\alpha f_{1t} + \beta f_{2t} + 0 f_{3t} + \varepsilon_t \quad t = 1, ..., \tau \tag{3.24}$$

$$x_t = -\alpha f_{1t} + \beta f_{2t} + \gamma f_{3t} + \varepsilon_t \quad t = \tau + 1, ..., T \tag{3.25}$$

Where $x_t = l_t - l_{t-1}$ is the innovation in illiquidity, $f_{1t} = s_t - s_{t-1}$ is the supply factor, $f_{2t} = d_t - d_{t-1}$ is the demand factor, $f_{3t} = 1$ is the regulation factor, and $\varepsilon_t = e_t - e_{t-1}$ is the differenced measurement errors. It is immediately obvious that the break in trend can be reformulated as a break in the loading on the regulation factor, which can be consistently estimated by our methodology, as shown in Section (3.3.2).

We simulate a panel of 180 liquidity measures to illustrate the power of our tests[49]. A detailed discussion of simulation and the Monte Carlo evidence of power and size of the tests can be found in Chen, Dolado and Gonzalo (2014). Figure 3.2 plots the simulated liquidity index, defined as the average of 180 standardized simulated liquidity measures. The blue solid line is the path with the structural break, and the green dotted line plots the counterfactual scenario where regulation has no effects by design. The star sign indicates the date when the structural break happens. The difference between the two paths is the regulation-induced liquidity gap. We can see that the magnitude of liquidity deterioration is very small at the beginning compared with the normal fluctuations of liquidity, and builds up very slowly. We conduct our structural break tests described in Section (3.3.2) and (3.3.3). The estimated break date is marked by the vertical dashed line. Despite of the small magnitude, both tests successful identify the date of the structural break.

We also use the dynamic factor model to estimate the counterfactual path of liquidity assuming there is no structural break. We first use the observed data before the break to estimate the loadings. Specifically, we regress each of the 180 liquidity measures on the estimated factors. Then we predict the counterfactual path of liquidity assuming the factor loadings in the post-break period are the same as the pre-break period. The red dash line shows the estimated path. Our estimation accurately traces out the true counterfactual path. Such accuracy is obtained because the large cross-section dimension ($N = 180$) of our liquidity measures compensate the relatively short time span for loading estimation (62 months).

3.4 Results for Market Liquidity of U.S. Corporate Bonds

For U.S. corporate bonds we present four different estimation strategies. We will begin by applying multiple breakpoint tests in levels to measures of market liquidity provisions. Subsequently, we will focus on a dynamic factor model and presents results of both single and multiple breakpoints in factor loadings, with the understanding that also further testing for factor breaks is available. Finally, we will focus on difference-in-differences matching results.

---

[49]To mimic our empirical application, we simulate 180 liquidity measures driven by two latent factors: a supply factor and a demand factor. The two factors follow AR(1) process with autocorrelation of 0.5, and cross-correlation of 0.5. The loading parameters on the two latent factors are drawn from N(0,1). A structural break occurs in July 2010 where 180 liquidity measures start to load on a new regulation factor, which follows AR(1) process with autocorrelation of 0.5 and an upward drift of 0.1. The loading parameters on the regulation factor follows N(0,0.2). The cross-correlation between regulation and supply and demand factor is also 0.5.

3.4.1 Multiple Breakpoint Tests for Liquidity Levels

We begin by studying break in levels of our main nine liquidity measures (or properly seven measures of liquidity levels and two measures of liquidity risk) employing the Bai and Perron (1998, 2003) estimation approach for multiple unknown breakpoints in the undifferenced and unstandardized time series. This simple test serves as a visual examination on the potential dates around which liquidity depletion may have occurred.

At the onset we do not separate bonds by underwriter, issue size, and credit rating. Rather we aggregate all bonds and plot their time series in Figure 3.3[50]. The estimated means for each sub-period (red dashed line) are also reported, where the break dates (a shift in the red dashed line) are estimated by the Bai and Perron (1998-2003) approach and are breaks significant at the 5% confidence level[51].

Concerning the dating of the structural breaks, the estimators should pick up at least the drastic reduction in liquidity produced by the near collapse of the U.S. financial system in September 2008 and the subsequent break towards more normal market liquidity levels at the end of 2009. Any detection of subsequent structural breaks towards lower levels of liquidity over the periods 2010-2014 needs instead to be carefully examined, as potential telltale indication of liquidity depletion concurrent with (and possibly caused by) regulatory intervention. The 2010-14 period covers important regulatory events such as the approval of Dodd-Frank, shutdowns of proprietary trading desks by major banks, Basel III, and the approval of the interim and the finalized Volcker Rule.

The double maximum tests indicate the presence of at least one structural break at 5% confidence level in all nine proxies[52]. The sequential $\sup F_T (\ell + 1|\ell)$ indicates three breakpoints for the IRC, IRC (standard deviation), Roll measure, and Non-block trades; one for the Amihud, Amihud (standard deviation), and Turnover (negative), and four for Size (negative) and Zero trading. As clarified by Figure 3.3, the Bai-Perron approach indicates clearly breaks in liquidity around the financial crisis. None of the structural breaks towards lower liquidity happen during the period of regulatory intervention. Instead, breaks towards higher liquidity are detected for seven out of nine liquidity measures.

We further compare the estimated mean liquidity in the subperiods before and after the crisis. With the possible exceptions of turnover and non-block trades, most of the liquidity measures indicate higher liquidity levels at the end of the sample period comparing to the pre-crisis level: the price impact of large transactions goes down (Amihud), bid-asked spreads tighten (IRC), the price reversal goes down (Roll), the median trade size stays stable, and trading becomes more frequent (Zero trading). Turnover and block trade are somewhat lower than the pre-crisis level, but the breaks occurred before or during the crisis, well before regulations came into place. In fact, using the aggregate bond turnover statistics from SIFMA we find that corporate bond turnover has been on a downward trend for more than ten years, and actually flattens out during the post-crisis period[53], suggesting factors other than post-crisis regulation may be the driving force. For example, an increasing share of corporate bond trading may have moved to bond ETFs which is not captured by TRACE data[54]. The reduction in the share of block trades may be driven by market structure transition from over-the-counter market to electronic trading platforms where transactions are conducted predominantly as non-block trades (Hendershott and Madhavan, 2015). Even the share of block trade seems to be lower, the median trade size is similar to the pre-crisis level.

While this is prima facie evidence against drastic reductions in liquidity following regulatory interven-

---

[50]In the online appendix Figure 2, we create an aggregate liquidity index using the average z-score of 9 liquidity measures. This approach helps to average out some noises from a particular liquidity measure but may lose some detailed information. We apply the same analysis on this liquidity index, and rearch similar conclusion if we examine each measure separately.

[51]The estimated break dates are reported in online appendix Table 2.

[52]In online appendix Table 3 and 4 we report the relevant statistics for the double maximum tests and the $\sup F_T (\ell + 1|\ell)$ tests.

[53]The result is reported in online appendix Figure 3.

[54]In online appendix Figure 3, we adjust the turnover by adding the trading volume of corporate bond ETFs. The trading volume from ETFs accounts a non-trivial share of decline in turnover, especially for high yield bonds.

tion, it is still possible that at the level of specific types of corporate bonds structural breaks may arise. In Figure 3.4 we present a graph tracing for each month the fraction of the 180 disaggregated market liquidity variables that are described to have a statistically significant (at 5% confidence level) break in that month and in what direction (i.e. towards lower liquidity -in blue- or higher liquidity -in red). The bulk of the structural breaks toward lower liquidity happens in July and August 2008, right before Lehman Brothers' failure. As it appears clear in Figure 3.4, if anything, around subsequent periods of regulatory intervention the disaggregate liquidity measures pointed systematically toward higher liquidity, not lower.

To understand the source of the disaggregate-level structural breaks, Figure 3.5 shows the decomposition of break dates by underwriting bank[55]. We can see that the bankruptcy of Lehman Brothers in September 2008 caused liquidity reductions for all underwriters. In comparison, the later recoveries are more heterogenous: bonds underwritten by JP Morgan and Goldman Sachs experienced earlier recovery in liquidity than bonds of other underwriters. This is consistent with anecdotal evidence that these two banks had relatively stronger balance sheets throughout the crisis.

The most important observation from this graph, however, is from the later period when banks start to shutdown their proprietary trading desks after the passage of the Dodd-Frank Act. Were proprietary trading indispensable for market making, one would expected to see bank-specific liquidity reductions line up with an announced trading desk shutdown by the same bank. This is hardly the case: no large bank specific liquidity reduction is observed after 2010 (all the bank-specific frequencies of liquidity reduction are below 5% after 2010)[56]. On the contrary, many banks experienced liquidity increases around July 2012, in the midst of regulatory interventions. There appears to be no clear evidence that the shutting down of proprietary trading desks was associated with an adverse impact on market liquidity.

3.4.2 Single Breakpoint Tests for the Dynamic Factor Model

This subsection shifts the attention to a dynamic factor model with the goal of assessing whether the underlying structure of correlation and of latent dynamics of liquidity across different bond types displays salient breaks during the period of crisis and post-crisis regulatory intervention. Comparing to the breakpoint tests for liquidity levels in the previous subsection, this approach allows more realistic modeling of the liquidity processes and captures more flexible forms of breaks, including breaks in trends, in serial correlation, and in factor loadings.

We discuss here the application of Chen, Dolado and Gonzalo (2014) using the 2005-14 monthly sample and our full matrix X of $N = 180$ differenced and standardized time series. A first preliminary step requires to estimate the number of factors over the full sample $T = 117$. According to our discussion in Section 3.3.2 this approach will not deliver a consistent estimate of the number of true factors in (3.5), but rather the sum of the true factors r and the number of big breaks in these factor loadings k_1. In online appendix Table 5 we report the full set of estimates based on Bai and Ng (2002) and Ahn and Horenstein (2014). Here we impose a $kmax = 10$ and notice that the estimates for $\{IC_{p1}, IC_{p2}, IC_{p3}, PC_{p1}, PC_{p2}, PC_{p3}, AIC_3, BIC_3, ER, GR\}$ range from 3 to 10. Although this range is not particularly tight, this is of little effect for the interpretation of our main findings in Figure 3.6.

Figure 3.6 reports the Sup-Wald and the Sup-LM test statistics of the full interval over which the unknown breakpoint is allowed to belong given a conservative $\pi_0 = 0.3$. Such sample restriction is due to power loss concerns for the Sup tests (Andrews, 1993). Our interval of search of breakpoints covers the period between January 2008 and January 2012. Figure 3.6 also reports the Andrews (1993) critical values

---

[55]In online appendix Figure 4 and 5, we show the decomposition by types of bonds and measures of liquidity. The results are consistent.

[56]A gradual shutdown of the trading desk would not be a problem for our test, since the estimated break points will show up sometime after the announcement date. However, we see none of this lagged liqudity reduction.

above which the structural break is significant at the 10% and 5% confidence. We perform the analysis for any possible number of factors in the range estimated in online appendix Table 5.

As evident from Figure 3.6, the Sup tests systematically pick breaks in factor loadings (at 5% confidence) when we allow a number of estimated factors above 4. Typically the Sup statistic indicates the breakpoint as occurring during the 2008-2009 recession or shortly after. This is informative because again such dating does not correspond to regulatory events of prominence, but rather corresponds to the financial crisis itself. In essence what the Chen, Dolado and Gonzalo (2014) methodology allows us to exclude is that a structural break in the underlying factor structure of the disaggregate liquidity occurred around dates of post-crisis regulatory activity[57].

So far the methodology in this subsection has focused on a single breakpoint, a restriction that, given the multitude of potential shocks affecting the U.S. financial system during our period of analysis, one should find unwarranted. We relax this restriction in the following subsection.

3.4.3 Multiple Breakpoint Tests for the Dynamic Factor Model

This subsection employs the Bai and Perron (1998, 2003) approach within the dynamic factor model, transposing the logic of Chen, Dolado and Gonzalo (2014) to the multiple breakpoint setting.

Figure 3.7 reports the results. Each panel represents a different factor models ranging from $\hat{r} = 2, ..., 10$ estimated factors, employing the Bai and Perron (1998, 2003) preferred approach to the first $\hat{r}$ PCA estimated factors of the matrix X of differenced and standardize disaggregate liquidity measures. The blue solid line represents the liquidity index, defined as the average of 180 standardized liquidity measures. The dashed vertical lines indicate estimated dates of breaks in dynamic factor model[58]. The double maximum tests indicates the presence of at least a structural break at the 5% confidence level in all nine dynamic factor models. The sequential $\sup F_T (\ell + 1|\ell)$ indicates at most two breakpoints for the models with $\hat{r} = 2, 3, 4, 5$, all essentially coincident with the start and end of the recession and the financial crisis. As in the previous section, such dating occurs well before regulatory events of prominence (the passage of the Dodd-Frank Act in July 2010, the announcement of the final rules of Basel III in July 2013, or the announcement of the finalized Volcker Rule in January 2014), but rather appears to correspond to dynamics within the confines of the financial crisis itself.

With $\hat{r} = 6, 7, 8, 9, 10$, more breakpoints in the factor loadings appear. Notably, there are breaks in late 2010 and 2011, which fall into the regulatory intervention period. To examine whether these breaks indicate deterioration or improvement in liquidity, we estimate the counterfactual path of liquidity assuming no structural breaks occur from the last estimated breakpoint onwards. Specifically, we first estimate the factor loadings using the data in the interval immediately before the structural break. Then we predict the counterfactual path of the average liquidity after the break assuming the factor loadings take the same value as before. For the models which do not detect breaks during the regulatory intervention period ($\hat{r} = 2, 3, 4, 5$), we use the break closest to the regulatory intervention to conduct the counterfactual analysis. We conduct this exercise for each of the 180 liquidity measures, and take the average to create a liquidity index.

The red dashed line in Figure 3.7 is the estimated counterfactual path of liquidity in absence of the last structural break. Comparing the observed and counterfactual path, we can tell that whether liquidity would be lower or higher in absence of the breaks. Consistently across all the nine specifications, these structural breaks during or right before the regulatory intervention period lead to slightly higher liquidity (lower illiquidity as the figure shows) comparing the counterfactual path. This is consistent with Figure 3.4 which shows the level of liquidity breaks towards higher liquidity, not lower around this time period. One

---

[57]In online appendix Table 6, we report the number of factors before and after the break.

[58]In online appendix Table 7, 8, 9, and 10 we report the estimated break dates, double maximum test statistics, the $\sup F_T (\ell + 1|\ell)$ test statistics, and the number of factors in each subperiod.

likely explanation could be the ability of our model to pick up an increasing role for electronic trading and for open-end mutual funds[59].

3.4.4 Difference-in-Differences Matching for Liquidity Levels

We now present a more standard estimation strategy based on a difference-in-differences exercise augmented by matching of corporate bonds based on pre-treatment covariates (Heckman, Ichimura, Smith, and Todd, 1998; Smith and Todd, 2005). Here, for reason that will become clear in the construction of the test, we will focus only on the finalization of the Volcker Rule in January 2014 as our treatment date. Given the limitation in our "post" sample of just 12 months available, we will take a symmetric 12-month window around January 2014[60].

We proceed as follows. First, we manually classify the top 40 underwriters into two groups –one covered by Volcker Rule and the other not covered based on the revised finalized version of Volcker Rule[61]. Then we identify the set of bonds which have at least one underwriter not covered by the Volcker Rule, that is a non-banking entity for which proprietary trading is not restricted. This set of bonds is a useful benchmark as at least one of the underwriters who typically make market on that bond is virtually unconstrained by the main regulatory restriction in the rule, and hence virtually free to provide liquidity services in case banking entities were so impaired. For each of these $3,106$ non-Volcker Rule bonds that are outstanding between January 2013 and December 2014, we find a match among all the Volcker Rule bonds issues in the same month, matures in the same month, has the same credit rating (investment grade/high yield), and has a relative size difference less than 50% of the average size of the pair[62].

Table 3.3 reports the results for a difference-in-differences model for each of our nine liquidity proxies where the treatment is administered to the Volcker Rule bonds after January 2014 and each regression controls for the reciprocal of issue age, the reciprocal of issue age squared, bond fixed effects, and month fixed effects. Standard errors are two-way clustered at the bond and month level. In seven out of nine measures the treatment does not predict reductions in liquidity with a confidence level of 5%. Only for IRC and IRC (standard deviation) we find a statistically significant effect. This is not particularly worrying since more than 80% observations have missing IRC and IRC (standard deviation) measure in the matched sample[63]. Overall, there seems to be no robust evidence of liquidity depletion as consequence of the Volcker Rule.

The regression evidence is also supported by the graphical representation. In Figure 3.8 we show the time series of the Volcker Rule bonds and non-Volcker Rule bonds around the time when the revised finalized version of the Rule was approved (the vertical line, January 2014). Both time series are normalized to take value of 0 at December 2013. Were evidence of liquidity depletion present in the data, one would expect to see systematically higher levels of the blue line after the treatment, a sign of reduced liquidity or heightened

---

[59]See also Dudley (2015). In online appendix Table 11, we show that some of the latent factors are indeed significantly correlated with innovations in bond mutual fund flows.

[60]Relative to the analysis above, the approach of this subsection is more restrictive, as it focuses on a single regulatory dimension and relies on a difference-in-differences type of identification, but it is also an approach much more familiar to applied econometricians. In addition, by relying on different identifying assumptions, complements nicely the macroeconometric estimation strategy above.

[61]See the the following document from Federal Register for details of the final rule: http://www.gpo.gov/fdsys/pkg/FR-2014-01-31/pdf/2013-31476.pdf

[62]There are fewer non-Volcker Rule bonds so we start our matching with them. If more than one bond satisfies the above criteria, we keep the one with smallest relative size difference. Since the Volcker Rule bonds are significantly larger than non-Volcker Rule bonds, many observations are dropped due to the last criterion on relative size. We ended up with a matched sample of 350 pairs of bonds.

[63]This is because the non-Volcker Rule bonds are usually very small and thinly traded. Therefore, liquidity measures which require a certain number of transactions of specific types are very noisy.

liquidity risk. This is hardly the case both in reporting unconditional time series as in Figure 3.8 or time series where bond and month fixed effects are conditioned out (not reported to save space).

3.4.5   Comments on the Change of Market Structure

With systematic evidence supporting the absence of structural deterioration in corporate bond liquidity, we will now conclude this section by going back to two of the most often cited evidences for liquidity depletion: the decline in dealer corporate bond inventories and the increase in agency trading.

Figure 3.9 shows the amount of corporate bonds held by dealer banks as the percentage of total corporate bond outstanding. We apply the Bai and Perron (1998, 2003) approach to estimate break points to this series, and three lessons can be learned from this test.

First, the estimation shows, as is obvious in observing the time series of the raw data, that the major reductions in dealer inventories occurred at the onset of the financial crisis (September 2008), far ahead of the post-crisis financial regulation. Therefore, at a minimum, there are other important factors driving the reductions of the inventories unrelated to the post-crisis financial regulation. One potential factor is the deleveraging of broker-dealers forced by rehypothecation lenders (Mitchell and Pulvino, 2012).

Second, the abnormally high level of bond holdings in 2007 seems the result of a pre-crisis run-up of risk-taking, as shown by a series of breaks towards greater holding amounts between 2002 and 2007. In this light, the dramatic reduction during the crisis appears actually more a "getting back to normal". In this sense, using the pre-crisis level as a baseline to calculate the change of inventory is somewhat misleading.

Third, there are two minor breaks, one in August 2011 and the other in March 2013, that fall into the period of regulatory intervention. However, as our tests on market liquidity have systematically shown, no structural reductions in market liquidity occurred during this period. This seems to suggest that some of the holdings may be held by the proprietary trading desks for risk-taking purposes, exactly the kind of activities that the Volcker Rule restrains. This possibility has also being raised by informal discussions (Brainard, 2015)[64]. The fact that dealer banks rapidly reduced their bond holding during 2008-09 crisis suggests that they demanded rather than supplied liquidity at the time when liquidity was most needed[65]. Another possibility is that this data series from the Federal Reserve overstates pre-crisis inventories because it improperly includes non-agency MBS. Indeed, one of the post-crisis breaks corresponds to the date of survey revision[66].

Another commonly cited evidence of liquidity depletion is the shift from principal-based transactions to agency-based transactions. Principal and agency transactions are two main types of trade that dealers may conduct. Principal trading occurs when a dealer uses its own inventory to fill the order for the client. The purpose behind principal trading is for the dealer to create extra profits (over and above the commission charged) for its own portfolios through price appreciations and bid-ask spreads. Traditionally, large banks have mainly focused on principal trading. Agency trading instead involves a dealer searching for the security demand by a client from other clients or dealers. It is an empirical question whether regulation has caused the shift to agency trading, and it is also unclear that such shift of business model would lead to liquidity deterioration.

Figure 3.10 plots the fraction of agency transactions over time[67]. We apply the Bai and Perron (1998, 2003) approach to estimate break points in the level of this series. Coincidently with the decline in bond

---

[64]In a speech by Federal Reserve Governor Lael Brainard at Salzburg Global Forum on July 1, 2015, he also mentioned that "since not all broker-dealer inventories are used for market-making activities, the extent to which lower inventories are affecting liquidity is unclear."

[65]We thank Albert Kyle for suggesting this point.

[66]See "Revised survey of primary dealers sheds new light on inventories," *The Credit Line*, April 18, 2013.

[67]TRACE does not disseminate the agency trade indicator. We create a proxy which equals to 1 if two or more transactions of the same bond with the same volume and at the same price happen at the same time. See Dick-Nielsen (2009) for a detailed discussion for measuring agency trades with the TRACE database.

inventory, we find that there is a secular increasing trend of agency-based transactions, and the bulk of the increase occurred before regulatory interventions. The timing casts doubts on the claim that the post-crisis regulation causes this change. Moreover, comparing to the time series of our liquidity measures in Figure 3.3, the increase in agency-based transactions does not line up with periods of liquidity reductions, suggesting the two are not necessarily equivalent[68].

A more interesting question is what explains the structural breaks towards higher liquidity levels during the regulatory intervention period. We suggest that post-crisis regulation, by encouraging competition in market-making, could be a contributing factor. The idea is that big banks used to enjoy a big funding advantage over non-bank entities in corporate bond market-making business due to explicit (e.g. deposit insurance) and implicit (e.g. too-big-too-fail status) subsidies from the government. The funding advantage of big banks generated an entry barrier for non-bank entities to compete in this capital intensive business. If post-crisis regulations by and large reduced the funding advantage of big banks, this might have led to a level playing field for non-bank entities to compete. As a result, more players can now enter the market, and increased competition should induce a downward pressure on the price of intermediacy.

There is evidence consistent with this explanation. The average number of competing market-makers trading a bond has increased by 40% from the period of July 2007-April 2009 to the period of May 2009-May 2014 (Bessembinder et al. 2016). Competition between trading venues has also intensified: bond ETFs and electronic trading platforms such as MarketAxess provide investors with cost-effective ways to trade corporate bonds outside the OTC market dominated by big banks. Some market commentators also express a similar view. For example, on July 26, 2015 The Wall Street Journal in an article titled "Overlooking the Other Sources of Liquidity" reports "*Missing from much of this debate, however, is recognition of the radical transformation that has taken place in many fixed-income markets as barriers to entry have fallen and new liquidity providers have stepped forward.*"[69] With more non-bank entities entering the market-making business, overall liquidity supply may increase, and sources of liquidity supply may become more diversified. In this sense, post-crisis regulation might have actually made market liquidity more resilient. This perspective is often missing in the post-crisis policy debate and definitely requires further investigation beyond the scope of this paper.

3.5 Results for Market Liquidity of U.S. Treasuries

This section extends our analysis to the U.S. Treasuries market. Much of the interest and the discussion pertinent to this market's liquidity can be ascribed to the salience of events like the flash crash of October 15th, 2014 when the yield of the U.S. 10-year note dropped by 34 basis points from 2.2% to 1.86% in the eight minutes between 9:33 and 9:45AM Eastern Time.

In Table 3.4 we report the summary statistics for this asset class, including Noise, On-the-run premium, Roll measure (all expressed in basis points) and Turnover (negative) over the April 2005-December 2014 sample[70], again calculated at the monthly frequency. The correlations among these proxies are intuitively positive, with the exception of Turnover (negative), as reported in Table 3.5. The reason for this counterintuitive negative correlation is given by the construction of the measure for the Treasuries. As the denominator in the Turnover variable is the total stock of public debt outstanding, the explosion of U.S. sovereign debt as

---

[68]Even if agency transactions may not directly impact liquidity, a legitimate concern is that it may bias the measurement of liquidity. To address this concern, we drop all agency trades and repeat our tests exclusively on principal-based transactions. We still find no systematic evidence of liquidity deterioration. The results are available upon request.

[69]See also an industry discussion panel titled "Are There Structural Issues in the U.S. Bond Market?" organized by the Brookings Institute for discussion on regulation and competition in market-making business.

[70]In the online appendix Figure 6, we extend our analysis to an earlier sample period from 1995 to 2005, which covers the collapse of LTCM, a liquidity crisis much smaller in scale comparing to the 2008-09 financial crisis.

consequence of the automatic stabilizers and the 2009 Fiscal Stimulus appear to severely affect the quality of this measure post 2009, an issue that will become clearer below.

We employ the Bai and Perron (1998, 2003) approach to estimate breakpoints in the level of the four liquidity time series: Noise, On-the-run premium, Roll measure and Turnover (negative)[71]. The corresponding double maximum tests indicates the presence of at least one structural break at the 5% confidence level in all four proxies, with the exception of the UDmax for the Noise variable. However, for the same variable WDmax reject the null that there is no break. The sequential $\sup F_T (\ell+1|\ell)$ indicates three breakpoints for the Noise and Roll measures, one for the On-the-run premium and four for the Turnover (negative). Figure 3.11 reports an informative visualization of when the breakpoints happen over time and in which direction the series breaks. For both the Noise and Roll measures this approach clearly captures the sudden deterioration of market liquidity around the 2008-09 financial crisis and a return to normality mid-2009. The Roll measures seems to suggest further liquidity amelioration in December 2011 (in fact close to the release of the first Proposed Volcker Rule published in November 2011). The On-the-run premium exhibits qualitatively very similar dynamics, as evident from the North-East panel in Figure 3.11, but our approach fails to pick up a structural break at the start of the crisis. The only proxy that seems to systematically break in terms of lower liquidity levels for Treasuries is Turnover (negative) in October 2008. However, looking at the components of this measure, this result appears mainly driven by two factors: 1. Treasury issuance dramatically increased after 2008. 2. The Federal Reserve balance sheet structurally increased, holding a very large portfolio of public debt due to the Quantitative Easing. Since the Fed typically is not actively trading, the turnover should intuitively drop.

3.6 Conclusions

This paper complements, both methodologically and substantively, a rigorous retrospective analysis of post-crisis regulatory intervention in domestic financial markets. Such analysis has been surprisingly bare in terms of systematic empirical evidence and it appears to be a necessary exercise in informing future legislative and rulemaking activities aimed at improving financial markets stability (Cochrane, 2014).

We specifically focus on the aftermath of the 2008-09 U.S. financial crisis and on the role played by the Dodd-Frank Act of 2010 and Basel III as potential triggers of liquidity shortages driven by retrenchment of financial institutions adversely affected by overreaching regulation. Several market participants have claimed this assessment to be crucial in the context of an informed cost-benefit analysis of regulatory intervention and rulemaking.

We initially focus on a large set of liquidity proxies with emphasis on the U.S. corporate bond market (an asset class likely to be adversely affected by regulatory tightening through disruption of ordinary market-making activities) and with particular attention paid to different underwriters, credit ratings, and issue sizes.

Our analysis is based on multiple estimation strategies, including standard breakpoint tests in levels, tests for structural breaks in dynamic factor models and difference-in-differences matching analysis. Reassuringly, the data display no statistical evidence of substantial deterioration in market liquidity after 2010. The tests presented are powerful enough to pick structural breaks in the data -they clearly pinpoint the crisis itself as a liquidity breakpoint- yet they consistently show no significant liquidity deterioration in the period of regulatory intervention covering the approval of the Dodd-Frank Act and Basel III, shutdowns of proprietary trading desks by major banks, or the proposal and finalization of the Volcker Rule. If anything, we detect evidence of liquidity improvement during periods of regulatory interventions, possibly due to the entry of non-banking participants. Evidence from the U.S. Treasuries market, by and large, confirms the

---

[71]Given the small number of time series available for the analysis of liquidity of Treasuries we do not employ dynamic factor model approaches in this Section. Online appendix Table 12, 13 and 14 report the estimated break dates, the double maximum test statistics and $\sup F_T (\ell+1|\ell)$ test statistics respectively.

absence of liquidity deterioration.

