Distribution Agreement

Signature:

________________________________ ________________

Zhao Li Date

Essays on Macroeconomics and Finance

By

Zhao Li
Doctor of Philosophy

Economics

Tao Zha
Advisor

Kaiji Chen
Committee Member

Vivian Yue
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Essays on Macroeconomics and Finance

By

Zhao Li

M.Sc., Emory University, 2014

B.A., Peking University, 2011

Advisor: Tao Zha, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the James T. Laney School
of Graduate Studies of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in Economics
2016

Abstract

Essays on Macroeconomics and Finance
By Zhao Li

This dissertation documents important issues on finance and macroeconomics. The first two chapters study the impact of uncertainty, which has recently become a hot research topic, although the mechanism of how it works remains an open question. I investigate the relationship between uncertainty and lending behavior in the syndicated loan market in order to propose new micro mechanisms. The Dealscan, Compustat, and CRSP databases are merged and the firm-level uncertainty is calculated from stock return volatility. Through comprehensive empirical studies, I ascertain that uncertainty substantially affects the quantity and price of the market. When uncertainty is higher, the loan shares tend to concentrate in the hands of lead lenders and the credit spread is higher. Along with empirical analyses, I propose a simple model to explain why higher uncertainty leaves the lead lenders with higher loan shares. The mechanism relies on the fact that uncertainty raises the unobservable investigation effort, which makes the benefit of shirking for the lead lenders higher. As a result, the participant lenders have to pull back their investment in order to leave the lead lenders with a higher share, which reduces the incentive of lead lenders to shirk responsibility. The first chapters thus propose new transmission channels of uncertainty's impact and sheds light on further theoretical and empirical works in this topic. The third chapter explores the BVAR forecasting methodologies for China. While it is well known that many models used for Western economies do not perform well in explaining and forecasting China's economic data, I challenge this convention by building rigorous econometric forecasting models for the Chinese economy. Different state-of-the-art Bayesian Vector-autoregression (BVAR) models are built, revised, and evaluated. It is found that the richer data set of additional macroeconomic data and sectoral data helps forecast the GDP, CPI, and interest rate. The large-scale BVAR model with 124 variables turns out to be the champion of forecasting models and it is more effective in extracting information from a large database than factor models and hierarchical models.

\

Essays on Macroeconomics and Finance

By

Zhao Li

M.Sc., Emory University, 2014

B.A., Peking University, 2011

Advisor: Tao Zha, Ph.D.

A dissertation submitted to the Faculty of the James T. Laney School
of Graduate Studies of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
in Economics
2016

Table of Contents

List of Figures

List of Tables

Preface

This dissertation contains three chapters, which document several important topics in macroeconomics and finance through both theoretical modeling and empirical analysis.

The focus of the first two chapters is on uncertainty. While a growing body of literature discusses the relationship between uncertainty and other key macroeconomic variables, the mechanism of how this relationship works remains an open question. I investigate the micro-foundation of uncertainty's impact through the analysis of one specific market — the syndicated loan market — and propose new transmission channels on how uncertainty affects the economy. These two chapters look into the quantity effect and price effect, respectively.

The first chapter discusses uncertainty's quantity effect, or the relationship between uncertainty and loan structure. In order to better understand the mechanism, here is a brief introduction about how a syndicated loan works. A syndicated loans means that a loan has a single borrower and multiple lenders. Naturally, different lenders play different roles and they can be divided into two main groups: lead lenders and participant lenders. The lead lenders initiate the loan, collect information, and monitor the firm. The participant lenders, on the other hand, are basically passive followers. Although they retain a share of the loan, they usually do not take on any specific responsibilities. The most important empirical finding in this chapter is that lead lenders tend to take higher loan shares when uncertainty rises. This is a very robust pattern that prevails in different empirical tests. The reason why it is important is because a loan structure is related to the total loan size: the loan provided by the lead lender is relatively stable whereas the participant lender's loan supply presents higher variability, which implies that when a loan share retained by lead lenders becomes higher due to uncertainty, the total loan supply tightens. As such, I am able to find a new channel that relates uncertainty with credit tightening. A theoretical model is also provided to explain how this happens. The core of the theoretical model is the unobservable investigation effort taken by the lead lender. When higher uncertainty results in a higher investigation effort, the loan share retained by the lead lender increases because the higher share will raise the potential loss if the lead lender shirks.

The second chapter studies uncertainty's price effect, or relationship between uncertainty and credit spread. With the merged data of the Dealscan, Compustat, and CRSP databases, I have access to individual loan data instead of the aggregate credit spread data and the credit spread data in this analysis measures the cost of the loan more precisely. It is found that higher uncertainty brings about higher credit spread, which confirms the existing theoretical models. The more interesting practice of this chapter is exploring uncertainty's impact from the side of the lender, which is usually overlooked in literature. I find that the credit spread is also higher with a higher level of lender uncertainty; however, the magnitude of the impact is smaller than that of the borrower. Last but not least, I find that the impact of the borrower's uncertainty is cyclical. There are at least two conflicting hypotheses on this issue and they are both well supported by economic theory. This paper provides some evidence to help validate these theories. It is found that when the GDP growth rate is lower, a more uncertain borrower will secure a loan at higher prices; this finding confirms the credit ration theory.

The third chapter explores the Bayesian VAR models and their ability to forecast China's economy. Admittedly, many theories and models used in Western economies fail to work on China, which has been attributed to the poor quality of the Chinese data. I argue that the challenge may lie in the fact that the right model has not yet been found. In this chapter, several state-of-the-art methodologies are evaluated and improved specifically for China's economy and I get some promising results. First, it is found that a simple revision of the parameter values in a small Sims&Zha model is able to bring about good results. More importantly, the large database contains important information that helps improve the model's performance. A large BVAR model built from a BGR algorithm turns out to be most effective. I also find ways of further improving the model's performance by proposing a Chinese version of the large BVAR model, which delivers very nice forecast for GDP, CPI, and the interest rate and is the champion of all models. In regards to another methodology, the BFAVAR model, it turns out that this model's performance is unbalanced: it does a better job in forecasting inflation and a worse job in forecasting GDP and the interest rate. The champions of these models are able to outperform, or at least stay competitive with, the BCEI forecasts, which gives us more confidence when applying the findings and methodologies in this chapter to guiding real-life decisions.

1 Uncertainty and Syndicated