3.7 Tables and Figures

2014m4: Enhanced
2010m7: The Dodd-Frank   2012m7: Announced the final supplementary leverage ratio
Act signed into law   rules for Basel II.5   adopted

2011m6: Announced
proprietary trading desk
shutdown by Bank of   2014m4: Effective date of the
America   Volcker Rule

2009m6: The end of recession

2015m7: Deadline of
2010m9: Announced   compliance of the Volcker
2008m9: The bankruptcy of   proprietary trading desk   Rule
Lehman Brothers   shutdown by Goldman Sachs

2007   2008   2009   2010   2011   2012   2013   2014   2015

2010m9: Announced   2013m7: Announced the final
proprietary trading desk   rules for Basel III
shutdown by JP Morgan

2010m9: Basel III approved   2014m1: The revised final
by member jurisdictions   version of the Volcker rule
approved

2007m1: The start of
recession   2011m1: Announced
proprietary trading desk
shutdown by Morgan Stanley

Figure 3.1: Timeline of Crisis and Post-Crisis Regulatory Activities
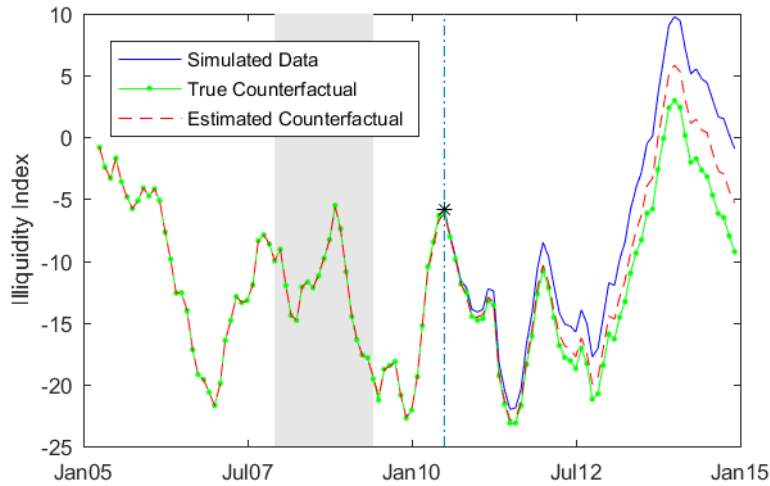
Figure 3.2: Simulated Illiquidity Index

Notes: This graph shows the average of 180 simulated liquidity measures over time. The blue solid line represents the average of 180 liquidity measures if regulation leads to a gradual deterioration in market liquidity, while the green dotted line represents counterfactual scenario where regulation has no effects. The dashed vertical line indicates the date of true and estimated structural break in the latent factor structure. The red dashed line is the estimated counterfactual path. The break date is estimated by Chen, Dolado, and Gonzalo (2014) and Bai and Perron (2003) approach with 5 percent significance level. The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.
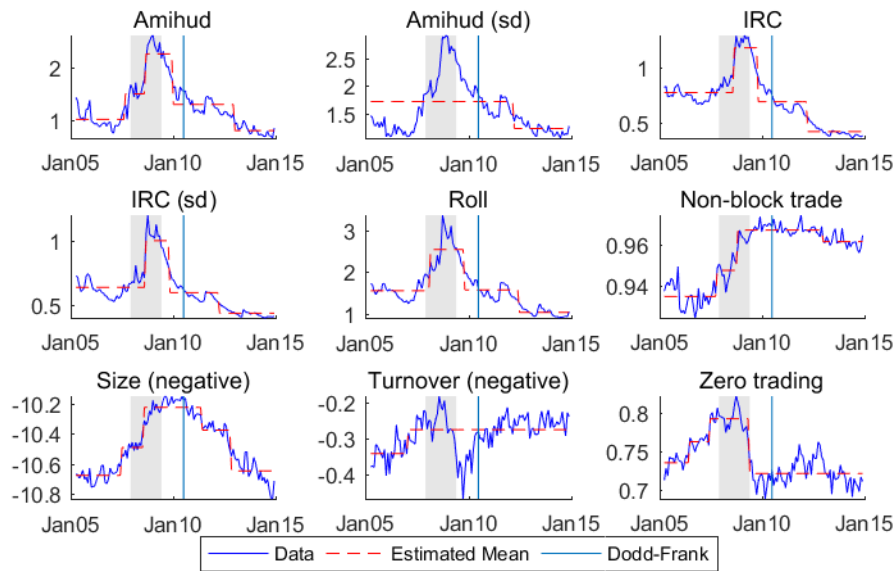
Figure 3.3: Time Series of Liquidity of U.S. Corporate Bonds (Aggregate-level)

Notes: This graph shows the time series of 9 aggregate-level liquidity measures of U.S. corporate bond market (blue line), and the estimated mean for each sub-period (red dashed line). The break dates (dates with a shift in the level of the red dashed line) are estimated by the Bai and Perron (1998-2003) approach with 5 percent significance level. The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.
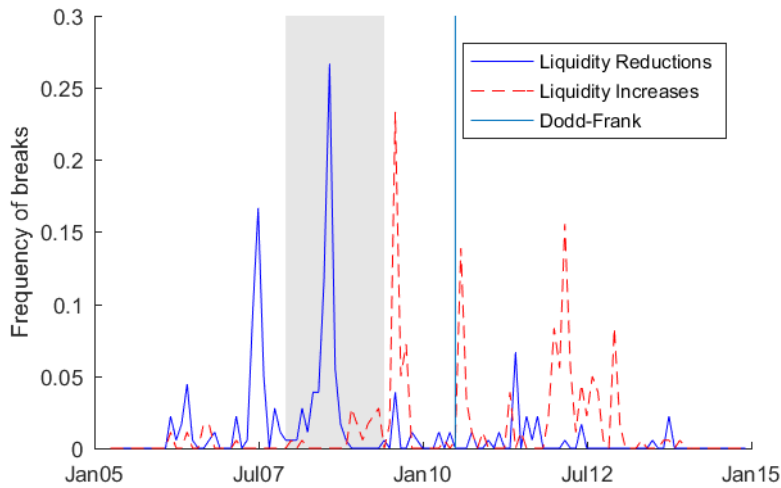
Figure 3.4: Breaks in the Means of Liquidity (Disaggregate-level)

Notes: This graph shows the frequency of break in the levels of 180 disaggregate-level liquidity measures for the U.S. corporate bond market over time. The x-axis shows the dates and the y-axis shows the corresponding fraction of the 180 liquidity measures which have a break at each date. The break dates are estimated using the Bai and Perron (1998-2003) approach with 5 percent significance level. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.
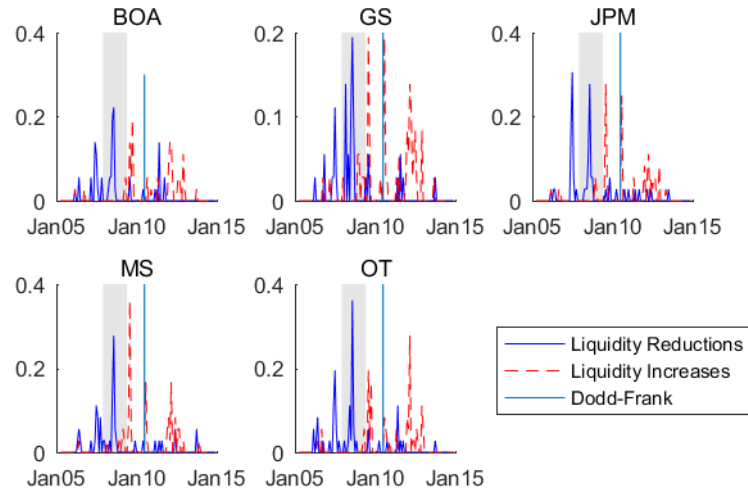
Figure 3.5: Breaks in the Means of Liquidity by Underwriter (Disaggregate-level)

Notes: This graph shows the decomposition of break dates by underwriter. The x-axis shows the dates and the y-axis shows the corresponding fraction of the 36 (=9×2×2) liquidity measures of each underwriter which have a break at each date. The break dates are estimated using the Bai and Perron (1998-2003) approach with 5 percent significance level. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The sample period is from April 2005 to December 2014. The data frequency is monthly.
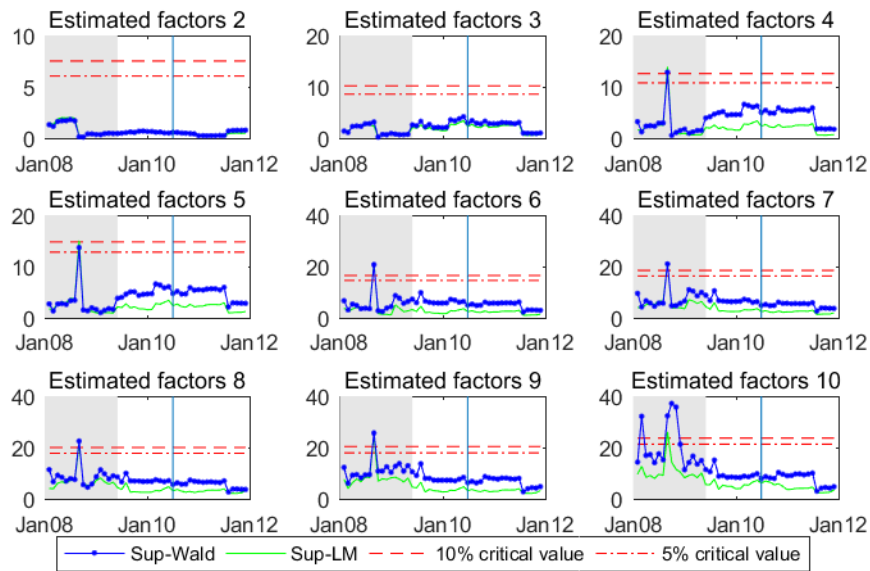
Figure 3.6: Breaks in the Means of Liquidity by Underwriter (Disaggregate-level)

Notes: This graph shows the test statistics of a single break in factor structure of 180 disaggregate-level liquidity measures employing the Chen, Dolado, and Gonzalo (2014) approach. The sample period is from April 2005 to December 2014. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The data frequency is monthly. The grey area indicates recession.
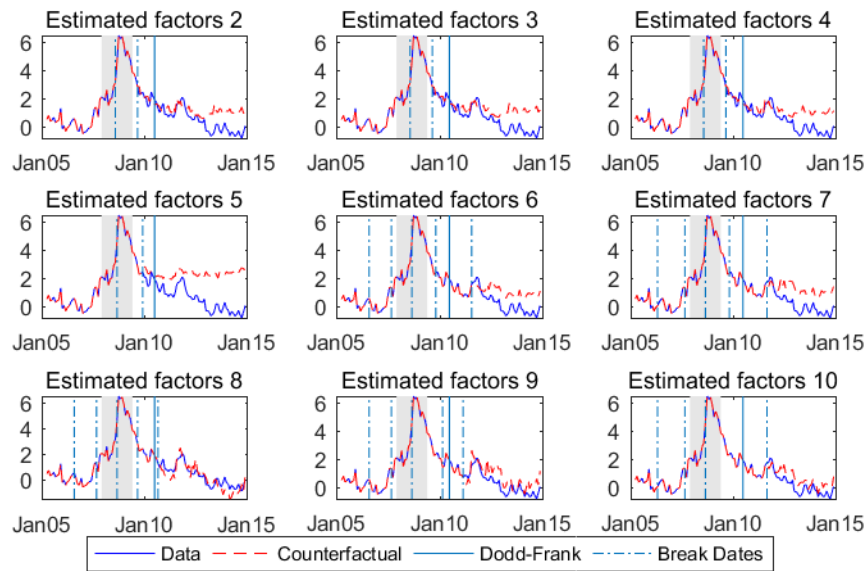
Figure 3.7: Liquidity Index of the U.S. Corporate Bond Market

Notes: This graph shows the average of 180 standardized liquidity measures of U.S. corporate bond market (blue solid line) and the estimated counterfactual path (red dashed line). The dashed vertical line indicates the dates of estimated structural breaks in the latent factor structure. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The break dates are estimated by Chen, Dolado, and Gonzalo (2014) and Bai and Perron (2003) approach with 5 percent significance level. The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.
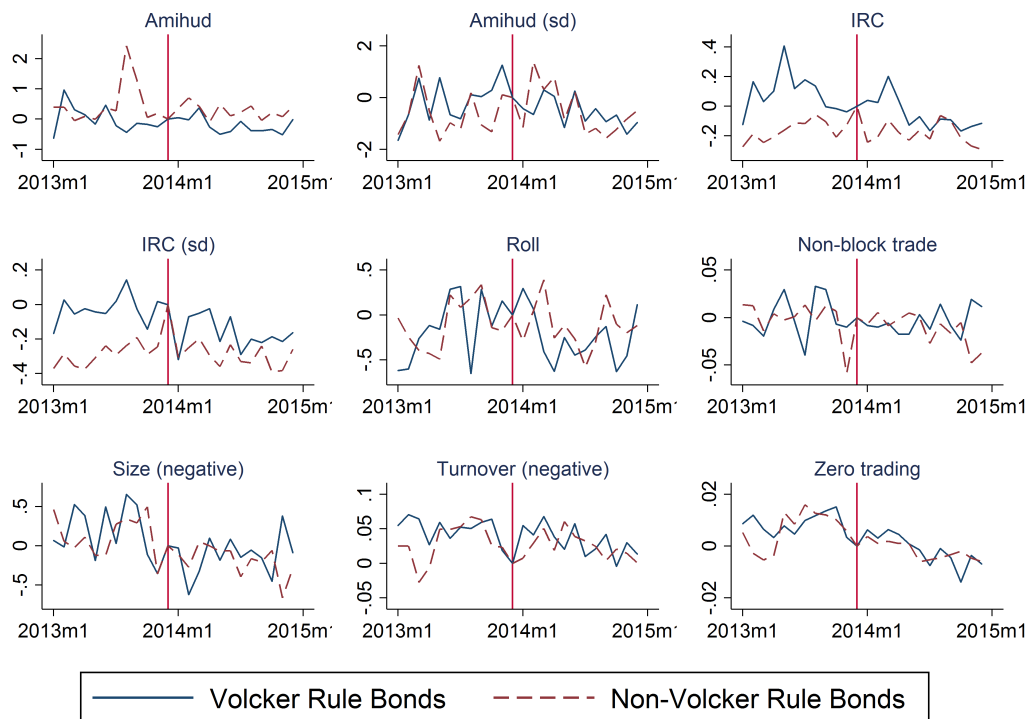
Figure 3.8: Liquidity of Volcker Rule and Non-Volcker Rule Bonds (Matched Sample)

Notes: This graph shows the time series of liquidity of Volcker Rule bonds and non-Volcker Rule bonds around the time when revised finalized version of the Volcker Rule was approved (January 2014). A non-Volcker Rule bond is defined as a bond which at least one of the underwriters is not subject to the Volcker Rule. A Volcker Rule bond is defined as a bond which all of the underwriters are subject to the Volcker Rule. Both time series are normalized to 0 in December 2013. The red vertical line indicates the date when the revised finalized version of the Volcker Rule was approved (2014m1). The sample period is from January 2013 to December 2014. The data frequency is monthly.
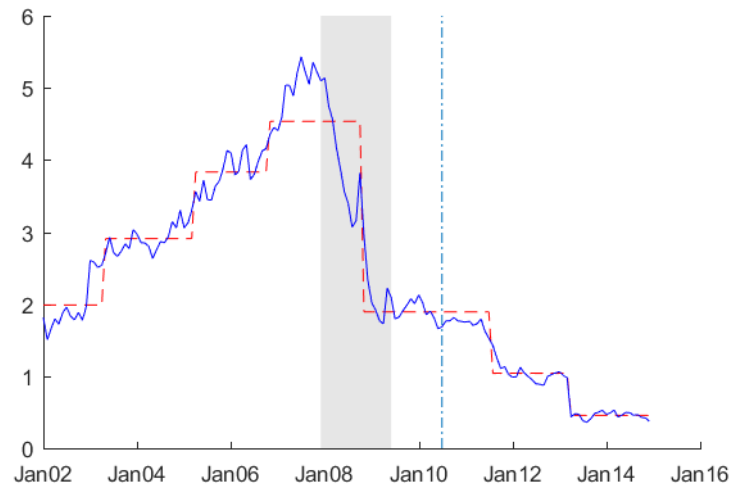
Figure 3.9: Primary Dealer Corporate Bond Holding

Notes: This graph shows the time series of the U.S. primary dealer corporate bond holding as the percentage of total corporate bond outstanding (blue line) and the estimated mean for each sub-period (red dashed line). The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The break dates (dates with a shift in the level of the red dashed line) are estimated by the Bai and Perron (1998-2003) approach with 5 percent significance level. The sample period is from January 2002 to December 2014. The data frequency is monthly. The grey area indicates recession.
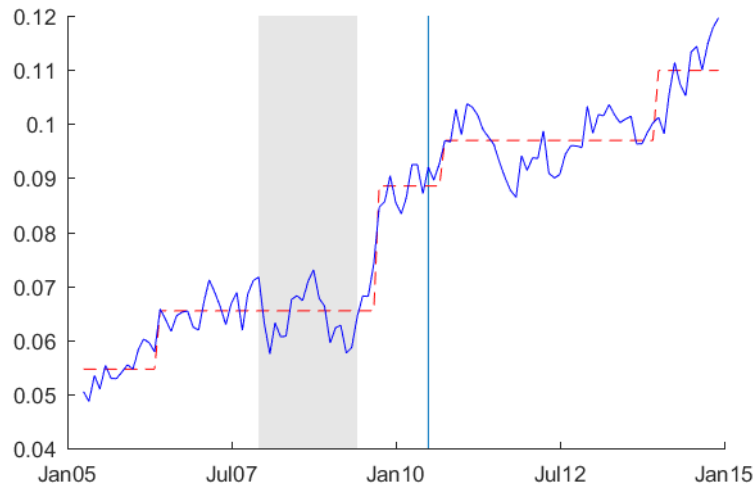
Figure 3.10: Fraction of Agency Transactions

Notes: This graph shows the fraction of agency transactions (blue line), and the estimated mean for each sub-period (red dashed line) over time. The break dates (dates with a shift in the level of the red dashed line) are estimated by the Bai and Perron (1998-2003) approach with 5 percent significance level. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.
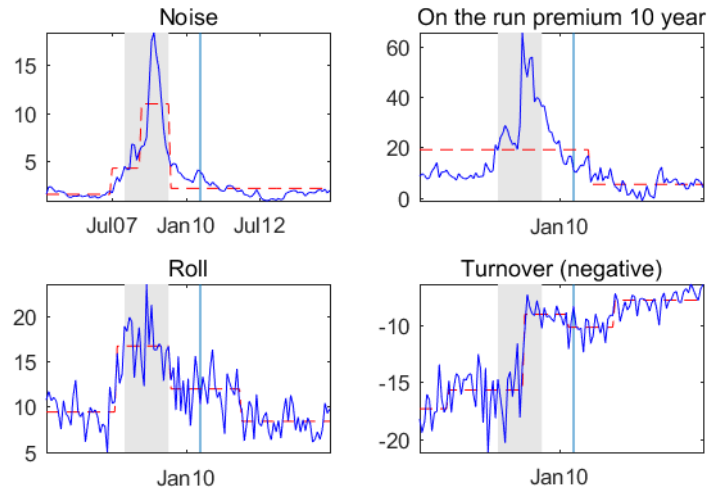
Figure 3.11: Time Series of Liquidity of the U.S. Treasury Bonds

Notes: This graph shows the time series of liquidity measures of U.S. Treasury market (blue line), and the estimated mean for each sub-period (red dashed line). The break dates (dates with a shift in the level of the red dashed line) are estimated by the Bai and Perron (1998-2003) approach with 5 percent significance level. The solid vertical line indicates the passage of Dodd-Frank Act (July, 2010). The sample period is from April 2005 to December 2014. The data frequency is monthly. The grey area indicates recession.

Table 3.1: Summary Statistics of the U.S. Corporate Bond Liquidity (Aggregate-level)

Measures	N	mean	sd	p10	p25	p50	p75	p90
Amihud	117	1.29	0.48	0.79	0.94	1.17	1.47	2.12
Amihud (sd)	117	1.57	0.48	1.10	1.19	1.43	1.83	2.35
IRC	117	0.70	0.24	0.40	0.49	0.70	0.82	1.08
IRC (sd)	117	0.61	0.18	0.42	0.47	0.58	0.67	0.88
Roll	117	1.59	0.54	0.96	1.23	1.52	1.81	2.41
Non-block trade	117	0.96	0.01	0.93	0.94	0.96	0.97	0.97
Size (negative)	117	-10.48	0.20	-10.71	-10.66	-10.51	-10.28	-10.18
Turnover (negative)	117	-0.29	0.05	-0.36	-0.32	-0.28	-0.25	-0.23
Zero-trading	117	0.74	0.03	0.71	0.72	0.74	0.76	0.79

Notes: This table shows the summary statistics of 9 aggregate-level liquidity measures for the U.S. corporate bond market. The sample period is from April 2005 to December 2014. The data frequency is monthly. The unit of Amihud, Amihud (sd), IRC, IRC (sd), and Roll is percentage point. The unit of Non-block trade, Turnover (negative) and Zero-trading is 1.

Table 3.2: Correlation Table of the U.S. Corporate Bond Liquidity (Aggregate Level)

	Amihud	Amihud (sd)	IRC	IRC (sd)	Roll	Non-block trade	Size (negative)	Turnover (negative)
Amihud (sd)	0.98							
IRC	0.88	0.84						
IRC (sd)	0.91	0.88	0.98					
Roll	0.93	0.93	0.96	0.97				
Non-block trade	0.29	0.33	-0.15	-0.06	0.01			
Size (negative)	0.75	0.76	0.51	0.51	0.58	0.65		
Turnover (negative)	-0.07	0.05	-0.28	-0.20	-0.10	0.27	0.00	
Zero-trading	0.40	0.43	0.54	0.52	0.56	-0.43	0.03	0.37

Notes: This table shows the correlations among 9 aggregate-level liquidity measures for the U.S. corporate bond market. The sample period is from April 2005 to December 2014. The data frequency is monthly.

Table 3.3: Difference-in-Difference Regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Amihud	Amihud (sd)	IRC	IRC (sd)	Roll	Non-block trade	Size (negative)	Turnover (negative)	Zero-trading
Volcker*Post	-0.231 [0.466]	-0.115 [0.359]	0.106* [0.0573]	0.113** [0.0474]	0.0177 [0.122]	0.00168 [0.00265]	0.0145 [0.0859]	-0.0160 [0.0118]	-0.00202 [0.00320]
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y
Bond F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	542	340	1807	1260	1992	2006	2006	11060	11060
Adjusted R-squared	0.458	0.398	0.238	0.325	0.245	0.312	0.524	0.353	0.879

Notes: This table shows the difference-in-difference regression of Volcker Rule bonds and non-Volcker Rule bonds around the time when revised finalized version of the Volcker Rule is approved (January 2014). A non-Volcker Rule bond is defined as a bond which at least one of the underwriters is not subject to the Volcker Rule. Each of the non-Volcker Rule bonds in our sample is matched to a Volcker Rule bond which issues in the same month, matures in the same month, has the same rating group (investment-grade/high-yield), and has a relative size difference less than 50 percent of the average size of the pair. The sample period is from January 2013 to December 2014. The data frequency is monthly. Control variables include the reciprocal of issue age, and the reciprocal of issue age squared. The standard errors are two-way clustered at the bond and month level. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 3.4: Summary Statistics of the U.S. Treasury Liquidity

Measure	N	mean	sd	p10	p25	p50	p75	p90
Noise	117	3.14	3.24	1.20	1.48	1.93	3.33	6.51
On the run premium	117	13.48	12.62	3.33	6.23	8.94	16.39	28.73
Roll	117	13.37	4.09	8.62	10.35	12.73	15.83	19.23
Turnover	117	-11.48	3.93	-17.64	-14.76	-9.79	-8.11	-7.39

Notes: This table shows the summary statistics of liquidity measures for the U.S. Treasury market. The sample period is from April 2005 to December 2014. The data frequency is monthly. The unit of Noise, On the run premium and Roll measure is basis point. The unit of Turnover (negative) is 1.

Table 3.5: Correlation Table of the U.S. Treasury Liquidity

	Noise	On the run premium	Roll
On the run premium	0.90		
Roll	0.62	0.72	
Turnover	0.03	-0.08	-0.37

Notes: This table shows the correlations between liquidity measures for the U.S. Treasury market. The sample period is from April 2005 to December 2014. The data frequency is monthly.

Chapter 4

Factions in Nondemocracies: Theory and Evidence from the Chinese Communist Party

4.1 Introduction

This paper presents a theoretical and empirical analysis of the internal organization of China's political linch-pin: the Chinese Communist Party (CCP). As the regime party of the People's Republic of China (PRC), the CCP is de jure and de facto the be-all and end-all of political activity in the second largest economy and the most populous country in the world today[72]. This motivates the interest of political economists in the CCP.

The nontransparent and often informal nature of elite interaction within a country lacking competitive elections and with a rich history of informal political jousting among factional leaders raises formidable obstacles to a rigorous politico-economic analysis. The economic literature on the internal organization (and, we will see, factional competition) at the highest levels of the Chinese government is limited[73]. Political scientists focused on China studies have been more attentive, but also often more qualitative and descriptive, at least until recently[74].

The CCP remains today "*a secretive, selective organization of about* 65 *million members who have positions of influence in all sectors of Chinese society...*" (Nathan and Gilley, 2003 p.7)[75]. Operations of the Politburo and the highest echelons of the CCP have been often described as opaque at best (Pye, 1980; Dittmer, 1995; Shih, 2008). As reported in Nathan (2016): "*Deng built a system of tacit norms by which senior leaders were limited to two terms in office, members of the Politburo Standing Committee divided leadership roles among themselves, and the senior leader made decisions in consultation with other leaders and retired elders.*"

Within this context, intra-elite competition is extremely hard to assess. The CCP officially rejects factional elite politics[76], but scholars since Nathan (1973) have emphasized how the faction –intended as patron-client clusters of mutually linked officials– represents the correct unit of analysis of elite politics in China. Since Nathan (1973), evidence supporting this interpretation has also steadily accumulated (Pye, 1981; Dittmer and Wu, 1995; Nathan and Gilley, 2003; Shih, 2004; Li, 2012; Li, 2013; Shih, 2016; Meyer, Shih, and Lee 2016). The present paper follows this line of inquiry, but with special attention paid to in-

---

[72]And plays a crucial role in steering economic activity in the country. See Bai, Hsieh, and Song (2016).

[73]The study of the political economy of China has several important exceptions, but often not precisely focused on national elite competition. Persico, Pueblita, and Silverman (2011) in their analysis of factional politics focus on the CCP, among their various case studies. Less relatedly, work such as Li and Zhou (2005) focuses on the promotion profiles of provincial leaders and so does Jia, Kudamatsu, and Seim (2015). Work by Lau, Qian, and Roland (2000) models the process of reform under Deng Xiaoping and the reform era.

[74]Descriptive discussion most pertinent to this paper includes Li (2012, 2013). Several quantitative exceptions are discussed in Shih (2016) with respect to scholarship in East Asian studies and political science, while less recent examples include Huang (2000), Shih (2004, 2007); Shih, Adolph and Liu (2012).

[75]At the time of writing. By 2016 the CCP membership has grown to 88.76 million.

[76]BBC, Monday January 5, 2015: "An editorial in Monday's flagship newspaper, The People's Daily, says cliques are akin to parasites and are 'harmful for both the country and the people.'" http://www.bbc.com/news/blogs-china-blog-30685782

dividual incentives, supplying an inherently economic model of behavior, where "*lower-level officials [...]
join factions in order to secure promotions and other regime goods from powerful patrons*" (Shih, 2016,
p.1) and where promotion dynamics throughout the party hierarchy are microfounded and characterized. A
theoretical contribution of this paper is in the formal model of factional interaction that we present.

In our model factions operate within a given party hierarchy. On the one hand, the advantage of factions
is that they provide support to their members in obtaining promotions up the pyramid. On the other, factions
allow the allocation of that support to be decided by senior affiliates, with the possibility of junior members
being blocked by higher ranked cofactionals keen to avoid promoting colleagues who will compete with
them for future openings. A faction member, though potentially benefiting from cofactional support, has
to bide his time and wait for the seniors in his faction to allow that support to materialize. The seniors
make this decision based on their own career objectives, so that a junior member's ascendancy through the
hierarchy is tethered to the rise of the relevant seniors above him. Unaffiliated (neutral) politicians face
no such restrictions, and this is why neutrals can also emerge in equilibrium. Though they do not enjoy
factional support, they are also not restricted in their capacity to contest openings higher up. The analysis
of the costs and benefits of joining factions is complicated by the dependence of promotion opportunities on
the factional composition of every level of the hierarchy at any point in time. This determines what kind of
openings may arise and who is in a position to block advancement at any level, a problem that we study in
detail.

Our theoretical results are important in matching empirical moments in terms of factional composition,
promotion rates, and the effects of changes in the factional identity of the top leadership in China. Absent
hard and verifiable information, we rely on the extant discussion of Chinese elite politics to identify a
minimal set of factions within the CCP. Factions have historically emerged within the CCP through close
personal connections with prominent patrons (e.g. in the cases of former General Secretary Jiang Zemin and
his successor, Hu Jintao) to mutually foster the career prospects of affiliated cadres, and do not necessarily
represent specific territorial or economic interest groups (Dittmer, 1995). As we discuss in Sections 4.2
and 4.3, this paper will lever only the most obvious factional links identified within the CCP, links based
on affiliation to the Communist Youth League of China (related to General Secretary Hu Jintao) or to the
so-called Shanghai Gang (affiliated most prominently with Jiang Zemin and bolstered by the special status
of Shanghai in Chinese politics).

Scholars such as Shih, Shan, and Liu (2010), Shih, Adolph, and Liu (2012), Jia et al. (2015) have
explored methodologies for the imputation of factional linkages based on place of birth, university ties, and
shared career profiles[77]. While we also focus on systematic biographical information, we remain wary of
potential mismeasurement in the identification of factional ties, as is likely for factional affiliation based
purely on place of birth or shared career paths. An important reason for this wariness will be evident in
our statistical analysis. Based on our factional definitions and within a proximate set of party officials of
almost equivalent rank in the same office and area (e.g. the number 1 and number 2 highest ranked party
members in a province), we show that members of a faction (let us call it *B*) are virtually never paired with
members of the same faction *B* at the same office. On the contrary, they are paired with members of a rival
faction (*R*) in excess to what would be predicted by random chance alone. For instance, if a province has
a *B* faction Party Secretary (ranked number 1), the Governor (his number 2) is likely to be an *R*, possibly a
neutral official, but most definitely not a *B* faction member. Thus simply sharing part of their career paths
may not be informative of factional affiliation for CCP elite officials. In fact, our evidence shows it may
mislead completely.

The statistical analysis of these systematic factional cross-patterns in top CCP positions is new to the
literature and will be discussed in Section 4.4. In addition to studying these cross-factional patterns, Section

---

[77]Shih (2008, p.66) discusses issues of measurement with the premise that "*Despite the centrality of factions in Chinese politics,
they are extremely difficult to observe in a systematic manner, especially in such an opaque political system.*"

4.4 reports statistically significant premia in terms of promotion rates and seat allocations to a leader's cofactionals. That factions may deliver advantages to their members is a necessary condition for our model's coherence. But the existence of precisely estimated leadership premia points also in the direction of factions both being reasonably identified within our analysis and of operative relevance within the CCP.

We formally explore and test for the presence of additional factions. This is possible within our setting thanks to the structural econometric approach we follow. We directly bring our model to the data, obtain estimates of the primitive parameters (such as leadership premia and parameters governing the contest functions for promotion) and formally test our mechanism against alternatives, including mechanisms based on pure seniority or meritocracy. Our factional model displays excellent in-sample and out-of-sample fit. We show how the estimated leadership premia in the CCP are quantitatively substantial, but quite far from winner-take-all levels, and that the intra-faction competition among faction members operates as a de facto endogenous dampening mechanism in slowing factional growth.

Our analysis includes several counterfactuals. We model possible institutional changes within the CCP, including the effect of increased leadership premia, which may indicate a break away from the *"collective leadership"* design envisioned by Deng. We also study the role of the identity of the top leadership, the factional role of princelings, and we try to explicitly assess General Secretary Xi Jinping's factional affiliation.

Besides the politico-economic literature on Chinese elite politics mentioned above, this paper speaks to the literature on the internal organization of autocratic regimes. Francois, Rainer, and Trebbi (2015, 2016) discuss at length the importance of its connection to the expanding literature on the political economy of development. Most related to our work (and one of the first rigorous analyses of factional politics within the economic literature) is Persico, Rodriguez-Pueblita, and Silverman (2011), who present a theoretical model of endogenous factional growth and link it qualitatively to evidence from factional local politics in Mexico within the Institutional Revolutionary Party.[78]

From a theoretical perspective, Dewan and Squintani (2015) model endogenous faction formation (an issue we address in our setting as well, when characterizing the decision of party members to join a faction). The authors develop a model where incentives for faction formation are ideological rather than economic (as in our setting and in Persico, Rodriguez-Pueblita and Silverman, 2011) and show how within their framework factions may serve welfare-enhancing purposes, limiting extremists within the party by tying them to moderate faction leaders. Factions are also shown to facilitate information sharing and party effectiveness in their model.

The remainder of this paper is organized as follows. In Section 4.2 we provide a brief institutional overview of the CCP. In Section 4.3 we discuss our data, operationalize factions, and provide a descriptive analysis of our samples. Section 4.4 produces a set of stylized facts, some novel, useful to frame and guide the theoretical analysis. In Section 4.5 we discuss our theoretical setup and Section 4.6 develops our estimator. Our main empirical results are reported in Section 4.7. Section 4.8 presents our counterfactual exercises. Section 4.9 concludes.

4.2 Institutional Background: the CCP

This section presents a brief institutional overview of the internal organization of the CCP in the reform era. It is in no way exhaustive, but only of assistance to the reader unfamiliar with Chinese politics in framing

---

[78]See also Belloni and Beller (1978). Persico et al. (2011) also point out to the relevance of factional politics well beyond Mexico's camarillas or the CCP, with references to studies of factionalism within the Japanese legislature (Cox et al., 1999, 2000) and the Italian parliament (Zuckerman, 1975; Kato and Mershon, 2006; Ceron, 2015; and Laver and Giannetti 2004). Factions in Australian politics are discussed in McAllister (1991). The US urban party machine factional structure, such as in the case of Tammany Hall, are subject of an entire and even earlier literature. See Myers (1917).

the analysis that follows[79].

In 2016 the Chinese Communist Party, with its 88.8 million members, is one of the largest political parties worldwide and one of the most enduring (founded in 1921). The CCP organization is strongly hierarchical in nature and the party reflects one-to-one the organization of the Chinese state, as typical in the architecture of Leninist regimes.

The top of the CCP hierarchy is shared by the figures of the General Secretary of the CCP and the second ranked member of the CCP, which respectively assume the roles of President and Premier of the State Council of the PRC. Both leaders belong in turn to the Politburo Standing Committee (PBSC), formed by the other 5 members and which represents the set of the highest ranked politicians in China. The PBSC is an expression of the 25-member Politburo (PB), the executive body of the Central Committee of the Chinese Communist Party. The Central Committee (CC) is de jure the highest political body in the CCP and currently consists of 205 full members and a set of 171 Alternate Central Committee (AC) members in junior standing relative to the full members (and without voting rights). All members of the CC and AC are ranked hierarchically. The CC and AC are elected during National Congresses of the CCP and the interim plenary sessions fill retirements or deaths, granting promotions (and occasionally administer demotions). Typically, CC members include ministerial-level officials and provincial ranking officials, including Provincial Party Secretaries (the highest CCP post in a Province) and Governor (the second ranked). It is important to notice that Provinces tend to display a political architecture that mimics the national government and the national party structure. Provincial leaders operate in the context of local party committees and local party congresses are held typically every five years. The CCP maintains a pyramidal structure, branching all the way down to the village level and the Village Party Branch Secretary.

While not all layers of the Chinese political hierarchy present nodes mapping into a diarchic structure, most do, typically separating party roles and administrative roles. Examples of diarchic arrangements include the presidency and premiership as the two highest ranking members of the Politburo Standing Committee; the PRC Presidency (President and Vice President); the State Council (Premier and Executive Vice Premier); and the top dyads at the provincial level (Provincial Party Secretary and Governor)[80]. We will occasionally refer to such pairs of positions as position 1 and 2.

The opportunity of entering the ranks of the CCP is closely guarded and party membership typically guarantees access and career opportunities beyond those available to common citizens[81]. For this reason, an elaborate recruitment process typically operates through the selection of successful university students and through family and work connections.

Membership of the Communist Youth League of China (CYLC), an ancillary organization to the CCP responsible for the youth (members are typically between 4 and 28 years of age), has traditionally operated as an entry point in the CCP. As discussed in Li (2012, 2013), individuals with a background in the CYLC are often referred to as members of the *tuanpai* (i.e. Youth League [faction]) and tend to originate, although by no means exclusively, from the less prosperous ("red") regions[82]. Li (2012) associates with the CYLC "populist" policies close to the rural poor and recent migrants to cities, as opposed to the policies preferred by more "elitist" groups comprised by CCP cadres close to former General Secretary Jiang Zemin and a group of party officials connected to the Shanghai municipal administration. Indeed, the economic and

---

[79]See also Chapter 1 in Nathan and Gilley (2003) for a less brief overview. For a comprehensive discussion of elite politics in China see references in Shih (2016).

[80]See Li (2014) for a discussion and examples. Other instances include the CMC (chairman and executive vice chairman), the CCP Secretariat, the NPC and CPPCC (chairman and executive vice chariman), the Supreme People's Court. Assuming the presence of such dyads across the whole hierarchy should be simply read as allowing for the presence of a close substitute in the party hierarchy for any member.

[81]The Organization Department of the CCP Central Committee on June 30th, 2016 in an official release indicated that 22 million Chinese residents had applied in 2015 and less than 4.5% of the applications were accepted. http://news.xinhuanet.com/english/2016-06/30/c_135478976.htm

[82]Prominent members include current Premier Li Keqiang and former General Secretary and President of the PRC Hu Jintao.

political role of Shanghai cannot be emphasized enough in CCP internal interactions, to the point that the term *Shanghai Bang* (Gang) has been often employed to identify the patronage cluster close to Jiang and to the economic interests of the coastal (blue) provinces (Li, 2002).

Whether additional factional groups besides the CYLC and the Shanghai Gang may be present within the CCP is unclear and disputed even among scholars of Chinese elite politics. For instance, some observers point at the anomaly of the exceptionally rapid careers of sons and daughters of prominent party officials and revolutionary veterans under Mao, often referred to as "princelings". The analysis below will discuss this specific group of CCP members in detail.

4.3   Data

We combine two biographical databases of Chinese politicians. The first data source is China Vitae, which collects biographical information on more than $4,494$ Chinese elites in government, politics, the military, education, business, and the media since 1992. Information provided by China Vitae includes gender, year of birth, place of birth, ethnicity, colleges attended, and career trajectory. Information in China Vitae comes from Chinese and English language web sites in China that are supported by or affiliated with the Chinese government.

Our second data source is a biographical database of CC members developed by Shih, Shan, and Liu (2008), and further updated by Lu and Ma (2015). This database contains all CC and AC members from the first Party Congress in 1921 to the eighteenth Party Congress in 2012. This data also provides biographical information and career trajectories similar to China Vitae. We focus our analysis on the period of 1956 to 2014, which starts from the first Party Congress since the founding of People's Republic of China (8th Party Congress in 1956) and ends with the most recent Central Committee (18th Party Congress in 2012), covering a total number of $1,853$ individuals.

We combine these two data sources to construct our estimation samples. Whenever there is inconsistency between the two data sources, (e.g. multiple politicians in the same position in the same year), we manually check with a third source, typically official websites affiliated with the Chinese government (e.g. www.xinhuanet.com; cpc.people.com.cn). We also collect provincial population and GDP data from China Data Online. The anti-corruption data originates from ChinaFile and China's Central Commission for Discipline Inspection (CCDI) website.

Following the literature on Chinese politics (Bo, 2008; Li, 2013a; Li, 2013b), we construct four affiliation indicators for the full sample of politicians: CYLC, Shanghai Gang, but also Military and Princeling status. A politician is classified as from the CYLC if he/she has held provincial and national level positions in CYLC. A politician is classified as from the Shanghai Gang if he/she has held official positions in the Shanghai municipal party apparatus, municipal government, municipal People's Congress, and municipal People's Political Consultative Conference. This again underlies the exceptionality of the Shanghai political machine. A politician is classified as from the Military if he/she served as military personnel in the Revolutionary Era (1921-1949), or has participated in the volunteer armies to Korea or Vietnam, or served as military personnel for more than half of its career after the founding of People's Republic of China. The restriction on the minimum time of military experience is to rule out civilian officials who work as the party secretary of a military region for a short period of time (e.g. Hu Jintao as the First Secretary of Guizhou Military District from 1985 to 1988), or civilian officials chair the Central Military Commission (e.g. Jiang Zemin as the chairman of the Central Military Commission from 1990 to 2005). A politician is classified as a Princeling if he/she is from a prominent political family, the so called "red aristocracy" (prominent examples include General Secretary Xi Jinping and disgraced former governor of Liaoning Bo Xilai). These four affiliations are not mutually exclusive (for example, Xi Jinping is both a princeling and an affiliated of the Shanghai Gang according to our definition) and not all party members in our sample are affiliated. In fact,

we allow for politicians in our sample to also be unaffiliated (neutral, indicated as *N*).

Theoretically one could consider CYLC, Shanghai Gang, military, and princelings alternative political factions. In Section 4.4 we show however than only two of these groups, CYLC and Shanghai Gang, truly exhibit the features of political factions within the CCP. Formal statistical tests will be also developed and brought in support of this thesis. To distinguish, we will refer to princelings and military as "groups" and CYLC and Shanghai Gang as "factions".

The military is virtually a parallel structure with limited political control, while the princelings as a group are extremely heterogeneous and appear to operate as a set of neutral and independently powerful actors (in fact, often times in deep rivalry among themselves, such is the case of Bo Xilai and Xi Jinping). While we will keep track of all types of affiliations in the analysis that follows, we emphasize here that our theoretical and empirical design will separate CYLC and Shanghai Gang faction members from all other political actors, including the military and princelings, which we will deem "neutral". Because of the traditional coloring associated with these two established factions, we will also occasionally refer to the CYLC as the Red faction, *R*, and to the Shanghai Gang as the Blue faction, *B*.

Table 4.1 provides summary statistics of demographics and careers of 4, 494 politicians who held important positions in government, politics, the military, education, business, and the media in China since 1992. The unit of observation is a position-individual pair. We classify the organizations into 12 categories: party apparatus, government, military, People's Congress, Chinese People's Political Consultative Conference (CPPCC), court, procuratorate, CYLC, business, media, education, and an unclassified category. The average duration of each position is about 4 years, and the age of starting each position varies from the early 30s (CYLC) to the late 50s (People's Congress). Individuals who hold these positions are predominately male, which reflects the large gender imbalance at the top levels of government and business in China[83]. Ethnicity is predominately Han, reflective of the ethnic composition in the Chinese population. The last four columns provide the frequency of the various affiliations in each type of organization. CYLC members tend to work in the party apparatus and media instead of the government system.[84] The Shanghai Gang is more evenly distributed across all types of organizations. Princelings are more likely to have experience in the military, but are less likely to work in the legal system (court and procuratorate), potentially due to the fact that the power of the judiciary is relatively muted in China.

We then turn our focus to a subset of elites, the members of Central Committees of the CCP. This is a group of around 400 people who comprise the CCP top leaders. Table 4.2 provides the demographics and the factional affiliation by sessions of the Central Committees. Similarly to the larger sample of elites, the CC members are predominantly male, in their mid-50s and mostly Han. Over the past 60 years, more members hold college or even post-graduate degrees. However, only 10 percent of them studied or worked abroad. More than 10 percent of them have worked as personal secretaries (*Mishu*) of prominent politicians, illustrating the importance of personal ties in Chinese politics. Conditioning on entering the Central Committee, around 20 percent of them are promoted into higher level in the four levels of the Central Committee, and around 50 percent will retire in next CC session. In terms of factional affiliation, CYLC, Shanghai Gang, and princelings each account for around 5 percent to 10 percent of members. The military has experienced a large downward trend, dropping from 56 percent in the 8th Central Committee to less than 20 percent in recent years.

4.4 CCP Factional Politics: Reduced Form Results

This section presents a set of facts on factional politics in China, the most important of which are novel, to

---

[83]25.1% of CCP members were women in 2016.

[84]This is consistent with the anecdotal discussion of Hoffmann and Enright (2008) that CYLC leaders often have experience in non-economic fields, such as party organization and propaganda

the best of our knowledge. These stylized facts are going to inform and motivate the theoretical analysis that follows.

i) National Political Actors. We begin by arguing qualitatively that the factional affiliations we posit (CYLC and Shanghai Gang) share properties that make them bona fide large national players within the CCP and are not merely political actors representing local constituencies.

In Figure 4.1 and Table 4.3 we describe the geographic distribution of members affiliated with the CYLC and the Shanghai Gang in provincial roles. As is evident, the representation across provinces is fairly broad and not limited to a particular local area, despite a small positive correlation between the presence of Shanghai Gang and the average GDP per capita of the province. On the other hand, individuals associated with princelings and the military group are distributed more unevenly: princelings are more likely to hold positions in rich costal areas – possibly due to their privileged status — while military members are more concentrated in poorer western provinces and places with strategic importance (e.g. Fujian, which neighbors Taiwan).

ii) Cross-Factional Mix. Useful to the understanding of factional dynamics within the CCP is the study of the peculiar factional mix which we observe when sampling the diarchic nodes pervading Chinese institutional design. These are pairs of positions of similar rank and operating in close institutional proximity to each other. Table 4.4 reports formal statistical tests of the factional composition of virtually all top two leadership posts in post-Deng China. In particular, we ask: given the factional affiliation of a politician sitting in one of the top two leadership positions of a national or provincial organ, what is the likelihood that the other position will be held by a cofactional member? It turns out it is extremely low.

Table 4.4 shows panel regressions of the factional affiliation of the number 1 official on the number 2 official's affiliation at the same node. The variables $CYLC1$ and $Shanghai1$ (respectively, $CYLC2$ and $Shanghai2$) are dummies which equal 1 if the number 1 official (respectively, number 2) is from that faction and 0 otherwise. We will also refer to such factions through the abbreviations R, B. The sample period is from 1992 to 2014. Columns 1-4 include all positions, and Columns 5-6 break down to provincial and national level positions. The provincial positions include 31 provincial and municipal units (secretary and governor)[85]. The national positions include the Politburo Standing Committee (two highest ranking members), PRC presidency (President and Vice President), the State Council (Premier and Executive Vice Premier), Central Military Committee (Chairman and Executive Vice Chairman), CCP Secretariat (two highest ranking secretaries), NPC (Chairman and Executive Vice Chairman), CPPCC (Chairman and Executive Vice Chairman), the Supreme People's Court (President and Executive Vice President).

Taking the top two leadership positions in any CCP (or PRC) organ, position 2 being filled by a R (respectively, a B) politician predicts negatively and significantly the likelihood of position 1 being filled by an R (respectively, a B) politician. The estimated negative coefficients indicate a statistically robust lower likelihood of same-faction pairs (R, R) or (B, B) relative to what would happen in case of pairings forming randomly between B, R, N. Interestingly, the evidence for princelings is much weaker, in line with further evidence below showing their lack of behavior as an organized faction. In Table 4.5 we further show that there is also a statistically precise excess likelihood of matching pairs in the form (R, B) and (B, R) relative to possible pairings with neutrals, N.

The presence of cross-factional pairs exceeds significantly what would emerge by random chance alone. To the best of our knowledge these facts on systematic cross-matching within Chinese elite politics are new. An implication of this evidence is that methodologies imputing factional affiliation based solely on shared professional paths may be highly deceptive, as discussed in the Introduction.

iii) Leadership Premia. A crucial feature of any theoretical model of factional politics is the ability of factions to deliver resources to their members. This seems a necessary condition that our factional definition should satisfy, a conceptual underpinning that we must be able to verify in the CCP data in order to justify

---

[85] Shanghai Municipality is excluded in the regression sample of Shanghai Gang.

our approach.

We will do this in what is possibly the starkest way: estimating premia in factional seat assignment and promotion rates of cofactionals of the country leader (i.e. the PRC President and General Secretary of the CCP). Again, we are not aware of any systematic analysis of this type for the CYLC and Shanghai Gang.

Table 4.6 shows a panel regression of promotion and retirement dummies on the factional affiliation of Central Committee members interacted with the faction of the General Secretary. The sample includes all members of the 8th to the 18th Central Committees (Politburo Standing Committee members are excluded from the promotion regression). Promotion is equal to 1 if a Central Committee member moves up in the rank defined by the four levels of Central Committee (1 PBSC, 2 PB, 3 CC, and 4 AC).

As is clear from the reduced form regressions, an *R* (respectively, a *B*) politician has substantially higher likelihood of promotion when an *R* (respectively, a *B*) leader is in power. On average CYLC and Shanghai Gang members exhibit promotion rates higher by 10 percentage points relative to neutral members (excluding military and princelings), as reported in Appendix Table 3. However, this result masks substantial heterogeneity. While CYLC and Shanghai Gang members' promotion rates hover around 4 percentage points higher than neutrals in times where the leadership is not from an individual faction, having a co-factional leader adds 20.6 percentage points to CYLC and 19.3 to Shanghai Gang, inducing a substantial, highly significant, leadership premium in the speed at which leader's cofactionals are promoted. Figure 4.2 provides a vivid visualization of the leadership premia in promotion rates.

We also perform an analysis looking at allocations of crucial posts to factional members. The dependent variables include: the share of official positions allocated to a faction constructed following the scheme of Bo (2010) and weighted by value (we will refer to it as "power score"); the share of seats of Alternate Central Committee members (AC); of the full Central Committee (CC); of the Politburo members (PB); and of the Politburo Standing Committee members (PBSC). These effects are reported in Table 4.7. Leadership premia are statistically significant, between 4 percentage points higher in terms of power score shares for the CYLC and around 2 percentage points for the Shanghai Gang. These estimates are not trivial, but quite far from winner-take-all levels. The leadership premia in the power score can be easily observed in the simple time series plots of Figure 4.3.

iv) Anti-Corruption Campaign. As in the allocation of rewards to cofactionals through leadership premia, we would also expect evidence of factional bias in the administration of punishment. We have limited systematic evidence in this respect, but it interestingly points in a direction consistent with the limited leadership premia discussed at point iii).

This novel evidence comes from the factional analysis of the CCP members hit by President Xi Jinping's anti-corruption campaign (initiated in 2012 and still ongoing as of 2016). A remarkable factional balance seems to be present in the administration of punishment, when looking at the detailed resumes of the so-called "tigers", a code name for high-ranking party members affected by the purge[86]. Table 4.8 shows that both CYLC[87] and Shanghai Gang cadres appear represented in the purged sample[88] and, importantly, both factions are represented in shares proportional to their overall representation in the upper echelons of the CCP, and not statistically significantly higher or lower. The reader may however notice a lower, but not significant, representation of Shanghai Gang members, the faction most likely to be associated with Xi (if

---

[86]As opposed to low-level politicians, "flies", involved in petty corruption. Tigers directly hit by the anti-corruption purge have included retired PBSC member Zhou Yongkang and retired PB member Xu Caiou.

[87]Links to the CYLC were evident in official news releases by The People's Daily which explicitly singled out specific subsets of this faction, particularly "The Shanxi Gang", officials linked to Ling Jihua, a disgraced protegé of Hu Jintao. http://www.bbc.com/news/blogs-china-blog-30685782

[88]We build a corruption dummy indicator for whether a political/military official is listed in the public anticorruption database of the Central Commission for Discipline Inspection and from ChinaFile. Table 4.8 shows the cross-section regression of corruption dummy on faction affiliation of an official. The sample includes all the individuals covered by China Vitae who have not retired in the year of 2007, the year of 17th party Congress. We dropped military personnel from the sample as the coverage of this group is relatively limited in China Vitae.

at all –see Section 4.8). These results appear also completely consistent with an independent analysis of the anti-corruption campaign presented in Lu and Lorentzen (2016).

v) Post-Deng era. Finally, we provide brief empirical justification for our focus on the post-Deng era. Mao Zedong and Deng Xiaoping have been often characterized as political "strong men" by many observers, as their legendary careers in the revolutionary era won them ultimate control over the military. In contrast, subsequent leaders, Jiang Zemin, Hu Jintao, and Xi Jinping, appear categorically different: civilian officers who rose through the party hierarchy relying on their ability and connections. This structural break is evident in the data.

Underlying the symbolic retirement of Deng in 1989, we document structural changes in the whole spectrum of political elites. Figure 4.4 shows the share of power score by factions or groups in the Central Committees of the CCP. Post-Deng China witnesses a significant decline in the influence of the military group, and a rise in factions such as CYLC and Shanghai Gang. Figure 4.5 breaks down the power score by four constituencies of the Central Committee: state organs, party apparatus, military, and regional governments. The pre-Deng era was ridden with volatile shifts across constituencies, with the most salient example being the Cultural Revolution between 1966 and 1976, during which state organs and party apparatus were virtually paralyzed. In contrast, the post-Deng era witnessed the stabilization of power shares for each constituency. Despite the lack of political reform often alleged by outside observers, the above evidence suggests that Chinese politics evolved to a new phase in which political strongmen became replaced by factional politics after Deng[89]. This is the period we focus on.

4.5 Model

Having produced a series of statistical regularities pointing in the direction of a systematic role for factional affiliation in the organization of the CCP (and the Chinese state more in general), we now proceed with the construction of a formal theory useful to understanding the incentive structure driving the data in the post-Deng era.

4.5.1 The Hierarchy of Positions

There is a L level hierarchy of leadership positions, ordered from the highest level 1, to the bottom, L. Each level, ℓ, of the hierarchy has a $M(\ell)/2$ leadership nodes. Each leadership node has a pair of leadership positions. The two positions at each node are ordered (position 1 and position 2). The hierarchy is broken up into regions, each of which nests a higher number of smaller regions below it. Level 1, the top level, has one node and hence two positions; $M(1) = 2$. It is the paramount leadership node for the country as a whole (currently, President Xi Jinping and Premier Li Keqiang). Level 2, the second layer in the hierarchy, has $M(2) > M(1)$ positions divided up into $M(2)/2$, and so on, with the number of positions strictly increasing down to level L. The nodes at the lowest level are the "entry" leadership positions, corresponding to the first step in a political life that we model.

Time is continuous. Each individual politician "dies" (or exogenously retires) with an instantaneous probability, δ, which also acts as the instantaneous discount rate. Upon a politician's demise, his or her position opens up for replacement. A politician's position also opens up when promoted to a position above, freeing the current spot. We assume that the flow utility from being in office is increasing in the position within the hierarchy. Denote by $u(\ell)$ the instantaneous utility generated at any position at level ℓ, with $\ell \in \{1,\dots,L\}$, so that $u(\ell) > u(\ell+1)$. Positions within a level are ranked, but the utility flow difference is

---

[89]Appendix Figure 3 shows additional evidence that age limits on Politburo members are strictly and systematically enforced in the post-Deng era, again another sign of break toward institutional regularization.

small. Position 1 at a node at any level ℓ is preferred to position 2, but to reduce complexity, simply refer to each as identically generating a flow of $u(\ell)$.[90]

Politicians cannot leapfrog levels of the hierarchy. An opening for either leadership position at a node in level ℓ is filled by applicants from the level immediately below, level $\ell + 1$. The only exception is positions at entry L (where there is no lower position). Though levels cannot be jumped, positions within a level can. Leaders can move from one level in the hierarchy to the next without having to progress through all the positions at their level. For example, a leader at position 2 in level $\ell + 1$ can be promoted to position 1 in level ℓ without having to first go through the intervening positions.

All eligible leaders from lower positions can apply for openings. It costs an arbitrarily small amount to do so. So, if there is an opening in any node of level ℓ, then all leaders from level $\ell + 1$ will apply. The winning applicant is said to be promoted up a step in the hierarchy.

4.5.2 Factions

There are two factions, denoted B (Blue) and R (Red), and the remaining individuals are neutrals, denoted N.[91] Factions exist to create promotion opportunities for their members and are organized in a hierarchy. A faction can support one and only one member applying for a single position. A faction member not supported by his faction for an opening cannot win promotion against a supported member.[92] For the time being, let us assume factions randomly choose whom to support amongst their eligible candidates.[93]

When a faction holds the paramount position, the effect of promotion support is enhanced, thus increasing the chances of the paramount leader's faction's candidate winning promotion vis-à-vis the other faction candidate and neutrals.

Factions write binding "contracts" with their members determining and restricting how factional support will be allocated. One can never quit a faction and the contract is a quid-pro-quo. On the way up the leadership hierarchy, the faction member will be helped in obtaining positions through the support from the faction infrastructure. If the paramount leader is from his faction, he will receive additional support. If he eventually becomes the paramount leader, the faction member will then offer the same support to the juniors that will follow him in return. This specific characterization of a faction aims at capturing in a stylized fashion the essential patron-client nature of such an organization, as emphasized in Nathan (1973).

Factions are organized geographically (for the sake of exposition and, to a certain extent, realism), in a way that mimics the allocation of power positions within the country. The most senior faction member is the individual with the highest leadership position in the hierarchy. Any faction member occupying a leadership position at level ℓ is senior to a faction member at level $\ell' > \ell$. Faction members are designated by their region. A member who has a position at the top of the government is in the region of the whole country, but a member holding a position at the top of a provincial government is a member of that province and is parallel in faction seniority to a member holding a similar position in another province. This person has factional seniority over all individuals below him in the leadership hierarchy within his province. So, if a member of faction B is the provincial leader in province a, he has factional seniority over any member of faction B who is a village leader in province a. He does not have factional precedence over a village leader in province c, or any other B member who is not in a.

---

[90]Formally: position 1 generates $\varepsilon \to 0^+$ extra utility relative to a position 2 at all ℓ in the same node.

[91]The presence of more than two factions is easily incorporated. Here, we maintain this assumption only for expositional purposes and in line with the empirical analysis that follows.

[92]Or against a neutral.

[93]We will relax this assumption below when we introduce a role for meritocracy and seniority in promotions. If a faction does not support a member, he could, in principle, quit the faction and contest positions as a neutral. We do not allow this, implicitly assuming that the costs of doing this are prohibitive – factions are like the mafia: able to severely punish people who do not fulfil commitments.

Vetoes

Factions exist to facilitate their members' rise through the leadership hierarchy. This requires having both as many members as possible and ensuring that members attain promotions. Each of these dimensions increases the probability of the faction being "powerful", i.e. attaining the paramount leadership. But given that factional support for a contested position can only be given to a single faction member, an individual may have personal incentives that run counter to his faction's objectives. For instance, a member may have an incentive to block the rise of cofactionals who could dilute his own factional support in future competitions for promotion. Factions guard against this by allowing for a seniority veto in allocating support for promotions. Support can be given if, and only if, no faction member *within the region* of the opening and senior to the candidate requesting support blocks it. Thus, when an opening arises in a region, each cofactional at equivalent or higher levels of seniority to the opening in that region can veto the provision of support. Vetoes importantly allow for individuals to block the rise of a member from the same faction who would directly compete with them for factional support in a future opening[94].

Such localized blocking of cofactional members will be very important in determining the shape of factional allocations throughout the hierarchy and the distribution of individuals across factions. The veto ensures that a faction member never has to support someone in his region from his own faction that will directly compete with him for subsequent promotions. At the same time, since the veto is regional, it does not provide so much blocking power that a high up faction member can freeze the advancement of anyone below him anywhere in the country. Providing these limited vetoes is the faction's way of balancing career incentives while lessening the costs of intra-faction rivalry, so that sufficient faction members in the hierarchy have a good chance of attaining the paramount leadership.

4.5.3 The State Variable

In principle, promotion probabilities at each point in time for each politician in the hierarchy will depend on their faction and the distribution of faction members across all other positions in the hierarchy. Hence, we will need to define the full distribution of positions by faction as the state variable of the system. Denote this by S^t at instant t. The state space is thus a $\sum_{\ell=1}^{L} M(\ell)\ell$ dimensional space, with each dimension taking one of three values B, R, N. The state does not change if no position opens up. However, each time an opening happens at a level ℓ, then one individual will be promoted from $\ell+1$ to ℓ to fill the open position, creating an opening at $\ell+1$ leading to one promotion from $\ell+2$, and so on, until the bottom of the hierarchy L, where a new politician enters and chooses his faction. Thus a single opening will lead to a cascade or, what we call, a "chain" of promotions. We assume that these chains occur instantaneously, and if at least one individual moving in a chain replaces an individual from a different faction, then S^t changes.

---

[94]Vetoes can be exercised for a promotion anywhere below in the hierarchy - as long as within one's region of pertinence. However, a politician at ℓ has no interest in vetoing any co-faction member below $\ell+1$. He can always veto them if, and when, they get to $\ell+1$. If vetoes cost even an arbitrarily small amount, they will not be exercised for promotions up to any level lower than ℓ, else they may be wasted (a politician may be promoted to $\ell-1$, potential rivals may retire, etc.). The single exception is where all politicians below are cofaction members. According to the model, a politician above would never let this happen and would have vetoed the rise of one of these cofactionals to avoid such a situation and ensure there is at least one individual below who can be promoted to his accompanying position not from his faction. Reassuringly, this is observed at all levels and for all periods in the data.

4.5.4 Paramount Leadership and Contests

In a competition for promotion with one member supported from each faction and one neutral politician, the probability of winning promotion, for a I faction member is given by the following contest function, $W(I)$:

$$
\begin{aligned}
W(I) &= \frac{i}{\beta + \rho + \eta}, && (4.1) \\
\text{where } i &= \beta, \text{ if } I = B; \\
i &= \rho, \text{ if } I = R; \\
i &= \eta, \text{ if } I = N.
\end{aligned}
$$

β, ρ, and η are parameters determining the strength of faction members in the contest function. Since a faction can only support a single member, the relative value of faction membership for a single politician, compared with being a neutral, depends on both the size of these parameters and the endogenous number of eligible candidates from that faction. Additionally, having the paramount leadership position in your faction helps getting a promotion for the faction's supported candidate. If the paramount leader comes from faction B, we allow $\beta^l > \beta$, and if from faction R we allow $\rho^l > \rho$, thus incorporating leadership premia in the model.[95]

Neutrals contesting a position operate as a somewhat disorganized faction. The overall likelihood of a position going to a neutral is unaffected by the number of neutrals contesting a position, provided there is at least one. Their total contest weight function is η. This treats neutrals symmetrically to factions and can be thought of as a proportional diluting of the neutral support in the same way a faction's support would be diluted were they to forward multiple candidates instead of one.

Promotions and Factional Distributions

The hierarchical structure of positions within the party is taken as given and constant over time.

Promotions arise to fill openings occasioned by a death/retirement or other promotions. As already explained, a single death can have many knock-on effects. At level 1, the instantaneous probability of an opening arising at any position is δ. Since this is the highest level we observe, only death/retirement removes the top leader. However, the instantaneous probability of an opening arising at a post at level 2 comprises the death hazard δ, plus the probability that there was an opening at level 1 and the individual at that level 2 post ascended to level 1 to fill it. This probability of promotion can, in principle, depend on both the factional affiliation of the individual at the post at level 2 and the faction of the individual at the post partnering the opening at level 1. Similarly, the instantaneous probability of an opening at a post at level 3 is δ plus the probability that the individual at the post at level 3 ascended to an opening at level 2 in the hierarchy, and so on. In the estimation Section 4.6 that follows these knock on promotions, or promotion chains, will be explicitly computed.

Let $p_I^{J^t}(\ell)$ denote the probability that an I faction member at level ℓ gets promoted to an opening paired with a J at level $\ell - 1$ at instant t, for $I, J = B, R,$ or N. Let $I^t(\ell)$ denote the number of positions held by faction I at level ℓ, at time t for $I = B, R,$ or N. By definition $M(\ell) = R^t(\ell) + B^t(\ell) + N^t(\ell)$. Since the instantaneous arrival rate of death is δ at any position, there are, in expectation, $I^t(\ell)\delta$ deaths arriving at a position paired with an I at level ℓ each t, and $M(\ell)\delta$ at level ℓ in general at each instant.

Let $\delta_I^t(\ell)$ denote the instantaneous arrival rate of promotions for an I politician at level ℓ. Let $\delta_I^{p^t}(\ell)$ denote the instantaneous arrival rate of a promotion for a politician sharing a node with an I politician at

---

[95]We allow for the possibility of no factional advantage, which might be especially likely at low levels of the hierarchy where the reach of the paramount leader could be muted. Note that it is also the case that a neutral's ascension to the paramount position does not advantage neutrals down the hierarchy.

level ℓ at time t. Consider first the simplest case, which is a promotion from level 2 to the top of the hierarchy $\ell = 1$. Since there are, in expectation, $I^t(1)\delta$ openings arriving for a position paired with an I due to a death, and since at level 1 there is no other way for an opening to arise, the instantaneous arrival of promotion for a I from level 2 is:[96]

$$
\begin{aligned}
&\delta_I^t(2) \\
&= R^t(1)\delta \times p_I^{R^t}(2) + N^t(1)\delta \times p_I^{N^t}(2) + B^t(1)\delta \times p_I^{B^t}(2).
\end{aligned}
\tag{4.2}
$$

We can now similarly compute the arrival of promotions from level 3 to level 2. Intuitively, the possibility of these arises when either a leader at level 2 dies, or is himself promoted to level 1, which in turn depends on a death at level 1, as specified in equation (4.2). Using these, we can compute the instantaneous arrival of promotions for an I from level 3 at t as depending on the probability of a position paired with an I being promoted or dying. The instantaneous death arrival of such an individual is δ, the probability of the paired partner being promoted is $\delta_I^{p^t}(2)$ for each of the factions I at level 2 at time t, hence:

$$
\begin{aligned}
&\delta_I^t(3) \\
&= R^t(2)\left(\delta + \delta_R^{p^t}(2)\right) \times p_I^{R^t}(3) + N^t(2)\left(\delta + \delta_N^{p^t}(2)\right) \times p_I^{N^t}(3) \\
&\quad + B^t(2)\left(\delta + \delta_B^{p^t}(2)\right) \times p_I^{B^t}(3).
\end{aligned}
$$

Similarly, continuing down the hierarchy, we have for any level $\ell > 2$:

$$
\begin{aligned}
&\delta_I^t(\ell) \\
&= R^t(\ell-1)\left(\delta + \delta_R^{p^t}(\ell-1)\right) \times p_I^{R^t}(\ell) + N^t(\ell-1)\left(\delta + \delta_N^{p^t}(\ell-1)\right) \times p_I^{N^t}(\ell) \\
&\quad + B^t(\ell-1)\left(\delta + \delta_B^{p^t}(\ell-1)\right) \times p_I^{B^t}(\ell).
\end{aligned}
\tag{4.3}
$$

(4.3) explicitly shows that the arrival rate of I promotions at level ℓ depends not only on the distribution at level $\ell - 1$, i.e. on $B^t(\ell-1), R^t(\ell-1), N^t(\ell-1)$, but also, through each of the $\delta_I^{p^t}(\ell-1)$, on $B^t(\ell-2), R^t(\ell-2), N^t(\ell-2)$. A convenient feature of our model specification is that higher levels of the hierarchy enter recursively, allowing the computation of arrival rates for all I factions all the way down the hierarchy.

Let us now consider the explicit form of the $p_I^{J^t}(\ell)$ using the contest function (4.1). We begin by assuming that the other faction $J \neq I$ will support one of its members for the position as well.

In this case $p_I^{J^t}(\ell)$ is determined as follows:

$$
p_I^{J^t}(\ell) \equiv \frac{1}{I^t(\ell)} \times
\begin{cases}
\frac{i}{i+k+k'} & \text{if } K^t(\ell),\, K'^t(\ell) > 0, \\
\frac{i}{i+k} & \text{if } K^t(\ell) > 0,\, K'^t(\ell) = 0, \\
\frac{i}{i+k'} & \text{if } K^t(\ell) = 0,\, K'^t(\ell) > 0, \\
1 & \text{if } K^t(\ell),\, K'^t(\ell) = 0,
\end{cases}
\tag{4.4}
$$

$$
\begin{aligned}
\text{where } \{I^t, i\} &= \{B^t, \beta\}, \{R^t, \rho\} \text{ or } \{N^t, \eta\}; \\
\{K^t, k\} &\neq \{I^t, i\}, \{K'^t, k'\}; \\
\{K'^t, k'\} &\neq \{I^t, i\}.
\end{aligned}
$$

Note that J (the faction of the politician that the opening at level $\ell - 1$ is paired with) does not enter directly into the probability of winning a promotion contest. But this is because specification (4.4) assumes that if

---

[96]This expression uses the fact that in continuous time simultaneous hazards do not arrive. That is, we put zero weight on the probability of a death opening occurring at the same instant in two positions.

members of another faction are present, one of them will always be supported in the contest for the position. As we now demonstrate, this will not always be the case, which will in fact simplify the expression above considerably:

Proposition 1. *i) A politician from faction J at level ℓ will veto the support of a cofactional member ascending to his level from $\ell + 1$ at t if there are members of both $I \neq J$ and neutrals, N, at level ℓ.*

ii) If there are no members of faction I at level ℓ, a politician from faction $J \neq I$ at level ℓ will veto a member of his own faction from $\ell + 1$ at t if the number of cofactional members at level ℓ is such that $J^t(\ell) < \frac{j + \eta}{i}$ where $i = \beta$, $j = \rho$ if $J^t = R$ and $i = \rho$, $j = \beta$ if $J^t = B$.

Proof. All proofs are in Appendix. □

If both R, B types are represented at a politician's level, he will gain by vetoing the ascension of a competitor from his own faction, as this increases the probability that his faction will support him for a subsequent opening at the level above. However, if all other factions are not already present, then he faces a trade-off. By vetoing a cofactional's promotion the party member still improves his chances of gaining factional support. But he also increases the chance that a member of a rival faction, which was not already present, gains entry to the group of competitors. This lowers the chances of him winning promotion conditional upon receiving the support of his faction. The sufficient condition in the statement of Proposition 1 ensures that the former effect dominates the latter. From now on, we proceed under the assumption that the sufficient condition for vetoes holds, so that we continue to see them throughout our observations. We will verify that this is indeed the case in the data, so we do not dwell on weaker necessary conditions for vetoes to hold further.

Following Proposition 1, vetoes generate a large amount of structure to the pattern of openings – meaning no two cofactional members will ever be paired at the same node. We have already verified in Section 4.4 that this is, in fact, a systematic feature of the data. Moreover, the prospects of promotion at any node depend not only on the distribution of openings immediately above, but also on the distribution of openings further up, as these determine the chances that a politician immediately above will himself be promoted. Promotion chances at all levels are affected by the full distribution of positions above. We can compute this explicitly using the recursive structure of the $\delta_I^t(\ell)$ terms and our results on vetoes.

Proposition 2. *The instantaneous arrival rate of promotions at each level of the hierarchy is as follows.*

Let $\mathrm{I}_B^t = 1$, iff $B^t(\ell) > 0$ and $\mathrm{I}_B^t = 0$, otherwise; $\mathrm{I}_R^t = 1$, iff $R^t(\ell) > 0$ and $\mathrm{I}_R^t = 0$, otherwise; $\mathrm{I}_N^t = 1$, iff $N^t(\ell) > 0$ and $\mathrm{I}_N^t = 0$, otherwise.

For an N member:

$$\delta_N^t(\ell) = \frac{\eta}{N^t(\ell)}$$
$$\times \left(R^t(\ell-1) \frac{\left(\delta + \delta_R^{p^t}(\ell-1)\right)}{\mathrm{I}_B^t \beta + \eta} + N^t(\ell-1) \frac{\left(\delta + \delta_N^{p^t}(\ell-1)\right)}{\mathrm{I}_B^t \beta + \mathrm{I}_R^t \rho + \eta} + B^t(\ell-1) \frac{\left(\delta + \delta_B^{p^t}(\ell-1)\right)}{\mathrm{I}_R^t \rho + \eta} \right).$$

For a B member:

$$\delta_B^t(\ell) = \frac{\beta}{B^t(\ell)}$$
$$\times \left(R^t(\ell-1) \frac{\left(\delta + \delta_R^{p^t}(\ell-1)\right)}{\beta + \mathrm{I}_N^t \eta} + N^t(\ell-1) \frac{\left(\delta + \delta_N^{p^t}(\ell-1)\right)}{\beta + \mathrm{I}_R^t \rho + \mathrm{I}_N^t \eta} \right).$$

For an R member:

$$\delta_R^t(\ell) = \frac{\rho}{R^t(\ell)}$$

$$\times \left(B^t(\ell-1) \frac{\left(\delta + \delta_B^{p^t}(\ell-1)\right)}{\rho + \mathrm{I}_N^t \eta} + N^t(\ell-1) \frac{\left(\delta + \delta_N^{p^t}(\ell-1)\right)}{\mathrm{I}_B^t \beta + \rho + \mathrm{I}_N^t \eta} \right).$$

For each one of these expressions we can see the negative dependence on the prevalence of one's own faction members. Take for example the last expression for R. The greater the number of other R's at level ℓ at t, the more diluted is an R's support (i.e. the lower the probability that any given R member will be chosen by the faction as the one to be supported), as per $R^t(\ell)$ in the denominator. Further, the more frequent the R's at level $\ell-1$ the harder it is to get an opening for which an R at ℓ will not be vetoed (e.g. at the extreme if $R^t(\ell-1) = M^t(\ell-1)$, then $\delta_R^t(\ell) = 0$). This is true for all levels of the hierarchy from the recursion of these equations.

The proposition highlights the possible down side of factional affiliation. Though factions have the potential to provide support for promotions such support is decided by cofaction members sitting above one in the hierarchy. They will never let a junior member contest with them for their own future promotions so, in a sense, the rise of the junior is tethered to, and thus depends upon, the rise of his cofactional seniors. If they do not rise, then, not only do they not generate the extra support that comes from the paramount leadership, they actively block their own juniors from ascending in their place. In other words, the proposition formalizes a form of natural pecking order in the factional structure, a feature that appears realistic in large organizations.

Finally, note that each statement of δ_I^t in Proposition 2 ignores the effect of a faction's holding of the paramount leadership on promotion (i.e. i^l). Effectively δ_I^t is written for the case of an N in paramount leadership. In the Appendix we state the full set of δ_I^t conditional upon paramount leadership affiliation.

4.5.5 Entry

Entry into the hierarchy of political positions occurs only at the lowest level, L. An entering politician at instant t decides which faction to join when starting his politician career, or to contest as a neutral, and bases this decision on the discounted expected utility he will receive via each one of the options. He maximizes his discounted expected utility stream:

$$V^t = \int_t^\infty e^{-\delta s} v^s ds$$

where v^t is the instantaneous utility at t. We formally consider this decision here. Recall that $u(\ell)$ denotes the politician per instant payoff to holding a position at level $\ell \in \{1, L\}$ in the hierarchy. So that if a politician holds a position at ℓ at instant t then $v^t = u(\ell)$. Define the corresponding value function for a politician of type $I = B, R, N$ at level ℓ at instant t by, $V_I^t(\ell)$. This is related to the promotion probabilities, $\delta_I^t(\ell)$, via the Bellman equation:

$$\delta \mathbb{E}^t V_I^t(\ell) = u(\ell) + \delta_I^t(\ell) \mathbb{E}^t \left[V_I^t(\ell-1) - V_I^t(\ell) \right] \tag{4.5}$$

The expectations operator appears in the expression because the value of being a type I politician at ℓ depends on the instantaneous probability of being promoted to level $V_I^t(\ell-1)$. Though this is known at instant t, via $\delta_I^t(\ell)$, the value of being at this higher level in turn depends on the evolution of $\delta_I^t(\ell)$. The evolution of these $\delta_I^t(\ell)$ promotion probabilities themselves depend on the state of the system, S^t, which is changing continuously in a stochastic manner due to deaths, openings, and promotions occurring through time via the contest function (4.1).

The entering politician at t chooses the faction with the highest expected utility stream:

$$\sup_{I \in \{B,R,N\}} \left\{ \mathbb{E}^t V_B^t(L), \mathbb{E}^t V_R^t(L), \mathbb{E}^t V_N^t(L) \right\}. \tag{4.6}$$

After entry, since a politician is fixed in his faction from then on, his choices are simple. He will apply for all promotions to which he is eligible, and he will veto according to Proposition 1. We consider the more difficult problem of the initial entry decision (4.6) now.

4.5.6 Equilibrium Behavior

Entering politicians will choose to enter the faction (or remain neutral) yielding the highest expected utility, which implies choosing the faction guaranteeing in expectation the fastest progression through the hierarchy. The most immediately relevant information for the agent will be the arrival of promotions if he/she registers as a I politician from level L to $L-1$, but one cannot specify, a priori, the relative weight an entering politician puts on the chances of being promoted at higher levels of the hierarchy compared to lower levels. Perhaps politicians care little about regional promotions, that occur early in their career, but greatly about promotions from the province to the central government. Conversely, politicians may put substantial value on their immediate entry prospects. Note that, indirectly at least, the relative performance of factions at higher levels already enters into a politician's evaluation of promotion at the lowest level, L, since openings immediately above depend negatively on the frequency of cofactional politicians all the way up the hierarchy; as discussed above after Proposition 2. At any point in time this valuation will depend on the full distribution of positions higher than the politician, that is on S^t, the high-dimensionality state space of the system. Without mapping the full form of expected hierarchy evolution, it is not possible to compute the value function $V_I^t(\ell)$ analytically. However, it is possible to establish a sufficient condition under which optimal entry ensures that along any time path all factional types and neutrals will be observed in equilibrium:

Proposition 3. *With $M(\ell)$ large enough for all ℓ, any equilibrium necessarily involves politicians in factions B, R, and N.*

Intuitively, with sufficiently many openings at all levels of the hierarchy, the value of entering via a faction (or as a neutral) that is not already present will eventually outweigh even the largest parametric disadvantages of that faction (or being a neutral). That is, for example, even if $\beta \ll \rho$ (so ceteris paribus it is better to enter as an R than a B), if there are sufficiently many positions in the hierarchy, a large number of R members and Proposition 1 will imply that the expected promotion rate will be faster if entering as a (rare) B member over entering as (one of the many) R. Thus, though we are not able to fully characterize optimal entry in an equilibrium, the sufficient condition of the proposition ensures that *any* equilibrium distribution of positions that we do observe will feature both factions and neutrals.

4.5.7 From Model to Data

Openings in the hierarchy occur at any point in time via the functions in Proposition 2. Other than through the effect of time on the changing distribution of factions across the hierarchy $\mathbf{S}^t$, which the model explicitly accounts for, the process leading to openings occurs independently of time (conditional on $\mathbf{S}^t$)[97].

Treating openings this way amounts to assuming that openings are independent events caused by exogenous factors, each triggering a chain of knock on effects. This assumption may be violated at the time of Chinese Communist Party Congresses, when there appears to be a large number of shuffles at different

---

[97]In what follows below we will dispense with the time index t for the empirical analysis.

levels of the hierarchy observed in a way that seems simultaneous, not sequential. Indeed, for the most part, the data is observed at low frequency, i.e. at each CCP Congress $T, T + 1, \ldots$ This implies that the promotion chains that our model postulates are not fully observable, so simulation methods will be necessary to link two subsequent $\mathbf{S}^T, \mathbf{S}^{T+1}$.

To operationalize the model in our specific empirical setting, we will assume that the simultaneity observed in exits and promotions reflects a particular structure, as follows.

First, we purge all individuals from all positions that we observe leaving the data in between snapshots $T, T + 1, \ldots$ That is, all individuals who are no longer present between times T and $T + 1$ are assumed to have retired at some point between two Congresses.

Second, openings are filled through a sequence of promotion chains. Each chain starts with the highest ranked exit in the sample and selects politicians to fill in the knock-on openings sequentially. This continues until all the exits and promotions between $\mathbf{S}^T$ and $\mathbf{S}^{T+1}$ are accounted for and all positions have been filled. Because there are many sets of promotion chains that can rationalize the observed openings in the data, Section 4.6 shows how simulation methods can be used to transparently address this issue in practice.

Third, for positions for which there is no explicit dyadic structure in the data, we draw at random a paired politician from the set of potential matches at the level at which the promotion occurs.

4.5.8 Discussion of the Model

Before moving to the estimation of the model, we offer here a brief discussion of an alternative modeling choice and justify our specific line of reasoning empirically.

Perhaps the best alternative to our individual career concerns model is a model that views the allocation of positions as the outcome of factional bargaining. In such a model the faction, as opposed to the individual politician, is the decision maker, and factions negotiate with each other over the allocation of positions in the hierarchy. Negotiations would favour the faction holding the paramount leadership position, and could thus easily exhibit the patterns of increased representation at all levels with leadership of a faction. The relative overall balancing could also be supported as an equilibrium outcome that ensures peace. If a leader comes from faction B he is not willing to completely expropriate faction R because he fears dissent from R. Dissent in extreme cases could take the form of revolt that would destabilize not just his own position, but, in the limit, the overall hold of the party. So positions could be still allocated to the other faction, as a price for peace. Reciprocally, the other faction might show similar restraint if it ascended to the paramount position. Anticipating this, the current leader would have further incentives to be moderate and inclusive in allocating positions.

Problems arise for this alternative story when the actual distribution of positions — and not just their overall number — is scrutinized further, as done in Section 4.4. There we observed that a pronounced pattern in the data was the omission of (B, B) and (R, R) pairs at leadership nodes. R members are more likely to be accompanied by B members than by N members and much less likely to be accompanied by another R — which is extremely rare. Why? One explanation consistent with factional negotiations is that each faction fears that the other faction may gain control of the node. If a B is in place, placing an R alongside him ensures that the B members do not gain permanent control of the node. But this sort of concern does not seem likely, as there does not appear to be evidence of such permanent nodal control in the data. We observe shuffling of cadres occurring regularly for Provinces for instance. There does not seem to be lock in of factions to posts. B members are replaced by R members at a node with the R subsequently replaced by another B. This additional evidence is available by the authors upon request.

But shuffling could itself be the strategy that factions employ to ensure that control does not get held too strongly. We may see B members replaced by R members in order to ensure that the B members do not hold the position at the node too strongly. But if this is the case, a further puzzle arises. A process of shuffling — though able to easily explain (B, R) nodes — would not especially favour these. We should also regularly

see (B,B) nodes and (R,R) nodes, which are then replaced by (R,R) and (B,B) nodes immediately after. If shuffling is used to avoid entrenchment, then there is no reason that intertemporal sharing of the nodes should not be sufficient to achieve this. There should be no particular reason to see the proliferation of (B,R) nodes that we observe in the data.

This seems to be the single most difficult fact to explain with a factional balancing model. Our individual career-concerns model explains this directly. In our view, the model we develop has a further advantage relative to a model that treats the factions as bargaining parties in that it sets the individual as the decision maker in a microfounded way.

4.6   Maximum Simulated Likelihood Estimation

This Section describes our estimation methodology. Define Y the observed data on career outcomes (i.e. promotions, exits, etc.) between two Congresses T and $T+1$ and X the observed data on the hierarchy plus a set of individual characteristics (i.e. X includes factions and position within the hierarchy/level $\mathbf{S}^T$, plus individual covariates).

We define $\mathtt{k}$ as a set of promotion chains, so that $\mathtt{k} = \{\mathtt{k}(1),\mathtt{k}(2),...\}$, where each chain $\mathtt{k}(c)$ of politicians (say, $s^0, s, s',$ and s'') is simply a set of politicians each belonging to different, but adjacent hierarchical levels ℓ, whose promotions were triggered by the exit of the highest ranking one of the chain (e.g. when s^0 dies or retires at $\ell = 1$, s is promoted from $\ell = 2$ to $\ell = 1$, then s' replaces s at level $\ell = 2$, and then s'' replaces s' at $\ell = 3$).

A chain starts from an opening at level $\ell - 1$ and involves promotions from ℓ all the way down to L.[98]

We impose that each politician promoted in the data belongs to exactly one chain and that each change between $\mathbf{S}^T$ and $\mathbf{S}^{T+1}$ is part of at least one chain $\mathtt{k}(c)$. (A politician promoted by two levels between T and $T+1$ will need to belong to two separate promotion chains.) Let $C = \#(\mathtt{k})$ be the number of promotion chains in set $\mathtt{k}$.

The unconditional likelihood of observing Y given X is:

$$f(Y|X) = \mathbb{E}_{\mathtt{k}}\left[f(Y|X,\mathtt{k})\right].$$

Define $Y_{\mathtt{k}(c)}$ as the set of career outcomes pertinent to the individuals involved in promotion chain number c of $\mathtt{k}$. Because the structure of the political hierarchy will change once a promotion chain is realized (i.e. the interim $\mathbf{S}$ will change), positions within the hierarchy/level and factional affiliations at all levels $X_{\mathtt{k}(c)}$ need to be modified after each chain $\mathtt{k}(c)$ is realized.

The conditional likelihood upon the realization of a set of promotion chains $\mathtt{k}$ happening over time is given by:

$$f(Y|X,\mathtt{k}) = \prod_{c=1}^{C} f(Y_{\mathtt{k}(c)}|X_{\mathtt{k}(c)},\mathtt{k}).$$

The likelihood contribution $f(Y_{\mathtt{k}(c)}|X_{\mathtt{k}(c)},\mathtt{k})$ of a chain $\mathtt{k}(c)$ of promotions initiated at $\ell - 1$ involves computing the conditional promotion probabilities of all individuals involved in $\mathtt{k}(c)$ at the various levels, down to L. A promotion from level ℓ to level $\ell - 1$ to be paired to a politician $K = R,B,N$ is a random event distributed over a discrete support formed of $M(\ell)$ points (individual politicians), $B(\ell)$ of which occurring with probability $p_B^K(\ell)$, $R(\ell)$ occurring with probability $p_R^K(\ell)$, and $N(\ell)$ occurring with probability $p_N^K(\ell)$. (We omit time indexes as they are unnecessary here.)

---

[98]Plus a new entry at the lowest level, which we do not model, as per our discussion of Proposition 3. The entry choice is not necessary for estimation and all parameters are identified without its addition.

Given the independence of the promotion events across levels, the construction of this likelihood is straightforward. Let I^ℓ be the faction of the individual belonging to $\mathrm{k}(c)$ at level ℓ and $J^{\ell-1}$ be the faction of the individual with which s/he is paired when promoted to level $\ell - 1$:

$$f(Y_{\mathrm{k}(c)}|X_{\mathrm{k}(c)}, \mathrm{k}) = \delta \prod_{l=\ell}^{L} p_{I^l}^{J^{l-1}}(l).$$

Going back to the example above of a chain of politicians $s^0, s, s',$ and s'' belonging to factions N, R, and B respectively, and assuming they all happen to get paired with N-type politicians, the likelihood contribution of this chain is:

$$f(Y_{\mathrm{k}(c)}|X_{\mathrm{k}(c)}, \mathrm{k}) = \delta \times p_N^N(2) \times p_R^N(3) \times p_B^N(4)$$

where each probability $p_{I^\ell}^{J^{\ell-1}}(\ell)$ is computed based on $X_{\mathrm{k}(c)}$, ordered from the top promotion to the level L promotion, as imposed by the sequential nature of the promotions comprised in each chain.

The Maximum Simulated Likelihood (MSL), for given number of simulated sets of promotion chains R_K,[99] is:

$$f(Y|X) = \frac{1}{R_K} \sum_{r=1}^{R_K} \prod_{c=1}^{C} f(Y_{\mathrm{k}_r(c)}|X_{\mathrm{k}_r(c)}, \mathrm{k}_r).$$

This is the estimator that we employ.

4.7 CCP Factional Politics: Structural Results

This section presents MSL estimates of the model and sample fit assessments. The sample includes all the members of the 14th-18th Central Committees in the post-Deng era. The simulation procedure in Section 4.6 was first implemented in a series of Montecarlo simulations and successfully probed for: i) identification of the structural parameters; ii) sensitivity to misspecification in the number of factions; and iii) sensitivity to misspecification in the contest function we use[100].

We begin our analysis with the most parametrically parsimonious model possible, one where we normalize $\eta = 1$ and the two faction parameters $\{\beta, \rho\}$ are estimated on top of a single leadership premium λ, defined as $\lambda = \beta^l/\beta = \rho^l/\rho$. The MSL results for this model are reported in Column 1 of Table 4.9. The estimated contest function parameters are 0.045 and 0.029 for CYLC and Shanghai Gang respectively, which are close to the average share of seats in the Central Committee. The estimated leadership premium λ is 2.553, implying that a faction candidate is more than twice as likely to be promoted when the paramount leader is from the same faction. The magnitude of the leadership premium is consistent with the reduced form evidence in Table 4.6. All parameters driving the promotion process across factions are precisely estimated.

Because it may seem restrictive to assume a common contest function across all levels of the CCP top echelons (which include heterogeneous layers in both size and jurisdiction, such as the top CCP positions and the PBSC, PB, CC, AC), Column 2 in Table 4.9 allows for level-specific parameters $\{\beta_k, \rho_k\}_{k=H,L}$ for the PB and higher versus CC and lower. The parameter estimates show that faction affiliation helps significantly more at higher levels than that at lower levels within the CCP: the estimated contest function parameters reach 0.162 and 0.193 at the PB and higher for CYLC and Shanghai Gang relative to CC and AC levels of 0.041 and 0.022.

---

[99]We employ 100 simulated chains sets for each CCP National Party Congress.

[100]All simulation results are available upon request.

One may also wonder whether the leadership premium differs across factions. Column 3 explores this possibility by allowing for faction-specific leadership premia $\{\lambda_R, \lambda_B\}$. The parameter estimates show that two factions have very similar premia (both are between 2 and 3). The improvement of log-likelihood is negligible, indicating that the two factions operate in a similar fashion. This result is also consistent with the reduced-form evidence in Table 4.6.

Column 4 in Table 4.9 combines both level-specific parameters $\{\beta_k, \rho_k\}_{k=H,L}$ and faction-specific leadership premia $\{\lambda_R, \lambda_B\}$. We conduct Likelihood Ratio (LR) tests for model 1, 2 and 3 against model 4 (numbering indicates the Column of reference). LR tests reject model 1 and 3, which impose a constant contest function across levels, against model 4, but do not reject model 2, which imposes a constant leadership premium across factions. In the following analysis, we will thus use the more parsimonious model 2 as our benchmark and refer to it as the baseline faction model.

Figure 4.6 provides a visual representation of the factions' seat shares by level as predicted by the model. The five bars represent the five levels of the Central Committee (the top two CCP positions, PBSC, PB, CC, and AC). The blue, white, and red parts represent the seat shares of the Shanghai Gang, Neutral, and CYLC respectively. The left panel is the data, while the right are the predictions of our baseline faction model. Our baseline faction model successfully replicates the distribution of factions across different levels of the hierarchy: faction members are relatively scarce in the lower levels, but become increasingly concentrated in the higher ones. This is related to the increasing contest function parameters estimated above, which imply an increasing advantage of factional affiliation as one progresses up the hierarchy. Notice that our model also captures the inertia of the factional composition of the various levels over time evident in the data thanks to the slow percolation of factional members up the hierarchy. The intuition is that promotions and retirements occur gradually over time. It takes time for a faction leader to grow his inner circle from the bottom of the hierarchy up. Interestingly, such dynamics can function as checks and balances on an incoming paramount leader. When a new leader first assumes power, he is likely to be surrounded by members from rival factions. There is also anecdotal evidence in line with this finding: Jiang himself once described his first few years as the General Secretary "*as standing on the brink of a deep ravine, or walking on thin ice*"[101]. Bo (2004) also suggests that the Shanghai Gang continued to exert strong influence in the first term of Hu Jintao. This finding will be particularly useful in understanding the upcoming second term of General Secretary Xi, expected by many observers to gain greater clout relative to his first term in office[102].

Our faction model also provides insights for the dynamics of power transition between factions. Figure 4.7 plots the aggregate share of promotions of each faction over time[103]. The share of promotions is defined as the ratio between the number of promotions for a faction and the total number of promotions. Again, the fit of the model is good. Figure 4.7 points also to a more subtle implication of our model: there are no discontinuous drops in the share of promotions of the paramount leader's faction right after he retires. When Jiang Zemin retired after the 15th Party Congress, a large share of the Shanghai Gang continued to be promoted to the 16th Central Committee. The pattern was repeated at Hu Jintao's transition to Xi Jinping at the 18th Party Congress. In reality there is uncertainty over the precise point at which the influence of the incoming paramount leader eclipses that of the departing incumbent and this influences promotion rates. Scholars have suggested that Deng retained considerable influence well after formal retirement in 1989; Jiang maintained informal and formal military oversight after stepping down as General Secretary. A retiring paramount leader may continue to shape the composition of the next Central Committee. Such intricate dynamics are captured by our simulation approach that draws different paramount leader transition dates across multiple simulations, smoothing out sharp discontinuities around the official power transition date.

---

[101]See Kuhn (2005).

[102]The 19th Congress is currently scheduled in the Fall of 2017.

[103]A more detailed breakdown by level of the Central Committee can be found in Table 4.11.

4.7.1 Adding Individual Covariates

So far we have assumed that faction members are selected to challenge a post randomly within a faction and level (modulo vetoes, of course). We can easily add individual characteristics, Z, to the within-faction selection process as well. Consider each row of the matrix Z_s to be a vector of characteristics for politician s. Define $q_{I,s}(\ell)$ as the probability that s of faction I is selected as the candidate of this faction at level ℓ, also define $\mathscr{A}_I(\ell)$ as the set of the members of faction I at level ℓ. We assume a within-faction selection probability of the logistic form[104]:

$$q_{I,s}(\ell) \equiv \frac{\exp(\gamma Z_s)}{\sum_{s' \in \mathscr{A}_I(\ell)} \exp(\gamma Z_{s'})}.$$

Therefore, the probability of winning promotion can be rewritten as $q_{I,s}(\ell) \times W(I)$. Notice that our baseline faction model is nested in this formula by setting coefficients of individual characteristics, γ, to 0. In this case we get back our random within-faction selection probability, $q_{I,s}(\ell) \equiv \frac{1}{I(\ell)}$. We refer to the above model as the faction model with individual characteristics.

The parameter estimates are reported in the Column 2 of Table 4.10. Comparing with the baseline faction model in Column 1, we see a reasonable improvement in model fit measured by log-likelihood. At the same time, however, we observe little change in the estimates of the parameters for the contest function and the leadership premium, suggesting that these parameters are indeed more related to the technology of factions than to individual covariates omitted in the baseline model. Examining the estimated coefficients of individual characteristics, we find that being a princeling or a male increases the probability of promotion, while having a graduate degree or being an ethnic minority hurts. The effect of age is non-linear: it has a positive effect at first, but eventually negatively affects promotion chances, in line with previously observed hard age limits enforced within the CCP.

4.7.2 Alternative Models

Given our main specifications, we are equipped for both in-sample and out-of-sample fit analysis of our structural model. It is useful in this respect also to present some alternative benchmarks to which we can compare our model's performance. First, we can use as the simplest alternative a model based on random promotion. This is done by setting:

$$p(\ell) \equiv \frac{1}{M(\ell)}.$$

Second, we implement a pure seniority-based promotion mechanism, setting for politician s:

$$p_s(\ell) \equiv \frac{\varphi(age_s)}{M(\ell)},$$

with $\varphi(.)$ a (third order) polynomial in age[105].

Figure 4.8 provides the scatter plots of model predicted shares of promotions by Party Congress and by level of the CCP against the data[106]. Our models (baseline faction and faction with individual characteristics) handily outperform both the random and the seniority models: the predicted shares by the faction models line

---

[104]Since our data only includes the top 5 levels of the party hierarchy (President/Premier, PBSC, PB, CC, AC), individual characteristics of the potential candidates eligible for promotion to AC are not always observable to us. As a result, we assume within-faction selection is random below the AC level.

[105]For the seniority and random model, we calibrate the probability of entering AC using the average share of each faction in the Central Committee.

[106]We combine level 1, 2 and 3 because there are two few observations in the first two levels.

up with the data nicely along the 45 degree line, whereas the shares predicted by random and pure seniority-based promotion models appear completely flat. This result is independent of which Party Congress we consider. More concretely, with only five structural parameters our baseline faction model reduces the mean squared errors of the predicted shares of promotion by more than 80 percent comparing to the random model, and more than 70 comparing to the seniority model. The right panel of Table 4.10 conducts formal specification tests. The Vuong statistics reject the random and the pure seniority-based promotion models against the faction model with individual characteristics[107].

What about political meritocracy? Because the CCP promotion model is by many referred to as a strictly meritocratic mechanism (Li and Zhou, 2005; Bell, 2015) and there is substantial debate as to whether systematic assessment of cadres based on economic performance plays a role in the CCP, we test our model against this third "purely meritocratic" mechanism. First, in order to find a suitable measure of performance, we need to restrict our analysis of promotions to provincial leaders in the Central Committee. We associate these prominent provincial CCP cadres with the economic performance (in terms of real GDP growth over their tenure) of their Province of service –precisely as in Li and Zhou (2005) and Jia et al. (2015)– and use this as a (admittedly rough) proxy for overall performance. Graphical evidence of the performance of our model is reported in Figure 4.9[108]. In the appropriate subset of promotions (i.e. those for which performance metrics are available), our model performs better than the purely meritocratic model in terms of sum of mean squared errors, which is reduced by 35 percent.

We further examine the out-of-sample fit of our model. Specifically, we re-estimate the model using only the 14th-17th Central Committees and predict the shares of promotion of each faction at 18th Central Committee. We compare the model predicted share of promotion with the actual data in the scatter plot in Figure 4.10. Our faction model again outperforms random and seniority-based models in terms of out-of-sample fit. The reduction in mean squared errors is 77 percent and 69 percent comparing our baseline faction model to the random model and seniority model respectively, reassuring us of its robustness.

4.8 Counterfactuals and Model Analysis

Within our econometric framework we can explore a set of counterfactual exercises and present an additional quantitative analysis of several questions relevant to the study of Chinese political economy.

4.8.1 Forgoing Collective Leadership

We begin by exploring an historical counterfactual on leadership premia in the post-Deng era. Our model explicitly recognizes such premia, but a wealth of anecdotal discussion in Chinese politics (and the empirical evidence of Section 4.4) suggests them to have been curtailed in the post-Deng era. This peculiarity of the post-Deng Chinese system, the emergence of so-called "collective leadership", has been frequently recognized in the literature. It is often indicated as the main structural break from the strongman political equilibria thought to have prevailed under Mao Zedong[109] and the paramount leadership of Deng Xiaoping (Tsou, 1995; Fewsmith, 2001; Shambaugh, 2008). This exercise is also useful in perspective of the present, latent changes in Chinese politics. Scholars like Nathan (2016) suggest President Xi may be "*overturning*

---

[107]The pure seniority-based model outperforms the baseline faction model in terms of log-likelihood. However, this is driven by the fact that only 10% of the politicians have factional affiliation. After we include individual characteristics in the factional model, the pure seniority-based model is easily rejected by the Vuong test.

[108]In this scatter plot, we do not break down the share of promotion by level because of the small number of observations in the subset of provincial leaders.

[109]"*During the Maoist era, factions were ideologically as well as personally defined, and remained fiercely loyal in what could become a winner-take-all game.*" Dittmer (2004, p.18)

Deng's system", as he "*has taken the chairmanship of the most important seven of the twenty-two leading small groups that guide policy in specific areas*" and "*tightened direct control over the military*".

Here, we will ramp up the limited role played by leadership premia in factional representation in China and present a counterfactual of what would have happened under heightened winner-take-all type factional competition. We run the model with twice as high a leadership premium λ.

Results are reported in Figure 4.11. A more detailed breakdown by level can be found in the third panel of Table 4.11. The counterfactual is implemented by simulating for each Congress T the share of promotion of each faction to the following Congress $T + 1$. Under the Jiang Zemin (Shanghai Gang) presidency, openings in the Politburo and the Central Committee are filled with more of the top leader's cofactionals. Under the CYLC leadership of Hu Jintao, numbers would have been comparable, swinging in the opposite direction with more *tuanpai* members promoted. The magnitude of the increase in the shares of promotions, however, is less than the increase in the leadership premium. The dampening effect emerges from the factional veto mechanism detailed in Proposition 1. As members of a faction become crowded at a certain level ℓ, new promotions from the same faction are more likely to be blocked by their cofactionals out of their own career concerns. Indeed, as shown in Table 4.11, the dampening effect is stronger in higher levels of the hierarchy where faction members are more concentrated (Figure 4.6). Therefore, individual incentives in intra-faction competition counterweigh the ability of a paramount leader to grow his own faction.

4.8.2   Li Keqiang Presidency

A second counterfactual we perform involves the choice of leadership ratified by the 2012 18th Party Congress. This is the event that brought Xi Jinping to the PRC Presidency. Nathan and Gilley (2003) present compelling documentary evidence that already ten years before the formal power transition Xi Jinping and Li Keqiang, the current PRC Premier, belonged to a select few with potential accreditation to the paramount post. Bo Xilai was also part of this highly selected group.

It is possible for us to study a counterfactual Li presidency. Figure 4.12 reports the aggregate share of promotion, and a more detailed breakdown by level can be found in the fourth panel of Table 4.11. Interestingly, given the estimated leadership premia, the promotion at PB level would have had a very limited increase in CYLC representation (Li's faction). More radical shifts would have been recorded in the promotion at the CC and AC though. Again this is a result of the slow percolation of factional representation induced by our model, compounded with the already high CYLC representation at the upper levels of the CCP at the end of Hu's last term in office.

4.8.3   Are Princelings a Faction?

The reader will notice that the analysis above posits factional affiliation of president Xi Jinping as a member of Shanghai Gang. This is in itself a matter of debate among scholars interested in Chinese elite politics. For instance, Li (2013) in his bi-factional representation of the Chinese top tiers defines Xi as a princeling associated with Jiang's camp (Shanghai Gang). In fact, Xi spent only seven months in any official role in Shanghai, but Jiang's substantial influence on Xi has been noted by many. Other researchers have pointed to President Xi as the leader of a new faction of his own, mostly with roots in Shaanxi, where Xi was born, and in Zhejiang Province, where he served as Party Secretary from 2002 to 2007[110]. Our model allows a formal statistical analysis of some of these questions.

---

[110]Some recurring affiliated politicians include current PBSC member and anti-corruption czar Wang Qishan, and potential PBSC future members such as Li Zhanshu, director of the CC General Office, and Politburo member Zhao Leji. Shih (2016) estimates, based on shared career experience, that less than 6 percent of current CC members have past ties with President Xi. This should however not be confused with a truly factional organization of the President's inner circle for which hard evidence is not available.

We begin by investigating whether our postulate of the princelings not behaving as a unified faction is warranted by the data. To assess this formally we implement Vuong specification tests between our baseline model and one where princeling status is coded as membership in faction P, with a specific parameter π regulating an expanded contest function of the type (4.1):

$$W(P) = \frac{\pi}{\beta + \pi + \rho + \eta}.$$

We also specify a faction-specific leadership premium, $\lambda_p = \pi^l/\pi$, which regulates the differential promotion probability when the paramount leader is from the princelings (e.g. Xi in the 18th Party Congress).

Results are reported in Table 4.12. The Vuong test indicates that the model where princelings are considered to be neutrals is preferred over one where princelings are treated as a separate faction. More importantly, the estimated leadership premium within the model imposing princelings as a faction, λ_p, is estimated to be less than 1. This means that, as princeling Xi reached the paramount position, other princelings did not appear to enjoy a higher premium in promotions. This finding prima facie violates one of the crucial features of factional politics – delivering resources to members of the faction once the faction leader is in power – and appears in stark contrast to what we have already observed for the broadly accepted factions, CYLC and Shanghai Gang, where we estimate $\lambda > 1$. In brief, the evidence rejects the hypothesis that princelings operate as a unified faction.

4.8.4 Is President Xi Jinping Affiliated to the Shanghai Gang?

Our structural approach allows also to produce formal tests for the analysis of factional affiliation of the top leadership. The case of Xi Jinping is emblematic because of both his strong ties to the CCP elite through family connections and his repeated rejection of intra-party factional politics (e.g. "cabals and cliques" mentioned in official transcripts on People's Daily, May 3rd, 2016[111]).

To this goal, we re-estimate the model assuming that Xi is an unaffiliated neutral, and compare the alternative model against our baseline specification where Xi is a Shanghai Gang member. The Vuong test shows that Xi is slightly more likely to be a Shanghai Gang member, although the statistical evidence is inconclusive. Our tests do not have enough power in this specific instance. Fortunately, such ambiguity is likely to be resolved after the 2017 19th Party Congress, which will unveil a wealth of data on new promotions within the CCP.

4.8.5 An Out-of-sample Forecast for the 2017 19th Party Congress

To conclude our quantitative exercises we employ our model to forecast the 19th Party Congress in 2017. Although admittedly speculative, to the best of our knowledge this is probably one of the very few rigorous quantitative environments allowing for an exercise of this kind. The model incorporates individual characteristics in this analysis to obtain more accurate forecasts[112].

The top panel of Table 4.13 shows that share of promotions by level of the Central Committee. Under the assumption that Xi is in fact a Shanghai Gang member, the Shanghai faction is expected to enjoy a higher share of promotions in the Politburo than the CYLC faction due to leadership premia. In contrast, promotions at lower levels are expected to be more comparable between the two factions due to the dampening effects stemming from vetoes.

Since there is still unresolved ambiguity regarding Xi's factional affiliation, we also conduct a forecast assuming Xi is a neutral in the bottom panel of Table 4.13. In this case, the Shanghai Gang would appear to lose its advantage in promotion for all the levels of the Central Committee.

---

[111]Available at http://en.people.cn/n3/2016/0503/c90000-9052676.html

[112]For individuals who newly enter AC at the 19th Party Congress whose characteristics are not readily available, we randomly draw the characteristics from the sample of the new entries of 18th Party Congress.

4.9 Conclusions

This paper contributes to an emerging literature on the political economy of economic development by focusing on elite organization in a nondemocracy. We specifically focus on modern China and on the internal organization of the Chinese Communist Party. The CCP, much like historical Leninist parties in Socialist countries, represents the linchpin of national politics and understanding its inner workings is central to any political economic analysis of the PRC.

We present a model of internal organization of this single-party regime, where explicit factional dynamics within the party enrich a problem of career concerns of political cadres. The model offers a series of novel insights on the role of factions in these regimes in a fully microfounded setting. Alternative modeling choices are also discussed.

The model is validated empirically employing a rich data set on the career profiles of top CCP members. In reduced form, a set of previously unexplored systematic empirical regularities in Chinese elite politics are probed and discussed. The extent of the 2012-2016 anti-corruption purge in shaping Chinese factional politics is also analyzed. In our structural estimation, we explore important counterfactuals pertinent to the Chinese historical case and use the model to answer a series of questions relevant to the political economy of the CCP. We hope that this framework may also prove useful to the understanding of the latent institutional shifts occurring within the CCP under General Secretary Xi.

In future research we hope to extend our analysis to the 2017 19th Party Congress. This will allow precision on all dimensions concerning the Xi Presidency.

Besides our application to Chinese politics, we plan to focus on similarly complex nondemocratic environments –the example of Russia comes to mind– where our model of hierarchical party organization may be to a certain extent transposable.

4.10 Proofs

For the proofs' notation we exclude time indexes unless necessary.

Proof of Proposition 1. Part (i). Suppose that, $I(\ell) > 0$, where $I \neq J$, and $N(\ell) > 0$. Consider the decision by a faction-J politician in a node at ℓ of whether to veto a cofactional's support for promotion to his node. With the promotion of a same-faction member from J to the politician's node, let $J^*(\ell)$ denote the total number of faction J members that would be present at level ℓ. Then, using equation (4.3) and (4.4), the promotion hazard parameter for this J politician at level ℓ (if the other faction K also vetoes co-faction members) if he does not veto becomes:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+i+\eta} \right),$$

with $j = \beta$ and $i = \rho$ or viceversa. If instead, the politician vetoes his cofactional, and a member of the other faction (or a neutral) ascends to his node, his promotion hazard becomes:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)-1} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+i+\eta} \right),$$

which is strictly greater.

If the other faction does not veto its members the respective expressions become:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+i+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+i+\eta} \right)$$

and

$$\delta_J(\ell) = \frac{j}{J^*(\ell)-1} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+i+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+i+\eta} \right).$$

And the latter hazard is clearly higher again. This proves part a).

Part (ii). Suppose that, $I(\ell) = 0$, and $N(\ell) > 0$. Suppose further that, with the promotion of a co-faction member to J's node there will be $J^*(\ell)$ members of J's faction at level ℓ, then the hazard parameter for promotion of this J politician is:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+\eta} \right). \tag{4.7}$$

If an N member instead ascends to his node, then the J member's promotion hazard is:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)-1} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{j+\eta} \right),$$

which exceeds (4.7), so he is clearly better off vetoing his own faction member.

However, if an $I \neq J, N$ ascends to his node, then the J member's promotion hazard becomes:[113]

$$\delta_J(\ell) = \frac{j}{J^*(\ell)-1} \left(I(\ell-1) \frac{(\delta + \delta_I^P(\ell-1))}{j+\eta} + N(\ell-1) \frac{(\delta + \delta_N^P(\ell-1))}{i+j+\eta} \right). \tag{4.8}$$

---

[113]Assuming that I's also veto. If they don't the sufficient condition is slightly altered, but qualitatively identical.

Since this I will contest $\ell - 1$ level openings (the second expression above), this lowers the chances of the J begin promoted to one of those. Assuming I factionals also veto, an I ascending to J's node lowers the chances of a J promotion the most if $N(\ell-1) = M(\ell-1)$. So a sufficient condition for J to exercise a veto assumes all higher positions are filled by members that are neutral, N. Under this assumption expression (4.8) becomes:

$$\delta_J(\ell) = \frac{j}{J^*(\ell) - 1}\left(M(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{i+j+\eta}\right),$$

and expression (4.7) becomes:

$$\delta_J(\ell) = \frac{j}{J^*(\ell)}\left(M(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{j+\eta}\right).$$

So J will veto a J coming from level $\ell+1$ provided that:

$$\frac{j}{J^*(\ell)}\left(M(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{j+\eta}\right) < \frac{j}{J^*(\ell)-1}\left(M(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{i+j+\eta}\right)$$

$$\Rightarrow \frac{J^*(\ell)-1}{J^*(\ell)} < \frac{j+\eta}{i+j+\eta}$$

$$\Rightarrow J(\ell) < \frac{j+\eta}{i}$$

where we use that $J^*(\ell) - 1 = J(\ell)$. ∎

Proof of Proposition 2. Let us define the indicator functions $I_B = 1$, iff $B(\ell) > 0$ and $I_B = 0$, otherwise; $I_N = 1$, iff $N(\ell) > 0$ and $I_N = 0$, otherwise; $I_R = 1$, iff $R(\ell) > 0$ and $I_R = 0$, otherwise.

Start with a neutral (N), who is at level ℓ in the hierarchy. $\delta_I(\ell-1)$ is determined from the hierarchy above:

$$\delta_N(\ell) = R(\ell-1)\left(\delta + \delta_R^p(\ell-1)\right)p_N^R(\ell)$$
$$+ N(\ell-1)\left(\delta + \delta_N^p(\ell-1)\right)p_N^N(\ell)$$
$$+ B(\ell-1)\left(\delta + \delta_B^p(\ell-1)\right)p_N^B(\ell).$$

Consider further that, differently from (4.4) where $p_N^R(\ell) = \eta/(I_B\beta + \eta + I_R\rho)$, now $p_N^R(\ell) = \eta/(I_B\beta + \eta)$ because in Proposition 1 each $R(\ell-1)$ is proven to veto any R possibly competing against N. For a similar reason, it holds that $p_N^B(\ell) = \eta/(I_R\rho + \eta)$.

We then have:

$$\delta_N(\ell)$$
$$= \frac{\eta}{N(\ell)}\left(R(\ell-1)\frac{\left(\delta + \delta_R^p(\ell-1)\right)}{I_B\beta + \eta} + N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{I_B\beta + I_R\rho + \eta} + B(\ell-1)\frac{\left(\delta + \delta_B^p(\ell-1)\right)}{I_R\rho + \eta}\right).$$

Similarly, for a faction B member this is given by:

$$\delta_B(\ell) = R(\ell-1)\left(\delta + \delta_R^p(\ell-1)\right)p_B^R(\ell) \tag{4.9}$$
$$+ N(\ell-1)\left(\delta + \delta_N^p(\ell-1)\right)p_B^N(\ell)$$
$$+ B(\ell-1)\left(\delta + \delta_B^p(\ell-1)\right)p_B^B(\ell)$$
$$= \frac{\beta}{B(\ell)}\left(R(\ell-1)\frac{\left(\delta + \delta_R^p(\ell-1)\right)}{\beta + I_N\eta} + N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\beta + I_R\rho + I_N\eta}\right),$$

where the last line uses the fact that vetoing from Proposition 1 implies $p_B^B(\ell) = 0$, while $p_B^R(\ell) = \beta / (\beta + I_N \eta)$ and $p_B^N(\ell) = \beta / (\beta + I_R \rho + I_N \eta)$.

Finally, for a faction R member this is:

$$
\begin{aligned}
\delta_R(\ell) &= R(\ell-1) \left(\delta + \delta_R^p(\ell-1) \right) p_R^R(\ell) \\
&\quad + N(\ell-1) \left(\delta + \delta_N^p(\ell-1) \right) p_R^N(\ell) \\
&\quad + B(\ell-1) \left(\delta + \delta_B^p(\ell-1) \right) p_R^B(\ell) \\
&= \frac{\rho}{R(\ell)} \left(B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1) \right)}{\rho + I_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{I_B \beta + \rho + I_N \eta} \right),
\end{aligned}
\tag{4.10}
$$

where the last line uses the fact that our vetoing results in Proposition 1 imply $p_R^R(\ell) = 0$, while $p_R^B(\ell) = \rho / (\rho + I_N \eta)$ and $p_R^N(\ell) = \rho / (\rho + I_B \beta + I_N \eta)$. ∎

Full Listing of $\delta_I(\ell)$ conditional on paramount leadership

For an N. If an N is paramount leader:

$$
\delta_N(\ell) = \frac{\eta}{N(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{I_B \beta + \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{I_B \beta + I_R \rho + \eta} + B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1) \right)}{I_R \rho + \eta} \right).
$$

If an R is paramount leader:

$$
\delta_N(\ell) = \frac{\eta}{N(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{I_B \beta + \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{I_B \beta + I_R \rho^l + \eta} + B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1) \right)}{I_R \rho^l + \eta} \right).
$$

If a B is paramount leader:

$$
\delta_N(\ell) = \frac{\eta}{N(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{I_B \beta^l + \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{I_B \beta^l + I_R \rho + \eta} + B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1) \right)}{I_R \rho + \eta} \right).
$$

where $I_B = 1$, iff $B(\ell) > 0$ and $I_B = 0$, otherwise; $I_R = 1$, iff $R(\ell) > 0$ and $I_R = 0$, otherwise.

For faction B member. If an N is paramount leader:

$$
\delta_B(\ell) = \frac{\beta}{B(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{\beta + I_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{\beta + I_R \rho + I_N \eta} \right).
$$

If an R is paramount leader:

$$
\delta_B(\ell) = \frac{\beta}{B(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{\beta + I_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{\beta + I_R \rho^l + I_N \eta} \right).
$$

If a B is paramount leader:

$$
\delta_B(\ell) = \frac{\beta^l}{B(\ell)} \left(R(\ell-1) \frac{\left(\delta + \delta_R^p(\ell-1) \right)}{\beta^l + I_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1) \right)}{\beta^l + I_R \rho + I_N \eta} \right).
$$

where $I_N = 1$, iff $N(\ell) > 0$ and $I_N = 0$, otherwise; $I_R = 1$, iff $R(\ell) > 0$ and $I_R = 0$, otherwise.

For a faction R member. If an N is paramount leader:

$$\delta_R(\ell) = \frac{\rho}{R(\ell)} \left(B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho + \mathtt{I}_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\mathtt{I}_B \beta + \rho + \mathtt{I}_N \eta} \right).$$

If an R is paramount leader:

$$\delta_R(\ell) = \frac{\rho^l}{R(\ell)} \left(B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho^l + \mathtt{I}_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\mathtt{I}_B \beta + \rho^l + \mathtt{I}_N \eta} \right).$$

If a B is paramount leader:

$$\delta_R(\ell) = \frac{\rho}{R(\ell)} \left(B(\ell-1) \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho + \mathtt{I}_N \eta} + N(\ell-1) \frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\mathtt{I}_B \beta^l + \rho + \mathtt{I}_N \eta} \right).$$

where $\mathtt{I}_N = 1$, iff $N(\ell) > 0$ and $\mathtt{I}_N = 0$, otherwise; $\mathtt{I}_B = 1$, iff $B(\ell) > 0$ and $\mathtt{I}_B = 0$, otherwise.

∎

Proof of Proposition 3. We first demonstrate that, if the system is stationary, so that $V_I^t(\ell) = V_I(\ell)$ and $\delta_I^t(\ell) = \delta_I(\ell)\ \forall I, \ell$, then $\delta_I(\ell) > \delta_J(\ell)$ implies $V_I(\ell) > V_J(\ell)$. So, (4.6) is solved by the I such that $\delta_I(\ell)$ is $\sup\{\delta_B(\ell), \delta_R(\ell), \delta_N(\ell)\}$.

The stationary analog of equation (4.5) where $V_I^t(\ell) = V_I(\ell)$ and $\delta_I^t(\ell) = \delta_I(\ell)\ \forall I, \ell$ is:

$$\delta V_I(\ell) = u(\ell) + \delta_I(\ell)\left[V_I(\ell-1) - V_I(\ell)\right],$$

which implies:

$$V_I(\ell) = \frac{u(\ell) + \delta_I(\ell) V_I(\ell-1)}{\delta + \delta_I(\ell)}$$

and

$$V_I(\ell-1) = \frac{u(\ell-1) + \delta_I(\ell-1) V_I(\ell-2)}{\delta + \delta_I(\ell-1)}.$$

By repeated substitution:

$$\begin{aligned}
V_I(\ell) =\ & \frac{u(\ell)}{\delta + \delta_I(\ell)} + \\
& \frac{\delta_I(\ell)\, u(\ell-1)}{\left(\delta + \delta_I(\ell)\right)\left(\delta + \delta_I(\ell-1)\right)} + \\
& \frac{\delta_I(\ell)\, \delta_I(\ell-1)\, u(\ell-2)}{\left(\delta + \delta_I(\ell)\right)\left(\delta + \delta_I(\ell-1)\right)\left(\delta + \delta_I(\ell-2)\right)} + \\
& \cdots + \frac{\delta_I(\ell)\, \delta_I(\ell-1)\cdots\delta_I(2)\, u(1)}{\left(\delta + \delta_I(\ell)\right)\left(\delta + \delta_I(\ell-1)\right)\cdots\left(\delta + \delta_I(1)\right)}.
\end{aligned}$$

This reduces to:

$$V_I(\ell) = \frac{u(\ell)}{\delta + \delta_I(\ell)} + \sum_{j=1}^{\ell-1} u(j) \times \frac{\prod_{k=j}^{k=\ell-1} \delta_I(k+1)}{\prod_{k=j}^{k=\ell}\left(\delta + \delta_I(k)\right)}.$$

Since flow payoffs are higher the higher the politician is in the hierarchy, i.e. $u(\ell-1) > u(\ell) \ \forall \ell$, then necessarily increasing the rate of promotion improves valuations, $\frac{dV_I(\ell)}{d\delta_I(\ell)} > 0 \ \forall \ell$. This implies that $\delta_I(\ell) > \delta_J(\ell)$ ensures $V_I(\ell) > V_J(\ell)$.

The proof proceeds next by establishing sufficient conditions for three parts. (i) The existence of neutrals given factions exist; (ii). The existence of a single faction given neutrals exist; (iii) The existence of a second faction, given neutrals and a first faction already exist.

In each part, a sufficient condition is provided for $\delta_I(\ell) > \delta_{J \neq I}(\ell)$ and $\delta_{K \neq I}(\ell)$ at a single level, ℓ. The sufficient condition established in each case is thus required to hold at all ℓ in order to ensure that an entering politician prefers entry as a type I.

Part (i). We establish a sufficient condition for there to be neutrals. Suppose, on the contrary, that there exist no N members. Necessarily, due to Proposition 1, without N's, all nodes will be filled by both a B and an R. Thus, under the supposition, the hierarchy remains stationary, so that, from the result above, it is sufficient to compute only the stationary $\delta_I(\ell)$ for each ℓ to determine the optimal I.

Assume, without loss of generality, that the paramount leadership position is held by a B. Consider level ℓ in the hierarchy. Necessarily the promotion hazard for an N at level ℓ is given by:

$$\delta_N(\ell) = \eta \left(R(\ell-1) \frac{(\delta + \delta_R^p(\ell-1))}{I_B \beta^l + \eta} + B(\ell-1) \frac{(\delta + \delta_B^p(\ell-1))}{I_R \rho + \eta} \right).$$

Due to optimal vetoes at each node, it must be that $R(\ell-1) = B(\ell-1) = M(\ell-1)/2$ and $I_B = I_R = 1$. The relationship between $\delta_R^p(\ell-1)$ and $\delta_B^p(\ell-1)$ is ambiguous. So consider both cases separately. First, assume that $\delta_R^p(\ell-1) \leq \delta_B^p(\ell-1)$, which will imply, due to the symmetry of the posited hierarchy, that $\delta_R^p(\ell) \leq \delta_B^p(\ell)$ too. Then, substituting for $I_B, I_R, R(\ell-1)$ and $B(\ell-1)$ yields:

$$\delta_N(\ell) = \eta \left(M(\ell-1)/2 \times \frac{(\delta + \delta_R^p(\ell-1))}{\beta^l + \eta} + M(\ell-1)/2 \times \frac{(\delta + \delta_B^p(\ell-1))}{\rho + \eta} \right).$$

Since $\delta_R^p(\ell-1) \leq \delta_B^p(\ell-1)$ then:

$$\delta_N(\ell) \geq \eta M(\ell-1)/2 \times (\delta + \delta_R^p(\ell-1)) \left(\frac{1}{\beta^l + \eta} + \frac{1}{\rho + \eta} \right),$$

and assuming, for now, that $\beta^l > \rho$ implies:

$$\delta_N(\ell) \geq \eta M(\ell-1) \times (\delta + \delta_R^p(\ell-1)) \left(\frac{1}{\beta^l + \eta} \right). \tag{4.11}$$

Now consider $\delta_B(\ell)$:

$$\delta_B(\ell) = \frac{2\beta^l}{M(\ell)} \left(M(\ell-1)/2 \times \frac{(\delta + \delta_R^p(\ell-1))}{\beta^l} \right)$$

$$= \frac{1}{M(\ell)} \left(M(\ell-1) \times (\delta + \delta_R^p(\ell-1)) \right).$$

Then $\delta_N(\ell) > \delta_B(\ell)$ if:

$$\eta M(\ell-1) \times (\delta + \delta_R^p(\ell-1)) \left(\frac{1}{\beta^L + \eta} \right) > \frac{1}{M(\ell)} \left(M(\ell-1) (\delta + \delta_R^p(\ell-1)) \right),$$

which rearranges to:

$$\frac{\eta}{\beta^l + \eta} > \frac{1}{M(\ell)}. \tag{4.12}$$

Now consider $\delta_R(\ell)$:

$$\delta_R(\ell) = \frac{2\rho}{M(\ell)} \left(M(\ell-1)/2 \times \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho} \right)$$

$$= \frac{1}{M(\ell)} \left(M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) \right).$$

Since, by supposition, $\delta_R^p(\ell-1) \leq \delta_B^p(\ell-1)$ it is possible to define $Z \geq 1$ such that $\frac{\delta + \delta_B^p(\ell)}{\delta + \delta_R^p(\ell)} \equiv Z$. Note that Z is invariant with respect to $M(\ell)$. To see why, note that with a symmetric hierarchy in which each node is filled by a B and R pair we have: $\delta_R^p(\ell) = \delta_B(\ell)$ and $\delta_B^p(\ell) = \delta_R(\ell)$. Thus, using equations (4.9) and (4.10) and the fact that in such a hierarchy $R(\ell) = B(\ell) = M(\ell)/2$, we have

$$\delta_R(\ell) = \left(M(\ell-1) \times \frac{\left(\delta + \delta_R(\ell-1)\right)}{M(\ell)} \right)$$

and

$$\delta_B(\ell) = \left(M(\ell-1) \times \frac{\left(\delta + \delta_B(\ell-1)\right)}{M(\ell)} \right).$$

So the ratio $\frac{\delta_B^p(\ell)}{\delta_R^p(\ell)} = \frac{\delta_R(\ell)}{\delta_B(\ell)} = \frac{M(\ell-1) \times \frac{(\delta + \delta_R(\ell-1))}{M(\ell)}}{M(\ell-1) \times \frac{(\delta + \delta_B(\ell-1))}{M(\ell)}} = \frac{\delta + \delta_R(\ell-1)}{\delta + \delta_B(\ell-1)}$, which is clearly independent of $M(\ell)$. Using the notation Z we then have:

$$\delta_R(\ell) = \frac{1}{M(\ell)} \left(M(\ell-1) \times \left(\delta + \delta_R^p(\ell-1)\right) Z \right).$$

Then $\delta_N(\ell) > \delta_R(\ell)$ if:

$$\eta M(\ell-1) \times \left(\delta + \delta_R^p(\ell-1)\right) \left(\frac{1}{\beta^L + \eta} \right) > \frac{1}{M(\ell)} \left(M(\ell-1) \left(\delta + \delta_R^p(\ell-1)\right) Z \right),$$

which rearranges to:

$$\frac{\eta}{\beta^l + \eta} > \frac{Z}{M(\ell)}. \tag{4.13}$$

which again holds for $M(\ell)$ large enough at all ℓ. So for sufficiently large $M(\ell)$, neutrals will be the preferred entering type, thus contradicting the maintained assumption that neutrals are not in the hierarchy. Assuming, alternatively, that $\beta^l \leq \rho$, instead of using the inequality in (4.11) we now have:

$$\delta_N(\ell) \geq \eta M(\ell-1) \times \left(\delta + \delta_R^p(\ell-1)\right) \left(\frac{1}{\rho + \eta} \right),$$

which, by following the same procedure as above, yields the analog to (4.12) as a sufficient condition for $\delta_N(\ell) > \delta_B(\ell)$, namely:

$$\frac{\eta}{\rho + \eta} > \frac{1}{M(\ell)}. \tag{4.14}$$

This again holds for sufficiently high $M(\ell)$, and again will hold for sufficiently high $M(\ell)$ for the R entrants subject to the scaling by factor Z. Again, entering politicians will choose to be neutral.

Now suppose the alternative relationship between $\delta_R^p(\ell-1)$ and $\delta_B^p(\ell-1)$, that is: $\delta_R^p(\ell-1) > \delta_B^p(\ell-1)$, and again first posit that $\beta^l > \rho$. Then let us use these two inequalities and substitute for $I_B, I_R, R(\ell-1)$ and $B(\ell-1)$ exactly as we did above. Equation (4.11) now yields:

$$\delta_N(\ell) = \eta \left(M(\ell-1)/2 \times \frac{\left(\delta + \delta_R^p(\ell-1)\right)}{\beta^l + \eta} + M(\ell-1)/2 \times \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho + \eta} \right)$$

$$> \eta M(\ell-1)/2 \times \left(\delta + \delta_B^p(\ell-1)\right) \left(\frac{1}{\beta^l + \eta} + \frac{1}{\rho + \eta} \right)$$

$$> \eta M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) \left(\frac{1}{\beta^l + \eta} \right).$$

Now $\delta_R(\ell)$ is given by:

$$\delta_R(\ell) = \frac{2\rho}{M(\ell)} \left(M(\ell-1)/2 \times \frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho} \right)$$

$$= \frac{1}{M(\ell)} \left(M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) \right).$$

Then $\delta_N(\ell) > \delta_R(\ell)$ if:

$$\eta M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) \left(\frac{1}{\beta^L + \eta} \right) > \frac{1}{M(\ell)} \left(M(\ell-1) \left(\delta + \delta_B^p(\ell-1)\right) \right).$$

A sufficient condition for this is:

$$\frac{\eta}{\beta^l + \eta} > \frac{1}{M(\ell)}.$$

This again holds for $M(\ell)$ high enough.

Now $\delta_B(\ell)$ is given by:

$$\delta_B(\ell) = \frac{2\beta}{M(\ell)} \left(M(\ell-1)/2 \times \frac{\left(\delta + \delta_R^p(\ell-1)\right)}{\beta} \right)$$

$$= \frac{1}{M(\ell)} \left(M(\ell-1) \times \left(\delta + \delta_R^p(\ell-1)\right) \right).$$

Since, by supposition it is now the case that, $\delta_B^p(\ell-1) \leq \delta_R^p(\ell-1)$ it is possible to define $K \geq 1$ such that $\frac{\delta + \delta_R^p(\ell)}{\delta + \delta_B^p(\ell)} \equiv K$. Similarly to the above, K is invariant with respect to $M(\ell)$. Substituting for K we have:

$$\delta_B(\ell) = \frac{1}{M(\ell)} \left(M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) K \right).$$

Then $\delta_N(\ell) > \delta_B(\ell)$ if:

$$\eta M(\ell-1) \times \left(\delta + \delta_B^p(\ell-1)\right) \left(\frac{1}{\beta^L + \eta} \right) > \frac{1}{M(\ell)} \left(M(\ell-1) \left(\delta + \delta_B^p(\ell-1)\right) K \right).$$

A sufficient condition for this is:

$$\frac{\eta}{\beta^l + \eta} > \frac{K}{M(\ell)}.$$

This again holds for $M(\ell)$ high enough. So new entrants will prefer to enter as neutrals over either faction for $M(\ell)$ large enough.

The analogous procedure under the alternative assumption $\beta^l \leq \rho$ yields a sufficient condition exactly as in (4.14):

$$\frac{\eta}{\rho + \eta} > \frac{1}{M(\ell)}.$$

Part (ii). We now establish a sufficient condition for there to exist at least a single faction. Suppose that all positions in the hierarchy are held by a neutral. Consider an entrant choosing to also be a neutral. In that case under the supposition, the system is again stationary and we have:

$$\delta_N(\ell) = \frac{N(\ell-1)}{M(\ell)} \left(\delta + \delta_N^p(\ell-1) \right).$$

But by entering as a B member the entrant would have:

$$\delta_B(\ell) = \beta N(\ell-1) \left(\frac{\delta + \delta_N^p(\ell-1)}{\beta + \eta} \right).$$

These rearrange to imply that $\delta_B(\ell) > \delta_N(\ell)$ provided that $M(\ell) > \frac{\beta + \eta}{\beta}$. The analogous sufficient condition for an R entrant is $M(\ell) > \frac{\rho + \eta}{\rho}$. This proves part (ii).

Part (iii). We establish a sufficient condition for two factions to exist. We proceed as above, by demonstrating a contradiction. If there is only one faction present, without loss of generality let it be B, and the other politicians are N, for sufficiently high $M(\ell)$, $\delta_R(\ell) > \delta_B(\ell)$ or $\delta_N(\ell)$, so that an entering politician will choose to enter as an R.

As previously, with only N and B in the hierarchy we have:

$$\delta_N(\ell) = \frac{\eta}{M(\ell) - B(\ell)} \left(N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\beta + \eta} + B(\ell-1)\frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\eta} \right), \qquad (4.15)$$

$$\delta_B(\ell) = \frac{\beta}{B(\ell)} \left(N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\beta + \eta} \right).$$

Either $\delta_N(\ell) > \delta_B(\ell)$, so that a new entrant would prefer to enter as an N over a B, or the converse, in which case he would choose to enter as a B over an N. Suppose first that $\delta_N(\ell) > \delta_B(\ell)$ and consider the promotion hazard for a single entering R:

$$\delta_R(\ell) = \rho \left(B(\ell-1)\frac{\left(\delta + \delta_B^p(\ell-1)\right)}{\rho + \eta} + N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\rho + \beta + \eta} \right).$$

If $\delta_N(\ell) > \delta_B(\ell)$ for an increase in $M(\ell)$, then necessarily the term $M(\ell) - B(\ell)$ in expression (4.15) increases with $M(\ell)$, since an extra politician would enter as an N instead of a B. But since $\delta_R(\ell)$ above is independent of $M(\ell)$, there exists an $M(\ell)$ sufficiently high so that $\delta_R(\ell) > \delta_N(\ell)$, and an entering politician would instead choose to be an R over being an N, contradicting the posited non-existence of R members in equilibrium.

Alternatively, suppose that $\delta_N(\ell) \leq \delta_B(\ell)$, then, for an increase in $M(\ell)$ necessarily the term $M(\ell) - N(\ell)$ increases with $M(\ell)$, as a politician would choose to enter as a B over being an N. Now consider the promotion hazard for a B:

$$\delta_B(\ell) = \frac{\beta}{M(\ell) - N(\ell)} \left(N(\ell-1)\frac{\left(\delta + \delta_N^p(\ell-1)\right)}{\beta + \eta} \right).$$

Again, since $\delta_R(\ell)$ is independent of $M(\ell)$, there exists an $M(\ell)$ high enough so that $\delta_R(\ell) > \delta_B(\ell)$, which implies that a new entrant will choose to enter as an R member, again contradicting the posited non-existence of R members. ∎

4.11 Tables and Figures
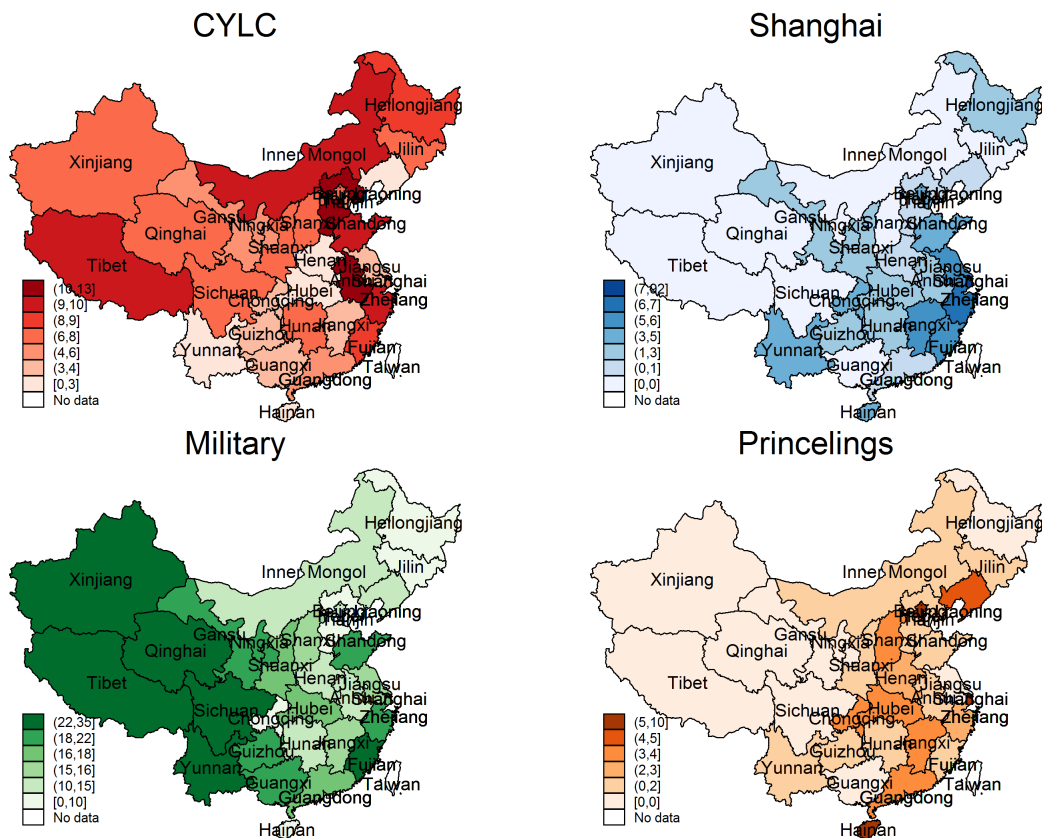
Figure 4.1: Geographic Distribution of Factions or Groups (1956-2014)

Notes: This graph shows the geographic distribution of factions or groups across provinces (municipalities) over the period of 1956 to 2014. The color scale represents the average share of faction or group in a province (municipality).
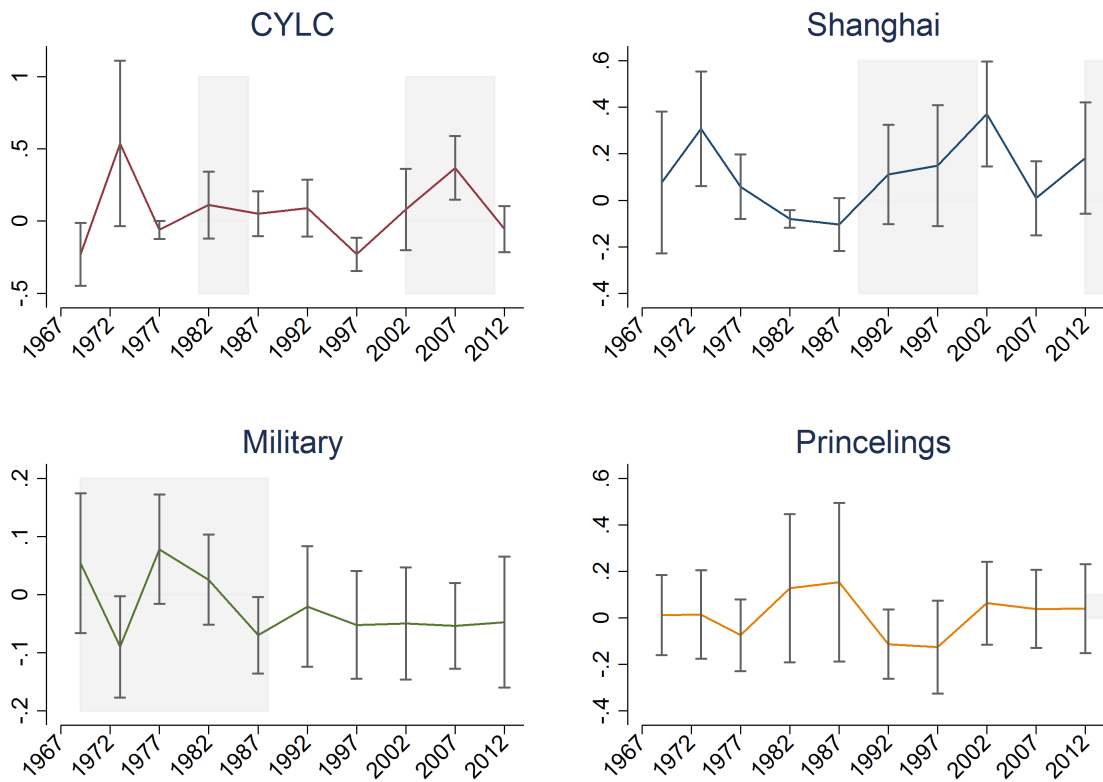
Figure 4.2: Leadership Premium in Promotion Rates of Each Faction or Group

Notes: This graph shows the leadership premium in promotion rates of each faction over the rest of members in the Central Committee over time. The leadership premium in promotion rates is defined as the regression coefficients of promotion dummy on faction or group affiliation. The regression is repeated for each session of Central Committee. The capped spikes indicate the standard errors of the estimates. The shaded area indicates that the General Secretary of CCP is from the same faction or group.
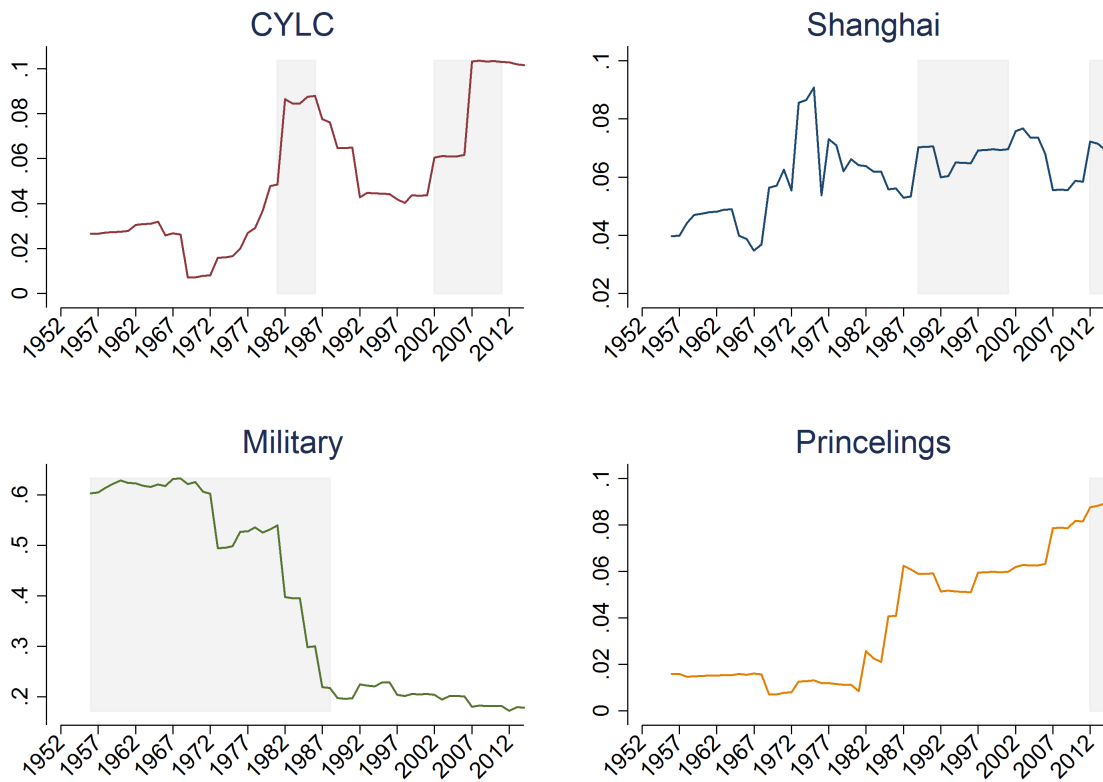
Figure 4.3: Leadership Premium in Power Score of Each Faction or Group

Notes: This graph shows the share of power score of each faction or group in the Central Committee over time. The power score is constructed following the scheme of Bo (2010). The shaded area indicates that the General Secretary of CCP is from the same faction or group.
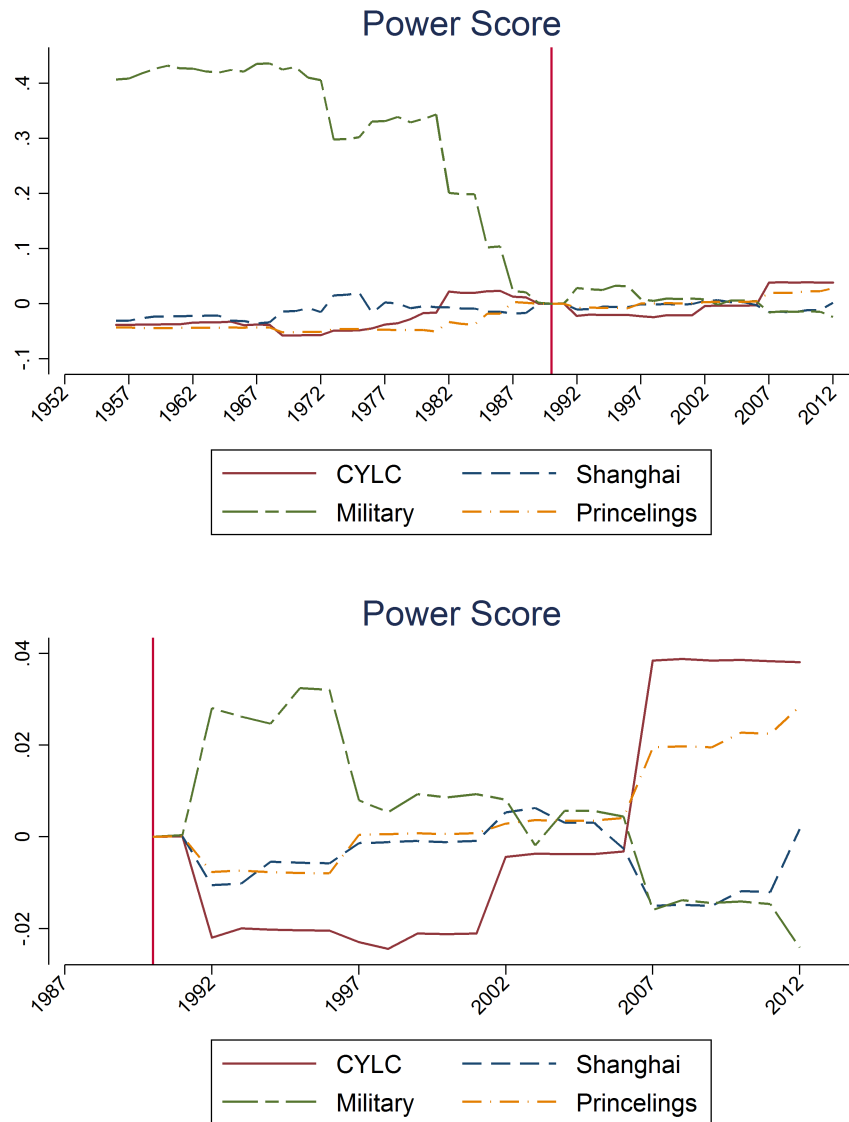
Figure 4.4: Power Score of Each Faction or Group in the Central Committee

Notes: This graph shows the share of power score of each faction or group in the Central Committee over time. The power score is constructed following the scheme of Bo (2010). The vertical line indicates the year of 1990, the first time when a civilian, Jiang Zemin, took over the Central Military Committee. The power score is normalized to zero in 1990. The upper panel shows the whole sample period from 1956 to 2012, the lower panel shows the post-Deng period from 1990 to 2012.
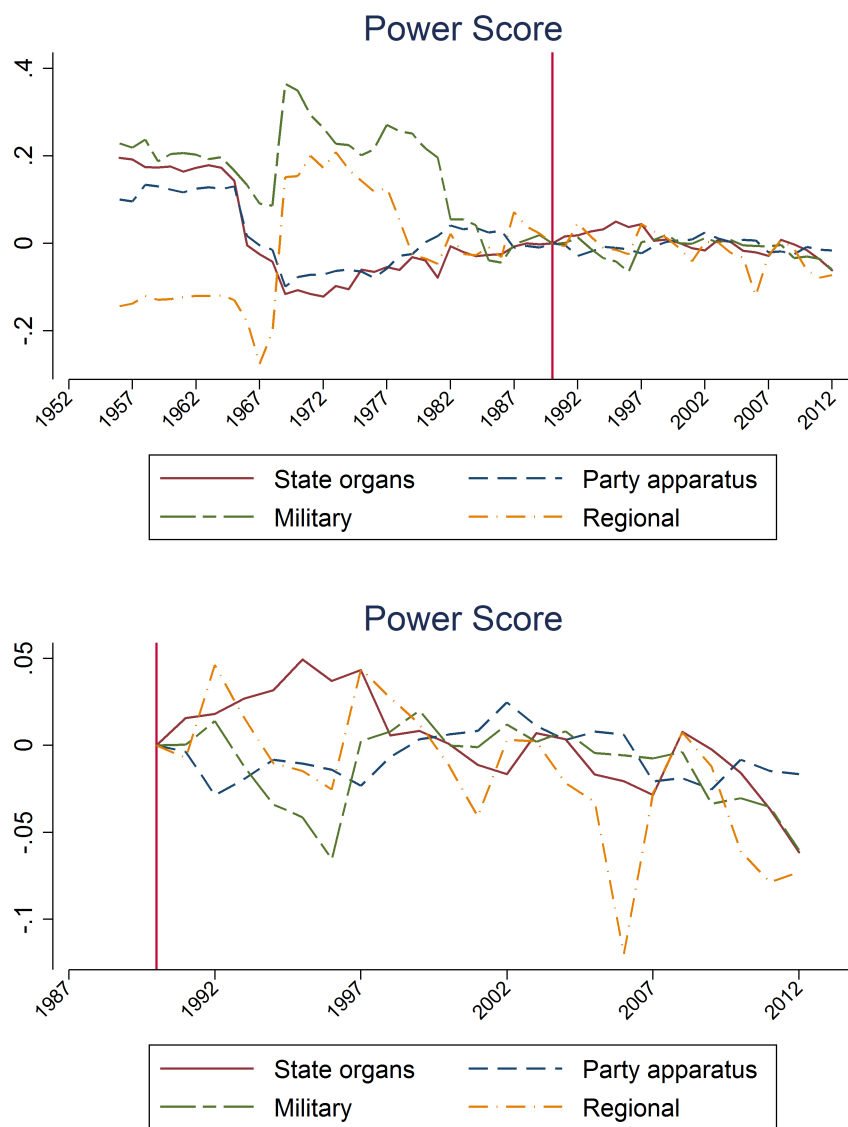
Figure 4.5: Power Score of Each Constituency in the Central Committee

Notes: This graph shows the share of power score for each constituency in the Central Committee over time. The power score is constructed following the scheme of Bo (2010). The vertical line indicates the year of 1990, the first time when a civilian, Jiang Zemin, took over the Central Military Committee. The power score is normalized to zero in 1990. The upper panel shows the whole sample period from 1956 to 2012, the lower panel shows the post-Deng period from 1990 to 2012.
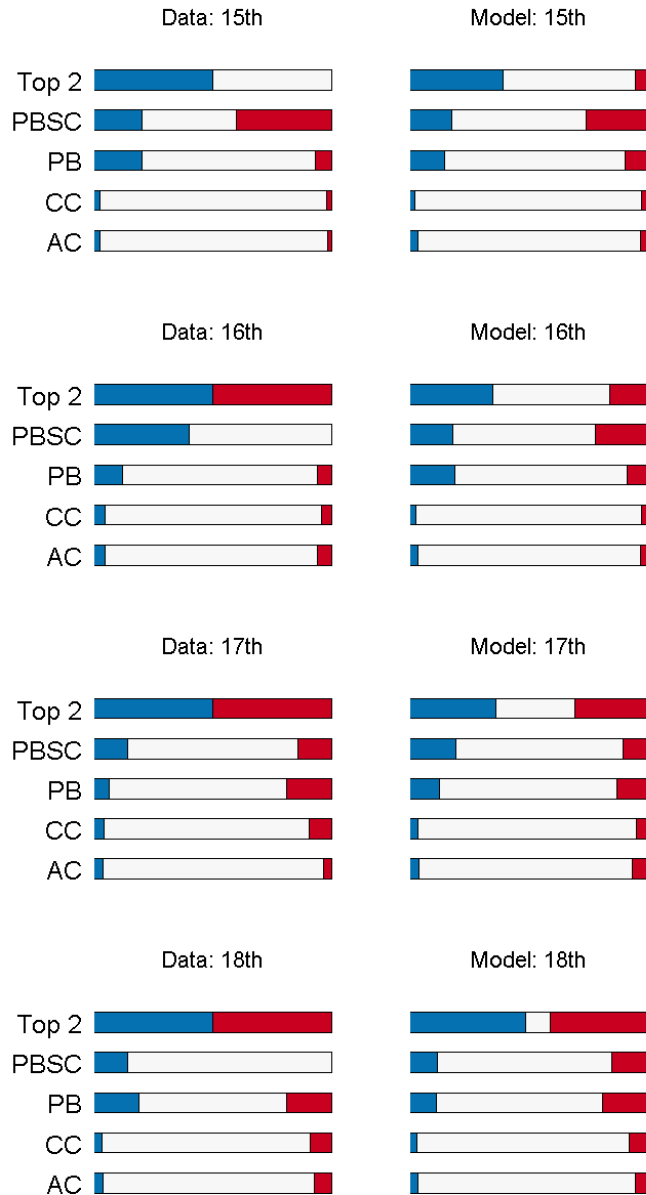
Figure 4.6: Seat Shares at Each Level of the Central Committee

Notes: This graph shows seat shares at each level of the Central Committee predicted by the baseline faction model and in the data. Each of the five bars represents the top two CCP positions, PBSC, PB, CC, and AC, from the top down, respectively. The blue/white/red bar represents the Shanghai Gang/Neutral/CYCL. The model is estimated using the 14th to 18th Central Committees and the results are averaged over 100 simulations for each Party Congress.
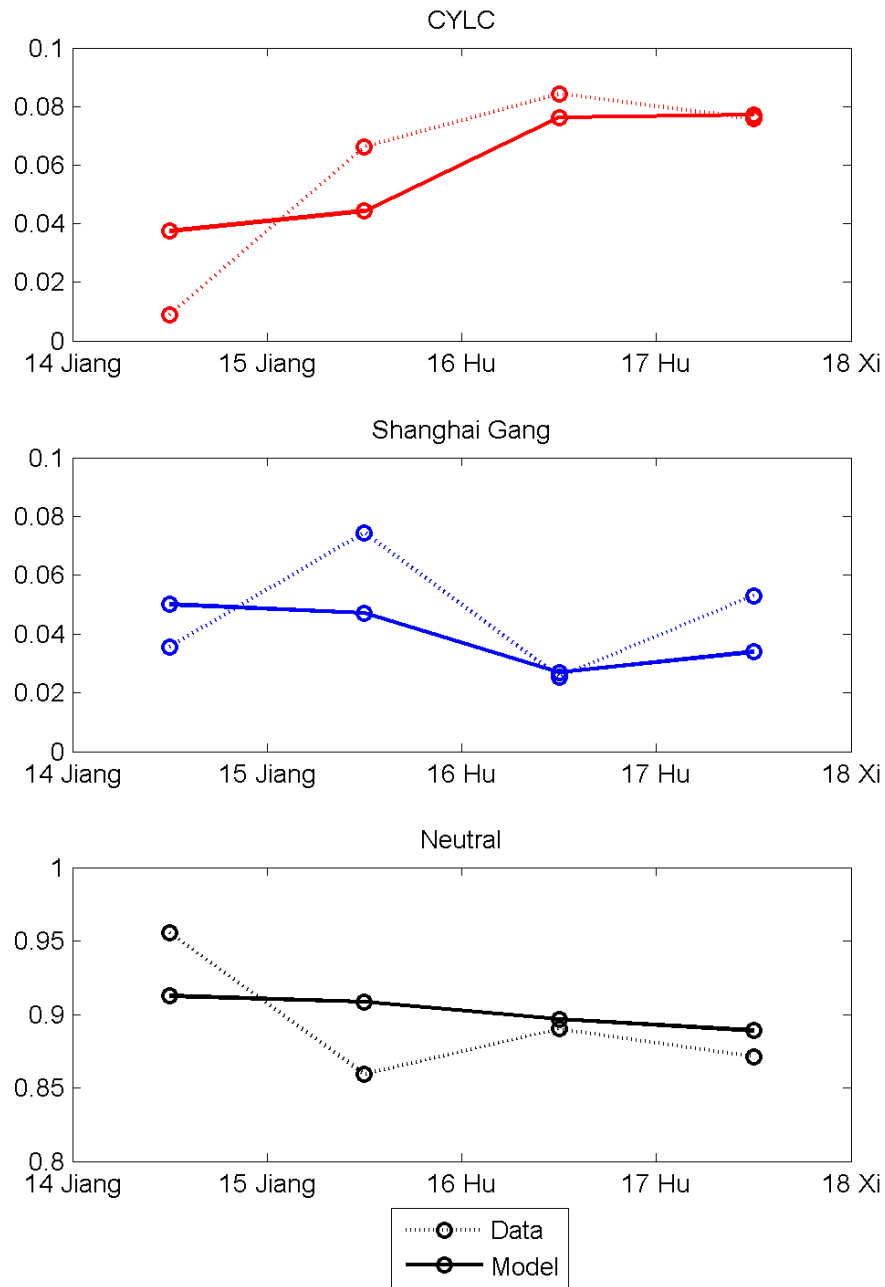
Figure 4.7: Aggregate Share of Promotions over Time

Notes: This graph shows the time series plot of the share of promotions of each faction over time. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The share of promotions is predicted by the baseline faction model estimated using the 14th to 18th Central Committees and the results are averaged over 100 simulations for each Party Congress.
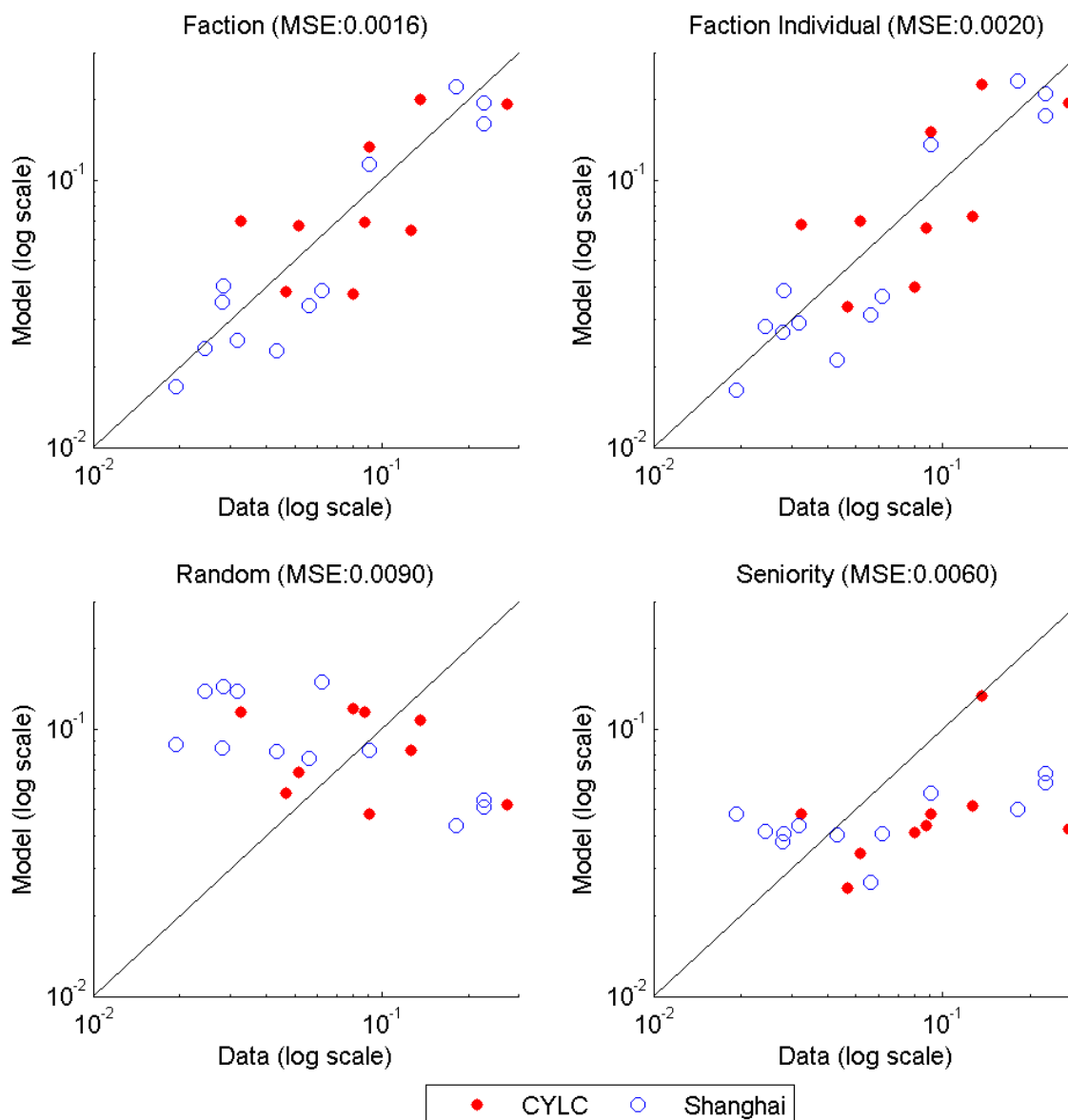
Figure 4.8: Model Fit (In Sample)

Notes: This graph shows the scatter plot of the model predicted share of promotions of each faction against the data. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The blue/red dot represents Shanghai Gang/CYLC. Each dot is a share of a faction at a given level of a given Party Congress. The estimation sample includes the 14th to 18th Central Committees and the results are averaged over 100 simulations for each Party Congress.
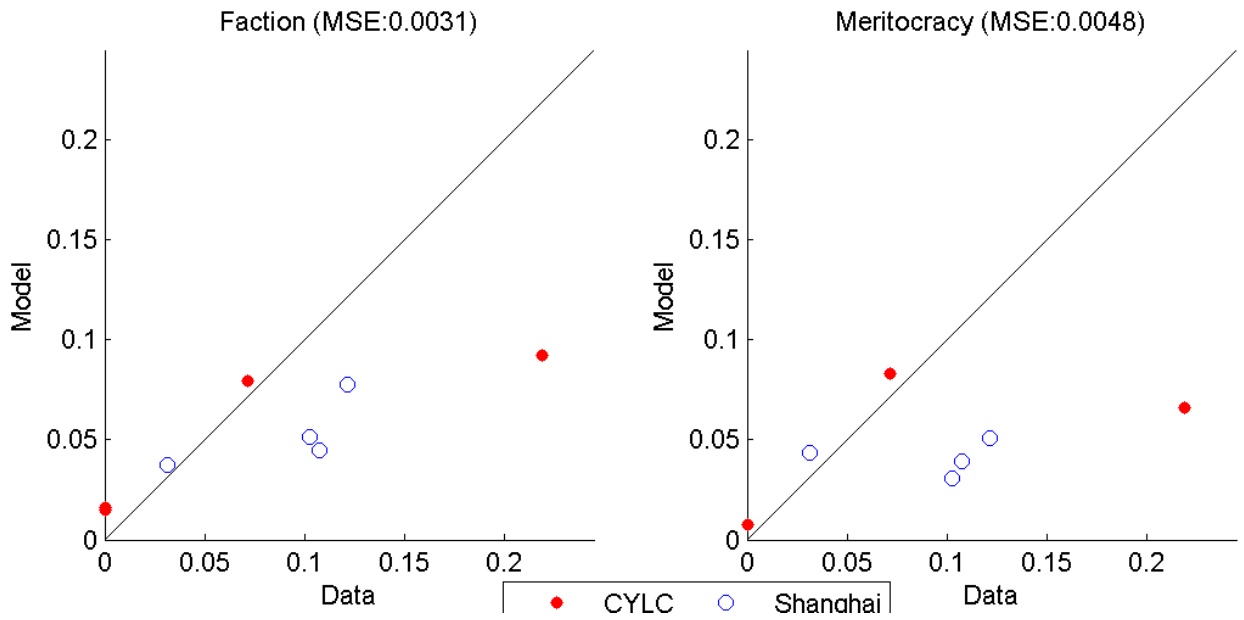
Figure 4.9: Meritocracy (In Sample)

Notes: This graph shows the scatter plot of the model predicted share of promotions of each faction against the data. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions. The blue/red dot represents Shanghai Gang/CYLC. Each dot is a share of a faction at a given level of a given Party Congress. The estimation sample includes the 14th to 18th Central Committees and the results are averaged over 100 simulations for each Party Congress.
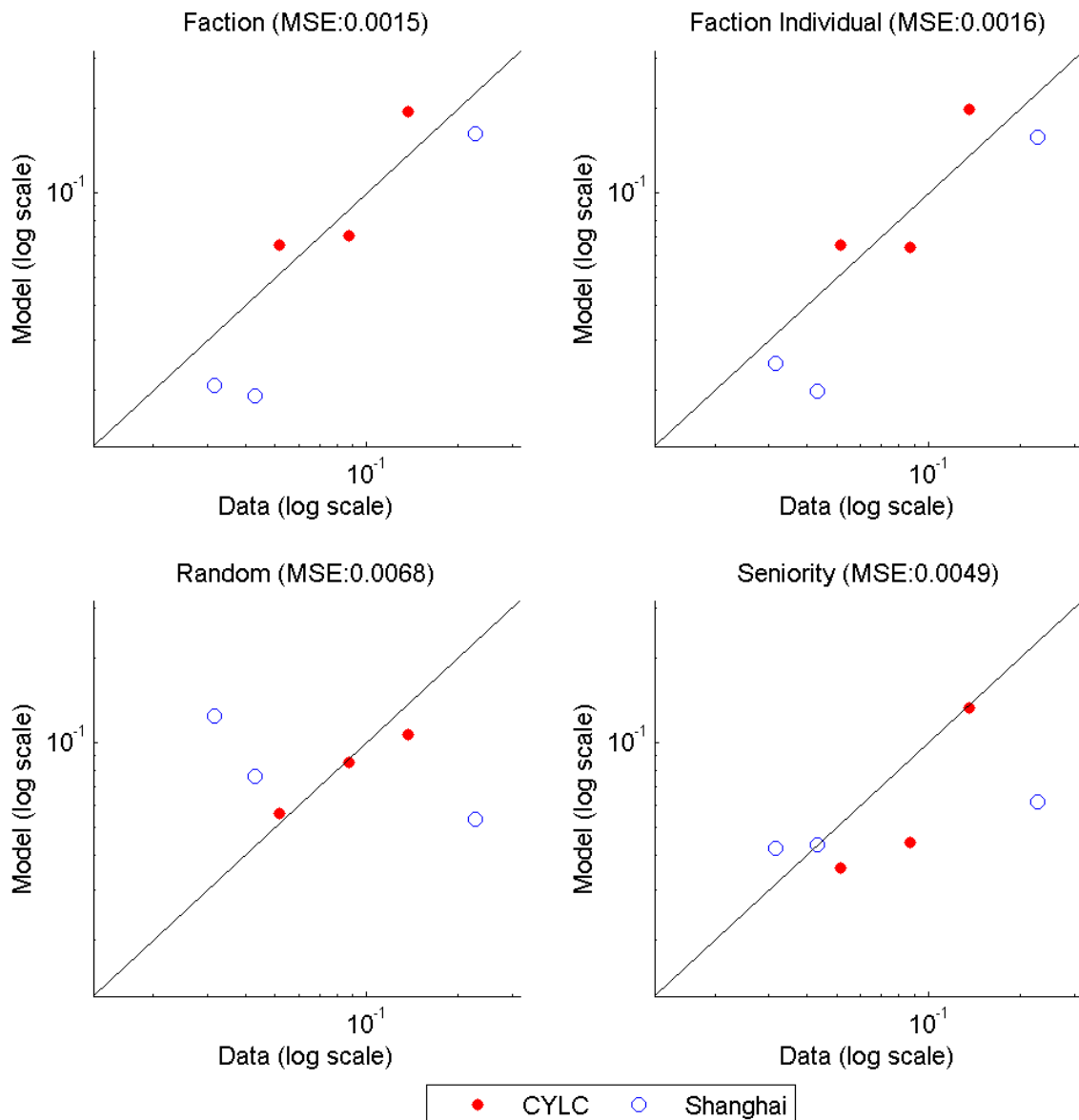
Figure 4.10: Model Fit (Out of Sample)

Notes: This graph shows the scatter plot of the model predicted share of promotions of each faction against the data. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The blue/red dot represents Shanghai Gang/CYLC. Each dot is a share of a faction at a given level of the 18th party congress. The estimation sample includes the 14th to 17th Central Committees and the results are averaged over 100 simulations for each Party Congress.
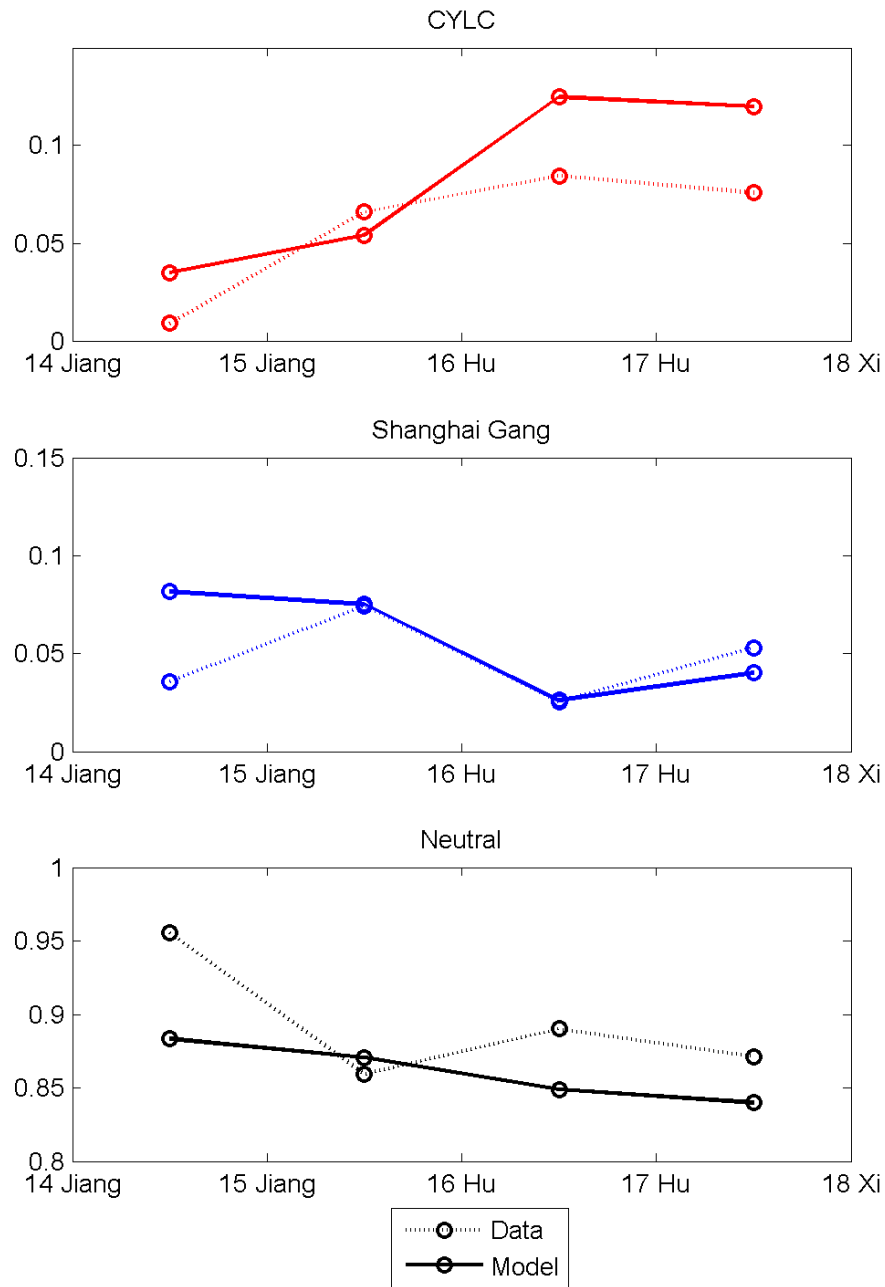
Figure 4.11: Counterfactual Aggregate Share of Promotions over Time (Leadership Premium × 2)
Notes: These graphs show the time series plot of the share of promotions of each faction over time. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The counterfactual simulations are conducted by doubling the leadership premium of the baseline faction model and the results are averaged over 100 simulations for each Party Congress.
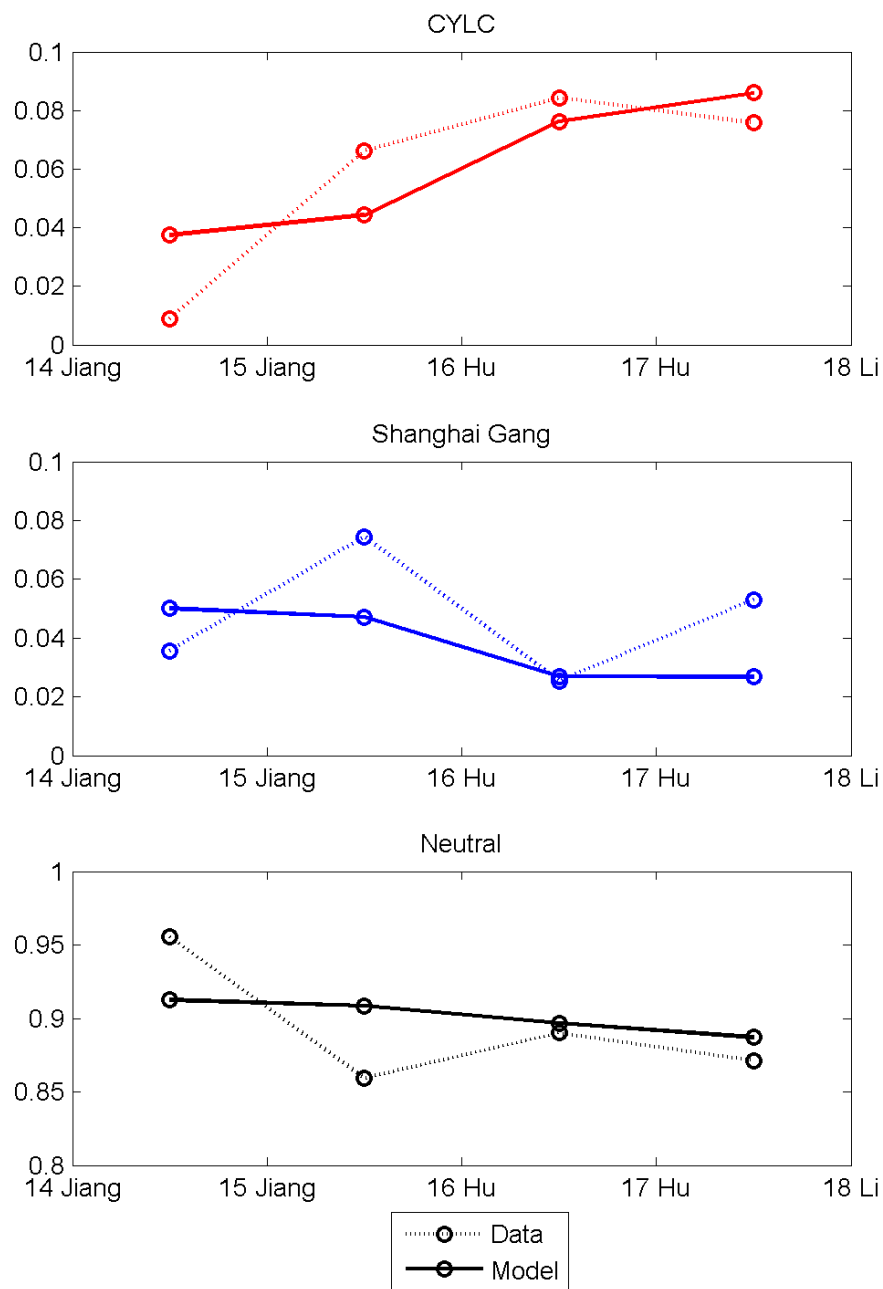
Figure 4.12: Counterfactual Aggregate Share of Promotions over Time (Li Keqiang Presidency)
Notes: These graphs show the time series plot of the share of promotions of each faction over time. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The counterfactual simulations are conducted by assuming Li Keqiang became the president in the 18th Party Congress and the results are averaged over 100 simulations for each Party Congress.

Table 4.1: Summary Statistics of Elites in China

Organizations	N	Duration	Age	Gender	Ethnicity	CYLC	Shanghai	Military	Princelings
Party Apparatus	10543	4.47	47.97	0.93	0.10	0.13	0.04	0.05	0.03
		(2.98)	(10.17)	(0.26)	(0.3)	(0.34)	(0.2)	(0.21)	(0.18)
Government	7099	3.98	46.88	0.93	0.05	0.07	0.03	0.01	0.04
		(12.54)	(9.71)	(0.25)	(0.22)	(0.25)	(0.17)	(0.09)	(0.19)
Military	2091	4.07	44.10	0.99	0.02	0.01	0.02	0.91	0.06
		(3.86)	(14.98)	(0.1)	(0.15)	(0.12)	(0.13)	(0.29)	(0.23)
People's Congress	1696	5.34	56.43	0.89	0.13	0.07	0.03	0.04	0.02
		(3.23)	(8.94)	(0.31)	(0.34)	(0.26)	(0.17)	(0.2)	(0.14)
CPPCC	1413	6.95	59.72	0.90	0.12	0.06	0.04	0.01	0.02
		(52.9)	(9.91)	(0.29)	(0.33)	(0.23)	(0.18)	(0.12)	(0.14)
Court	213	4.46	37.16	0.97	0.22	0.08	0.02	0.00	0.00
		(3.87)	(8.59)	(0.17)	(0.41)	(0.26)	(0.14)	(0)	(0)
Procuratorate	53	5.34	46.51	0.92	0.10	0.11	0.08	0.00	0.00
		(3.16)	(11.89)	(0.27)	(0.31)	(0.32)	(0.27)	(0)	(0)
CYLC	521	3.85	31.91	0.87	0.12	0.77	0.01	0.00	0.03
		(2.88)	(7.14)	(0.33)	(0.32)	(0.42)	(0.11)	(0.06)	(0.16)
Business	4590	4.27	42.99	0.93	0.04	0.04	0.04	0.01	0.02
		(4.62)	(13.66)	(0.25)	(0.2)	(0.19)	(0.19)	(0.09)	(0.14)
Media	500	4.72	40.74	0.98	0.06	0.10	0.04	0.01	0.03
		(4.61)	(11.73)	(0.15)	(0.24)	(0.3)	(0.2)	(0.12)	(0.18)
Education	3781	2.88	34.04	0.92	0.03	0.04	0.03	0.02	0.02
		(4.23)	(12.38)	(0.26)	(0.16)	(0.21)	(0.17)	(0.14)	(0.13)
Unclassified	3558	3.94	40.00	0.90	0.07	0.06	0.03	0.03	0.03
		(4.39)	(13.58)	(0.3)	(0.26)	(0.24)	(0.17)	(0.16)	(0.16)

Notes: This table shows summary statistics of demographics and career paths of 4,494 elites who hold import positions in government, politics, the military, education, business, and media in China since 1992. The unit of observation is position-individual pair. We report means and standard deviation, in parentheses below. N is the number of observations in each type of organization. Duration is the length of tenure in the position. Age is the age when an individual first started the job. Gender equals 1 if an individual is male, 0 otherwise. Ethnicity equals 1 if a member is from an ethnic minority, 0 otherwise. CYLC/Shanghai/Military/Princelings equals 1 if an individual is from CYLC/Shanghai/Military/Princelings faction/group, 0 otherwise. The data source for this table is China Vitae.

Table 4.2: Summary Statistics of Central Committee Members

C.C.	Year	N	Gender	Age	College	Graduate	Mishu	Ethnicity	Abroad	Promotion	Retirement	CYLC	Shanghai	Military	Princelings
8	1956	173	0.95	51.23	0.40	0.00	0.34	0.03	0.07	0.14	0.51	0.03	0.05	0.56	0.02
			(0.21)	(6.97)	(0.49)	(0)	(0.48)	(0.17)	(0.25)	(0.35)	(0.5)	(0.18)	(0.21)	(0.5)	(0.13)
9	1969	278	0.92	51.82	0.33	0.01	0.09	0.03	0.07	0.11	0.26	0.01	0.05	0.56	0.01
			(0.26)	(11.71)	(0.47)	(0.1)	(0.29)	(0.16)	(0.25)	(0.31)	(0.44)	(0.1)	(0.23)	(0.5)	(0.08)
10	1973	318	0.88	52.36	0.36	0.01	0.09	0.02	0.08	0.09	0.42	0.02	0.07	0.44	0.01
			(0.32)	(12.94)	(0.48)	(0.1)	(0.29)	(0.15)	(0.27)	(0.29)	(0.49)	(0.12)	(0.25)	(0.5)	(0.11)
11	1977	331	0.90	56.88	0.38	0.01	0.06	0.02	0.08	0.07	0.62	0.03	0.06	0.44	0.02
			(0.3)	(11.95)	(0.49)	(0.09)	(0.24)	(0.14)	(0.27)	(0.26)	(0.49)	(0.17)	(0.23)	(0.5)	(0.12)
12	1982	344	0.93	59.61	0.58	0.04	0.12	0.04	0.09	0.13	0.62	0.08	0.06	0.33	0.02
			(0.25)	(8.08)	(0.49)	(0.2)	(0.32)	(0.2)	(0.29)	(0.34)	(0.49)	(0.26)	(0.23)	(0.47)	(0.15)
13	1987	285	0.93	56.11	0.79	0.09	0.15	0.05	0.11	0.15	0.40	0.05	0.06	0.21	0.05
			(0.26)	(6.55)	(0.41)	(0.28)	(0.36)	(0.22)	(0.32)	(0.36)	(0.49)	(0.22)	(0.23)	(0.41)	(0.22)
14	1992	318	0.92	56.87	0.88	0.11	0.13	0.07	0.10	0.18	0.45	0.03	0.03	0.21	0.04
			(0.26)	(6.18)	(0.32)	(0.32)	(0.34)	(0.25)	(0.31)	(0.38)	(0.5)	(0.16)	(0.18)	(0.41)	(0.19)
15	1997	343	0.93	56.51	0.95	0.17	0.11	0.06	0.11	0.18	0.48	0.02	0.04	0.20	0.05
			(0.26)	(5.72)	(0.22)	(0.38)	(0.31)	(0.23)	(0.31)	(0.39)	(0.5)	(0.15)	(0.19)	(0.4)	(0.22)
16	2002	356	0.93	56.05	0.98	0.31	0.07	0.07	0.10	0.16	0.47	0.05	0.06	0.19	0.05
			(0.26)	(5.33)	(0.15)	(0.46)	(0.26)	(0.26)	(0.3)	(0.36)	(0.5)	(0.22)	(0.23)	(0.39)	(0.23)
17	2007	366	0.90	56.15	0.87	0.52	0.10	0.07	0.11	0.21	0.48	0.07	0.04	0.17	0.06
			(0.29)	(5.68)	(0.34)	(0.5)	(0.29)	(0.25)	(0.31)	(0.41)	(0.5)	(0.26)	(0.2)	(0.37)	(0.23)
18	2012	373	0.91	56.50	0.87	0.68	0.17	0.07	0.10	0.00	0.00	0.09	0.05	0.17	0.05
			(0.29)	(4.73)	(0.34)	(0.47)	(0.38)	(0.26)	(0.3)	(0)	(0)	(0.29)	(0.21)	(0.37)	(0.23)

Notes: This table shows summary statistics of the members of the 8th -18th Central Committees. We report the mean and the standard deviation, below in parentheses. Gender equals 1 if a member is male, 0 otherwise. College equals 1 if a member has a college degree, 0 otherwise. Graduate equals 1 if a member has a post-graduate degree, 0 otherwise. Abroad equals 1 if a member has studied or worked abroad, 0 otherwise. Mishu equals 1 if a member has been worked as a personal secretary of prominent politicians, 0 otherwise. Ethnicity equals 1 if a member is an ethnic minority, 0 otherwise. Promotion equals to 1 if a member will be promoted in the next session of Central Committee, 0 otherwise. Retirement equals to 1 if a member will retire after the current session of Central Committee, 0 otherwise. CYLC/Shanghai/Military/Princelings equals 1 if a member is from CYLC/Shanghai/Military/Princelings faction/group, 0 otherwise.

Table 4.3: Geographical Distribution of Factions and Groups

Dependent Variable: Average Share of Faction or Group				
	(1)	(2)	(3)	(4)
	Shanghai	CYLC	Military	Princelings
GDP per capita	0.644**	-0.652	-2.141***	1.517***
	[0.265]	[0.623]	[0.741]	[0.319]
Constant	1.705***	7.309***	19.97***	0.693*
	[0.533]	[0.915]	[1.875]	[0.374]
Observations	30	31	31	31
Adjusted R-squared	0.040	0.011	0.053	0.396

Notes: This table shows the cross-section regressions of the share of each faction in provinces (municipalities) on the average provincial (municipal) GDP per capita over the period of 1956-2014. The share of a faction in a province is defined as the ratio of the number of faction members who have worked in this province (municipality) over the total number of central committee members who have worked in the same place during their careers. Robust standard errors are reported in the bracket. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.4: Factional Mix

	(1) All	(2) All	(3) All	(4) All	(5) Provincial	(6) National
Dependent Variable:			CYLC1			
CYLC2	-0.139**	-0.185***	-0.189**	-0.245***	-0.136*	-0.499**
	[0.0568]	[0.0594]	[0.0755]	[0.0723]	[0.0693]	[0.143]
Year F.E.	N	Y	N	Y	Y	Y
Position F.E.	N	N	Y	Y	Y	Y
Observations	794	794	794	794	648	145
Adjusted R-squared	0.016	0.070	0.193	0.254	0.242	0.180
Dependent Variable:			Shanghai1			
Shanghai2	-0.105***	-0.132***	-0.353*	-0.378**	-0.0319	-0.802*
	[0.0319]	[0.0346]	[0.180]	[0.175]	[0.0466]	[0.341]
Year F.E.	N	Y	N	Y	Y	Y
Position F.E.	N	N	Y	Y	Y	Y
Observations	773	773	773	773	627	145
Adjusted R-squared	0.006	0.011	0.382	0.392	0.187	0.278
Dependent Variable:			Princelings1			
Princelings2	-0.0535	-0.0595	-0.132**	-0.134**	-0.155*	-0.0411
	[0.0505]	[0.0523]	[0.0571]	[0.0545]	[0.0806]	[0.114]
Year F.E.	N	Y	N	Y	Y	Y
Position F.E.	N	N	Y	Y	Y	Y
Observations	794	794	794	794	648	145
Adjusted R-squared	0.001	0.020	0.133	0.154	0.202	0.227

Notes: This table shows panel regressions of the factional affiliation of the number 1 official on the number 2 official in the same political office. The top/middle/bottom panel shows results for CYLC/Shanghai/princelings respectively. Variable CYLC1 (CYLC2) is a dummy which equals to 1 if number 1 (2) official is from the CYLC faction. Shanghai1, Shanghai2, Princelings1 and Princelings2 and defined similarly. Column 1-4 include all positions, and Column 5-6 break down to provincial and national level positions. The provincial positions include 31 provincial and municipal units (secretary and governor). The position in Shanghai Municipality is excluded in the regression sample for Shanghai Gang. The national positions include Politburo Standing Committee (two highest ranking members), PRC presidency (President and Vice President), the State Council (Premier and Executive Vice premier), Central Military Committee (Chairman and Executive Vice Chairman), CCP Secretariat (two highest ranking secretaries), NPC (Chairman and Executive Vice Chairman), CPPCC (Chairman and Executive Vice Chairman), the Supreme People's Court (President and Executive Vice President). Standard errors are clustered at both position unit and year level. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.5: Factional Mix (Shanghai vs. CYCL)

	(1) All	(2) All	(3) All	(4) All	(5) Provincial	(6) National
Dependent Variable: Shanghai1						
CYLC2	0.164**	0.166*	0.187**	0.196**	0.00169	0.779***
	[0.0810]	[0.0818]	[0.0747]	[0.0761]	[0.0241]	[0.169]
Year F.E.	N	Y	N	Y	Y	Y
Position F.E.	N	N	Y	Y	Y	Y
Observations	773	773	773	773	627	145
Adjusted R-squared	0.046	0.047	0.376	0.381	0.186	0.489
Dependent Variable: CYCL1						
Shanghai2	0.368*	0.315	0.396*	0.338*	0.101	0.758*
	[0.195]	[0.197]	[0.207]	[0.197]	[0.207]	[0.357]
Year F.E.	N	Y	N	Y	Y	Y
Position F.E.	N	N	Y	Y	Y	Y
Observations	773	773	773	773	627	145
Adjusted R-squared	0.043	0.069	0.207	0.239	0.232	0.200

Notes: This table shows panel regressions of the factional affiliation of the number 1 official on the number 2 official in the same political office. Variable Shanghai1 (Shanghai2) is a dummy which equals to 1 if number 1 (2) official is from the Shanghai faction. Variable CYLC1 (CYLC2) is a dummy which equals to 1 if number 1 (2) official is from the CYLC faction. The sample period is from 1992 to 2014. Column 1-4 include all positions, and Column 5-6 break down to provincial and national level positions. The provincial positions include 31 provincial and municipal units (secretary and governor) excluding Shanghai Municipality. The national positions include Politburo Standing Committee (two highest ranking members), PRC presidency (President and Vice President), the State Council (Premier and Executive Vice premier), Central Military Committee (Chairman and Executive Vice Chairman), CCP Secretariat (two highest ranking secretaries), NPC (Chairman and Executive Vice Chairman), CPPCC (Chairman and Executive Vice Chairman), the Supreme People's Court (President and Executive Vice President). Standard errors are clustered at both position unit and year level. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.6: Leadership Premia in Promotion and Retirement

	Promotion		Retirement	
	(1)	(2)	(1)	(2)
CYLC	0.0397	0.0299	-0.111**	-0.132***
	[0.0450]	[0.0456]	[0.0439]	[0.0430]
CYLC*CYLC Secretary	0.206**	0.242**	-0.0797	-0.101
	[0.0943]	[0.0959]	[0.0818]	[0.0836]
Shanghai	0.0144	0.0281	-0.0353	-0.0614
	[0.0371]	[0.0373]	[0.0493]	[0.0498]
Shanghai*Shanghai Secretary	0.193***	0.170**	-0.0394	0.0212
	[0.0717]	[0.0727]	[0.0724]	[0.0737]
Princelings	0.0294	0.0368	-0.120**	-0.106**
	[0.0471]	[0.0468]	[0.0489]	[0.0484]
Princelings*Princelings Secretary	0.0158	-0.0125	-0.0161	-0.0772
	[0.101]	[0.103]	[0.112]	[0.116]
Military	-0.0414**	-0.0392**	0.0229	0.0160
	[0.0185]	[0.0190]	[0.0280]	[0.0287]
Military*Military Secretary	-0.0239	-0.0313	-0.109***	-0.0465
	[0.0207]	[0.0262]	[0.0324]	[0.0392]
Controls	Y	Y	Y	Y
Year F.E.	N	Y	N	Y
P-value (CYLC*CYLC Secretary=Shanghai*Shanghai Secretary)	0.8275	0.5902	0.7131	0.283
Observations	2998	2998	3113	3113
Adjusted R-squared	0.066	0.068	0.121	0.155

Notes: This table shows panel regressions of promotion and retirement indicators on the faction or group affiliation of Central Committee members interacting with the affiliation of the General Secretary. The sample includes all the members of the 8th to 18th Central Committees, except Politburo Standing Committee members are excluded from the promotion regressions. Promotion is a dummy which equals to 1 if a Central Committee member moves up in the rank defined by the four levels of Central Committee (1 PBSC, 2 PB, 3 CC, and 4 AC), 0 otherwise. Retirement is a dummy which equals to 1 if a Central Committee member retires from the Central Committee, 0 otherwise. Robust standard errors are reported in brackets. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.7: Leadership Premia in Power Score and Seat Shares

	CYLC						Shanghai				
	(1)	(2)	(3)	(4)	(5)		(1)	(2)	(3)	(4)	(5)
	Power score	AC seats	CC seats	PB seats	PBSC seats		Power score	AC seats	CC seats	PB seats	PBSC seats
Secretary	0.0420***	0.0233**	0.0340***	0.0525	0.0955*		0.0105***	-0.00802	-0.0103*	0.0786***	0.195***
	[0.00876]	[0.0102]	[0.00995]	[0.0327]	[0.0555]		[0.00382]	[0.00739]	[0.00587]	[0.0206]	[0.0398]
Observations	59	59	59	59	59		59	59	59	59	59
Adjusted R-squared	0.390	0.212	0.274	0.072	0.100		0.139	0.032	0.076	0.382	0.394

	Military						Princelings				
	(1)	(2)	(3)	(4)	(5)		(1)	(2)	(3)	(4)	(5)
	Power score	AC seats	CC seats	PB seats	PBSC seats		Power score	AC seats	CC seats	PB seats	PBSC seats
Secretary	0.274***	0.259***	0.231***	0.410***	0.476***		0.0516***	0.0178***	0.0271***	0.169***	0.361***
	[0.0695]	[0.0724]	[0.0623]	[0.0813]	[0.0821]		[0.00784]	[0.00484]	[0.00744]	[0.0234]	[0.0243]
Observations	59	59	59	59	59		59	59	59	59	59
Adjusted R-squared	0.533	0.541	0.485	0.583	0.558		0.165	0.046	0.044	0.179	0.465

Notes: This table shows regressions of the power scores and seat shares of each faction or group on the affiliation of the General Secretary. The dependent variables are the power score (Score), the share of Alternate Central Committee members (AC), the share of full Central Committee members (CC), the share of Politburo members (PB), and the share of Politburo Standing Committee members (PBSC). The independent variable Secretary is a dummy which equals to 1 if the General Secretary is from the same faction, 0 otherwise. The top left panel (column 1-5) reports the results for the CYLC faction. The top right panel (column 6-10) reports the results for the Shanghai faction. The bottom left panel (column 1-5) reports the results for the Military group, the bottom right panel (column 6-10) reports the results for the Princeling group. The sample period is from 1956 to 2014. Newey-West standard errors with 5 lags are reported in brackets. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.8: Anticorruption and Factional Affiliation

Dependent Variable	Corruption
CYLC	0.0272
	[0.0237]
Shanghai	-0.0229
	[0.0237]
Princelings	0.0189
	[0.0321]
Gender	0.0139
	[0.0167]
Ethnicity	-0.0191
	[0.0168]
AC	-0.0350**
	[0.0136]
CC	-0.00920
	[0.0129]
PB	0.0125
	[0.0407]
PBSC	0.0328
	[0.0583]
age	-0.00596***
	[0.000649]
Observations	2240
Adjusted R-squared	0.032

Notes: This table shows the cross-sectional regression of a corruption dummy on the faction or group affiliation of an official. Corruption is defined as 1 if the official is investigated or prosecuted according to ChinaFile and the China's Central Commission for Discipline Inspection (CCDI) website, and 0 otherwise. The sample includes all the individuals except military personnel covered by China Vitae who have not retired in the year of 2007, the year of 17th Party Congress. Robust standard errors are reported in brackets. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively.

Table 4.9: Parameter Estimates of the Faction Model

	(1)		(2)		(3)		(4)
ρ	0.045***	ρ_H	0.162**	ρ	0.042***	ρ_H	0.153**
	[0.008]		[0.063]		[0.009]		[0.062]
β	0.029***	β_H	0.193***	β	0.033***	β_H	0.241**
	[0.006]		[0.068]		[0.010]		[0.103]
λ	2.553***	ρ_L	0.041***	λ_R	2.791***	ρ_L	0.037***
	[0.511]		[0.007]		[0.720]		[0.008]
		β_L	0.022***	λ_B	2.178***	β_L	0.027***
			[0.005]		[0.758]		[0.009]
		λ	2.526***			λ_R	2.898***
			[0.514]				[0.761]
						λ_B	1.956***
							[0.699]
Log-LL	-2766		-2747		-2766		-2746
Log-LLR	-19.305		-0.378		-19.142		-
P-value	0.000		0.385		0.000		-

Notes: This table shows the parameter estimates of the faction model for different specifications. The sample includes all the members of the 14th to 18th Central Committees. Standard errors are reported in brackets. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively. The bottom panel shows log-likelihood, log-likelihood ratio, and p-value of the log-likelihood ratio tests for each specification against model (4) as the alternative hypothesis. The estimator employs 100 simulations for each Party Congress.

Table 4.10: Parameter Estimates of Alternative Models

	Baseline Faction Model		Faction with Individual Characteristics		Random		Seniority
ρ_H	0.162** [0.063]	ρ_H	0.174** [0.069]	ρ entry	0.043	ρ entry	0.043
β_H	0.193*** [0.068]	β_H	0.201*** [0.072]	β entry	0.043	β entry	0.043
ρ_L	0.041*** [0.007]	ρ_L	0.043*** [0.008]			Age1	0.464*** [0.105]
β_L	0.022*** [0.005]	β_L	0.023*** [0.005]			Age2	-1.213*** [0.127]
λ	2.526*** [0.514]	λ	2.390*** [0.531]			Age3	-0.428*** [0.050]
		Princeling	0.413** [0.202]				
		Military	0.129 [0.122]				
		College	-0.152 [0.164]				
		Graduate	-0.222* [0.119]				
		Minority	-0.813*** [0.208]				
		Gender	0.926*** [0.237]				
		Age1	0.361*** [0.109]				
		Age2	-1.201*** [0.136]				
		Age3	-0.421*** [0.055]				
Log-LL	-2747		-2617		-2763		-2660
Log-LLR	-129.976		-		-		-
P-value	0.000		-		-		-
Vuong	-		-		-13.429		-7.026
P-value	-		-		0.000		0.000

Notes: This table shows the parameter estimates of four alternative models of CCP promotion dynamics. The sample includes all the members of the 14th to 18th Central Committees. The probability of entry for seniority and random model is calibrated using the mean faction shares in the sample. Standard errors are reported in brackets. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively. The estimator employs 100 simulations for each Party Congress. The bottom panel shows log-likelihood, log-likelihood ratio, p-value of the log-likelihood ratio tests, Vuong test statistics, and the p-value of the Vuong tests for each model against the model "faction with individual characteristics" column as the alternative hypothesis.

Table 4.11: Share of promotion of Each Faction by Level of the Central Committee

		Data			Baseline Faction Model			Counterfactual (Leadership Premium×2)			Counterfactual (Li Keqiang Presidency)		
		B	N	R	B	N	R	B	N	R	B	N	R
14th													
	PB	18.18%	81.82%	0.00%	22.41%	68.25%	9.33%	33.76%	59.30%	6.95%	22.41%	68.25%	9.33%
	CC	2.80%	96.26%	0.93%	3.48%	93.19%	3.33%	6.09%	90.59%	3.32%	3.48%	93.19%	3.33%
	AC	2.83%	96.23%	0.94%	4.01%	92.65%	3.34%	6.60%	90.24%	3.16%	4.01%	92.65%	3.34%
15th													
	PB	22.73%	68.18%	9.09%	19.42%	67.28%	13.30%	32.43%	53.14%	14.43%	19.42%	67.28%	13.30%
	CC	5.61%	89.72%	4.67%	3.40%	92.77%	3.83%	6.16%	89.29%	4.55%	3.40%	92.77%	3.83%
	AC	6.19%	85.84%	7.96%	3.85%	92.40%	3.75%	5.60%	89.51%	4.89%	3.85%	92.40%	3.75%
16th													
	PB	9.09%	63.64%	27.27%	11.50%	69.25%	19.25%	9.96%	61.59%	28.45%	11.50%	69.25%	19.25%
	CC	1.94%	85.44%	12.62%	1.68%	91.83%	6.49%	1.81%	86.49%	11.70%	1.68%	91.83%	6.49%
	AC	2.44%	94.31%	3.25%	2.34%	90.61%	7.05%	2.27%	86.57%	11.16%	2.34%	90.61%	7.05%
17th													
	PB	22.73%	63.64%	13.64%	16.27%	63.70%	20.03%	16.70%	54.83%	28.47%	14.28%	63.40%	22.32%
	CC	4.31%	90.52%	5.17%	2.31%	90.94%	6.75%	3.27%	85.34%	11.39%	1.82%	90.54%	7.64%
	AC	3.17%	88.10%	8.73%	2.53%	90.52%	6.95%	3.02%	86.51%	10.47%	1.90%	90.46%	7.64%
18th													

Notes: This table shows the share of promotions of each faction by level of the Central Committee in the data and predicted by the different models. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The sample includes all the members of the 14th to 18th Central Committees. The first panel shows the share of promotions of each faction in the data. The second panel shows the prediction by the baseline faction model. The third panel shows the counterfactual prediction in which the leadership premium is doubled comparing to the baseline faction model. The last panel shows the counterfactual prediction in which Li Keqiang becomes President in the 18th Party Congress.

Table 4.12: Tests of Xi's Factional Affiliation

	Baseline Faction Model		Princelings as Faction		Xi as Neutral
ρ_H	0.162**	ρ_H	0.178**	ρ_H	0.164**
	[0.063]		[0.074]		[0.064]
β_H	0.193***	β_H	0.153**	β_H	0.195***
	[0.068]		[0.067]		[0.069]
ρ_L	0.041***	π_H	0.364***	ρ_L	0.044***
	[0.007]		[0.124]		[0.008]
β_L	0.022***	ρ_L	0.050***	β_L	0.027***
	[0.005]		[0.009]		[0.006]
λ	2.526***	β_L	0.027***	λ	2.150***
	[0.514]		[0.006]		[0.437]
		π_L	0.059***		
			[0.010]		
		λ	1.876***		
			[0.394]		
		λ_p	0.564		
			[0.358]		
Log LL	-2747		-2866		-2748
Vuong	-		-15.850		-0.197
P-value	-		0.000		0.422

Notes: This table shows the parameter estimates of three models of CCP promotion dynamics. The sample includes all the members of the 14th to 18th Central Committees. Standard errors are reported in brackets. The estimator employs 100 simulations for each Party Congress. ***,**,* indicates 1 percent, 5 percent, and 10 percent significance level respectively. The bottom panel shows log-likelihood, Vuong test statistics, and the p-value of the Vuong tests for each model against the baseline faction model as the alternative hypothesis.

Table 4.13: Out-of-sample Forecast of 19th Central Committee

	B	N	R
Xi as Shanghai Gang			
PB	24.18%	66.37%	9.45%
CC	3.84%	92.72%	3.44%
AC	4.35%	91.52%	4.13%
Xi as Neutral			
PB	14.12%	75.53%	10.35%
CC	2.20%	94.03%	3.77%
AC	2.28%	93.77%	3.95%

Notes: This table shows the aggregate share of promotions of each faction at each level of the Central Committee in the 19th Central Committee predicted by the faction model with individual characteristics. The share of promotions is defined as the ratio between the number of promotions of a faction and the total number of promotions to this level. The sample used to estimate the parameters includes all the members of the 14th to 18th Central Committees. The forecast employs 100 simulations for this Party Congress.

Chapter 5

Conclusion

This thesis studies theoretical mechanisms and empirical consequences of government actions on financial market in order to better understand the organization of the financial sector and the inner working of the governments.

This first essay documents a new channel of monetary policy transmission through the shadow banking system. Analyzing U.S. money supply data from 1987 to 2012, I find that shadow bank deposits expand significantly when the Federal Reserve tightens monetary policy. This channel partially offsets the reduction of commercial bank deposits and dampens the impact of monetary tightening. I construct a structural model of bank competition and show that this new channel is a result of deposit competition between commercial and shadow banks in a market with heterogeneous depositors. Facing more yield-sensitive clientele, shadow banks pass through more rate hikes to depositors during periods of monetary tightening, thereby poaching deposits from commercial banks. Fitting my model to institution-level data from both commercial banks and money market funds, I show that the shadow bank channel reduces the impact of monetary policy on money supply by 40 percent. My results suggest a cautious stance towards the use of monetary tightening as a tool for promoting financial stability, because monetary tightening may unintentionally drive more deposits into the uninsured shadow banking sector, thereby amplifying the risk of bank runs.

The second essay examines the effects of the post-crisis financial regulations, encompassing the Dodd-Frank Act and Basel III, on market liquidity of the U.S. fixed income market. We estimate structural breaks in a large panel of liquidity measures of corporate and Treasury bonds. Our methodology does not require a priori knowledge of the exact timing of breaks, can capture not only sudden jumps but also breaks in slow-moving trends, and displays excellent power properties in presence of confounding factors. Against the popular claim that post-crisis regulations hurt liquidity, we find no evidence of liquidity deterioration during periods of regulatory intervention. Instead, breaks towards higher liquidity are often identified. We discuss their connection to the post-crisis rise in agency trading.

The third essay investigates theoretically and empirically the factional arrangements and dynamics within the Chinese Communist Party (CCP), the governing political party of the People's Republic of China. Our empirical analysis ranges from the end of the Deng Xiaoping era to the current Xi Jinping presidency and covers the appointments of both national and provincial officials. We present a set of new empirical regularities within the CCP and a theoretical framework suited to model factional politics within single-party regimes.

Future work In the future work, I plan to extend my thesis in several dimensions. For instance, the first essay studies monetary transmission in a partial equilibrium setting. It would be very interesting to extend the model to a general equilibrium framework where I can study the interaction between market power and monetary policy on firm investments and aggregate output. The second direction is to apply the structural IO framework to study other financial industry such as mutual funds. Previous literature on mutual fund usually assumes a representative investor which has perfectly elastic demand for mutual fund investment. A direct result of such assumption is that the representative investor will only care about abnormal returns generated by mutual fund managers. However, in practice, mutual fund investors could have very heterogeneous preference over factor loadings of different mutual funds. This may generate market segmentations and different competitive environment for different mutual funds. It would be very interesting to explore the implications of such framework.

Bibliography

Acharya, Viral V., Philipp Schnabl, and Gustavo Suarez, 2013, Securitization without risk transfer, *Journal of Financial Economics* 107, 515–536.

Adams, Robert M, Kenneth P Brevoors, and Elizabeth K Kiser, 2007, Who competes with whom? the case of depository institutions, *Journal of Industrial Economics* 141–167.

Adrian, Tobias, and Hyun Song Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418–437.

Ahn, Seung C, and Alex R Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203–1227.

Ajello, Andrea, Thomas Laubach, David López-Salido, and Taisuke Nakata, 2015, Financial stability and optimal interest-rate policy, *Working Paper* .

Amato, Jeffery D, and Eli M Remolona, 2003, The credit spread puzzle, *BIS Quarterly Review, December* .

Amihud, Yakov, and Haim Mendelson, 1980, Dealership market: Market-making with inventory, *Journal of Financial Economics* 8, 31–53.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of financial Economics* 17, 223–249.

Andrews, Donald WK, 1994, Empirical process methods in econometrics, *Handbook of econometrics* 4, 2247–2294.

Aragon, George O, and Philip E Strahan, 2012, Hedge funds as liquidity providers: Evidence from the lehman bankruptcy, *Journal of Financial Economics* 103, 570–587.

Bai, Chong-En, Chang-Tai Hsieh, and Zheng Michael Song, 2014, Crony capitalism with chinese characteristics, *University of Chicago, working paper* .

Bai, Jushan, and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191–221.

Bai, Jushan, and Pierre Perron, 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 47–78.

Bai, Jushan, and Pierre Perron, 2003, Computation and analysis of multiple structural change models, *Journal of applied econometrics* 18, 1–22.

Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *The Journal of Finance* 66, 911–946.

Barr, Michael S, 2012, Financial crisis and the path of reform, the, *Yale J. on Reg.* 29, 91.

Bernanke, Ben S, 2011, The effects of the great recession on central bank doctrine and practice, *Keynote address at the Federal Reserve Bank of Boston 56th Economic Conference Long Term Effects of the Great Recession* 1, 8–19.

Bernanke, Ben S, and Alan S Blinder, 1988, Credit, money, and aggregate demand, *The American Economic Review* 435–439.

Berry, Steven, 1994, Estimating discrete-choice models of product differentiation, *The RAND Journal of Economics* 242–262.

Berry, Steven, James Levinsohn, and Ariel Pakes, 1995, Automobile prices in market equilibrium, *Econometrica: Journal of the Econometric Society* 841–890.

Bertrand, Marianne, Matilde Bombardini, and Francesco Trebbi, 2015, *Lobbying the Dodd-Frank Act of*, mimeo UBC.

Bo, Zhiyue, 2004, The 16th central committee of the chinese communist party: formal institutions and factional groups, *Journal of Contemporary China* 13, 223–256.

Borio, Claudio, and Haibin Zhu, 2012, Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism?, *Journal of Financial Stability* 8, 236–251.

Brunnermeier, Markus K, and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial studies* 22, 2201–2238.

Carey, Mark, Mitch Post, and Steven A. Sharpe, 1998, Does corporate lending by banks and finance companies differ? evidence on specialization in private debt contracting, *The Journal of Finance* 53, 845–878.

Chamberlain, Gary, 1982, Multivariate regression models for panel data, *Journal of Econometrics* 18, 5–46.

Chen, Liang, Juan J Dolado, and Jesús Gonzalo, 2014, Detecting big structural breaks in large factor models, *Journal of Econometrics* 180, 30–48.

Chen, Ye, Hongbin Li, and Li-An Zhou, 2005, Relative performance evaluation and the turnover of provincial leaders in china, *Economics Letters* 88, 421–425.

Coates, IV, and C John, 2014, Cost-benefit analysis of financial regulation: Case studies and implications, *Yale Law Journal, Forthcoming* .

Cochrane, John H, 2014, Challenges for cost-benefit analysis of financial regulation, *The Journal of Legal Studies* 43, S63–S105.

Comerton-Forde, Carole, Terrence Hendershott, Charles M Jones, Pamela C Moulton, and Mark S Seasholes, 2008, Time variation in liquidity: The role of market maker inventories and revenues .

Cox, Gary W, Frances McCall Rosenbluth, and Michael F Thies, 1999, Electoral reform and the fate of factions: The case of japan's liberal democratic party, *British Journal of Political Science* 29, 33–56.

Dang, Tri Vi, Gary Gorton, and Bengt Holmström, 2016, The information sensitivity of a security, *Working Paper* .

Di Maggio, Marco, and Marcin Kacperczyk, 2016, The unintended consequences of the zero lower bound policy, *Working Paper* .

Dick, Astrid A., 2008, Demand estimation and consumer welfare in the banking industry, *Journal of Banking and Finance* 32, 1661–1676.

Dick-Nielsen, Jens, Peter Feldhütter, and David Lando, 2012, Corporate bond liquidity before and after the onset of the subprime crisis, *Journal of Financial Economics* 103, 471–492.

Dittmer, Lowell, 2003, Leadership change and chinese political development, *The China Quarterly* 176, 903–925.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2016, The deposits channel of monetary policy, *Working Paper* .

Driscoll, John C, and Ruth A Judson, 2009, Sticky deposit rates: Data and implications for models of price adjustment, Technical report, Working Paper.

Dudley, W., 2015, Regulation and liquidity provision, *Remarks at the SIFMA Liquidity Forum, NYC, 9/30/2015* .

Duffie, Darrell, 2012, Market making under the proposed volcker rule, *Rock Center for Corporate Governance at Stanford University Working Paper* .

Edwards, Amy K, Lawrence E Harris, and Michael S Piwowar, 2007, Corporate bond market transaction costs and transparency, *The Journal of Finance* 62, 1421–1451.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos, 2015, Deposit competition and financial fragility: Evidence from the us banking sector, *Fama-Miller Working Paper* 15–33.

Feldhütter, Peter, 2012, The same bond at different prices: identifying search frictions and selling pressures, *Review of Financial Studies* 25, 1155–1206.

Francois, Patrick, Ilia Rainer, and Francesco Trebbi, 2014, The dictator's inner circle, Technical report, National Bureau of Economic Research.

Francois, Patrick, Ilia Rainer, and Francesco Trebbi, 2015, How is power shared in africa?, *Econometrica* 83, 465–503.

Frieden, Jeffry, 2015, The political economy of adjustment and rebalancing, *Journal of International Money and Finance* 52, 4–14.

Garman, Mark B, 1976, Market microstructure, *Journal of financial Economics* 3, 257–275.

Giannetti, Daniela, and Michael Laver, 2005, Party cohesion, party factions and legislative party discipline in italy, in *Joint Workshop Sessions of the European Consortium for Political Research: Granada*, Citeseer.

Glushkov, Denys, 2007, Corporate bonds and ratings: Overview and research issues, in *WRDS User Meeting, Wharton School, University of Pennsylvania*.

Goldstein, Michael A, and Edith S Hotchkiss, 2007, Dealer behavior and the trading of newly issued corporate bonds, in *Afa 2009 san francisco meetings paper*.

Gorton, Gary, and Andrew Metrick, 2010, Regulating the shadow banking system, *Brookings Papers on Economic Activity* 2010, 261–297.

Gorton, Gary, and Andrew Metrick, 2012, Securitized banking and the run on repo, *Journal of Financial Economics* 104, 425–451.

Goyenko, Ruslan, Avanidhar Subrahmanyam, and Andrey Ukhov, 2011, The term structure of bond market liquidity and its implications for expected bond returns, *Journal of Financial and Quantitative Analysis* 46, 111–139.

Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of financial Economics* 66, 361–407.

Gürkaynak, Refet S, Brian Sack, and Jonathan H Wright, 2007, The us treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291–2304.

Gustafsson, Elisabeth, 2014 .

Hanson, Samuel Gregory, David Stuart Scharfstein, and Aditya Vikram Sunderam, 2015, An evaluation of money market fund reform proposals, *Working Paper* .

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, 1998, Characterizing selection bias using experimental data, Technical report, National bureau of economic research.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd, 1997, Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *The review of economic studies* 64, 605–654.

Ho, Katherine, and Joy Ishii, 2011, Location and competition in retail banking, *International Journal of Industrial Organization* 29, 537–546.

Ho, Thomas, and Hans R Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial economics* 9, 47–73.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2013, Noise as information for illiquidity, *The Journal of Finance* 68, 2341–2382.

Huang, Jing, 2006, *Factionalism in Chinese communist politics* (Cambridge University Press).

Iyer, Rajkamal, and Manju Puri, 2012, Understanding bank runs: The importance of depositor-bank relationships and networks, *American Economic Review* 102, 1414–45.

Jia, Ruixue, Masayuki Kudamatsu, and David Seim, 2015, Political selection in china: The complementary roles of connections and performance, *Journal of the European Economic Association* 13, 631–668.

Kacperczyk, Marcin, and Philipp Schnabl, 2013, How safe are money market funds?, *The Quarterly Journal of Economics* 128, 1073–1122.

Kaiser, R. G., ????, How america's essential institution works, and how it doesn's .

Kashyap, Anil K, and Jeremy C Stein, 1995, The impact of monetary policy on bank balance sheets, in *Carnegie-Rochester Conference Series on Public Policy*, volume 42, 151–195, Elsevier.

Kashyap, Anil K., and Jeremy C. Stein, 2000, What do a million observations on banks say about the transmission of monetary policy?, *American Economic Review* 90, 407–428.

Krishnamurthy, Arvind, 2002, The bond/old-bond spread, *Journal of Financial Economics* 66, 463–506.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2015, The impact of treasury supply on financial sector lending and stability, *Journal of Financial Economics* 118, 571–600.

Kyle, Albert S, 1985, Continuous auctions and insider trading, *Econometrica: Journal of the Econometric Society* 1315–1335.

Li, Cheng, 2002, The shanghai gang: Force for stability or cause for conflict, *China Leadership Monitor* 1, 1–18.

Li, Cheng, 2012, The battle for china top nine leadership posts, *The Washington Quarterly* 35, 131–145.

Li, Cheng, 2013, A biographical and factional analysis of the post-2012 politburo, *China Leadership Monitor* 41, 1–17.

Li, Hongbin, and Li-An Zhou, 2005, Political turnover and economic performance: the incentive role of personnel control in china, *Journal of Public Economics* 89, 1743–1762.

Longstaff, Francis A., 2004, The flight-to-liquidity premium in u.s. treasury bond prices, *Journal of Business* 77, 511–526.

Longstaff, Francis A, Sanjay Mithal, and Eric Neis, 2005, Corporate yield spreads: Default risk or liquidity? new evidence from the credit default swap market, *The Journal of Finance* 60, 2213–2253.

McCabe, Patrick E, Marco Cipriani, Michael Holscher, and Antoine Martin, 2013, The minimum balance at risk: A proposal to mitigate the systemic risks posed by money market funds, *Brookings Papers on Economic Activity* 2013, 211–278.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal, 2013, *Political Bubbles: Financial Crises and the Failure of American Democracy* (Princeton University Press).

Mian, Atif, Amir Sufi, and Francesco Trebbi, 2008, The political economy of the us mortgage default crisis, Technical report, National Bureau of Economic Research.

Mian, Atif, Amir Sufi, and Francesco Trebbi, 2014, Resolving debt overhang: Political constraints in the aftermath of financial crises, *American Economic Journal: Macroeconomics* 6, 1–28.

Moreira, Alan, and Alexi Savov, 2016, The macroeconomics of shadow banking, *National Bureau of Economic Research Working Paper* .

Nelson, Benjamin, Gabor Pinter, and Konstantinos Theodoridis, 2015, Do contractionary monetary policy shocks expand shadow banking?, *Bank of England Working Paper* .

Nevo, Aviv, 2000, A practitioner's guide to estimation of random-coefficients logit models of demand, *Journal of Economics & Management Strategy* 9, 513–548.

Nevo, Aviv, 2001, Measuring market power in the ready-to-eat cereal industry, *Econometrica* 69, 307–342.

Newey, Whitney K, and Kenneth D West, 1986, A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Parlatore, Cecilia, 2016, Fragility in money market funds: Sponsor support and regulation, *Journal of Financial Economics* .

Persico, Nicola, Jose CR Pueblita, and Dan Silverman, 2011, Factions and political competition, *Journal of Political Economy* 119, 242–288.

Posner, Eric A, and E Glen Weyl, 2013, Benefit-cost analysis for financial regulation, *American Economic Review* 103.

Pozsar, Zoltan, Tobias Adrian, Adam B Ashcraft, and Haley Boesky, 2010, Shadow banking, *Working Paper* .

Radecki, Lawrence J., 1998, The expanding geographic reach of retail banking markets, *Economic Policy Review* 15–34.

Reynaert, Mathias, and Frank Verboven, 2014, Improving the performance of random coefficients demand models: The role of optimal instruments, *Journal of Econometrics* 179, 83–98.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market., *Journal of Finance* 39, 1127–39.

Scharfstein, David, and Adi Sunderam, 2016, Market power in mortgage lending and the transmission of monetary policy, *Working Paper* .

Schmidt, Lawrence DW, Allan G Timmermann, and Russ Wermers, 2016, Runs on money market mutual funds, *Working Paper* .

Shih, Victor, 2004, Factions matter: personal networks and the distribution of bank loans in china, *Journal of Contemporary China* 13, 3–19.

Shih, Victor, Christopher Adolph, and Mingxing Liu, 2012, Getting ahead in the communist party: explaining the advancement of central committee members in china, *American Political Science Review* 106, 166–187.

Smets, Frank, 2013, Financial stability and monetary policy: How closely interlinked?, *Sveriges Riksbank Economic Review* 3, 121–160.

Stein, Jeremy C, 2012, Monetary policy as financial stability regulation, *The Quarterly Journal of Economics* 127, 57–95.

Stein, Jeremy C, 2013, Overheating in credit markets: origins, measurement, and policy responses, *speech, February* 7.

Stock, James H, and Mark Watson, 2009, Forecasting in dynamic factor models subject to structural instability, *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry* 173, 205.

Stock, James H, and Mark W Watson, 2011, Dynamic factor models, *Oxford Handbook of Economic Forecasting* 1, 35–59.

Sunderam, Adi, 2015, Money creation and the shadow banking system, *Review of Financial Studies* 28, 939–977.

Teles, Pedro, and Ruilin Zhou, 2005, A stable money demand: Looking for the right monetary aggregate, *Economic Perspectives* 29, 50.

Thakor, Anjan, 2012, The economic consequences of the volcker rule, *Report by the US Chamber Center for Capital Market Competitiveness* .

Williams, John C., 2014, Financial stability and monetary policy: happy marriage or untenable union?, *FRBSF Economic Letter* .

Yellen, Janet, 2014, Monetary policy and financial stability, *The 2014 Michel Camdessus Central Banking Lecture* .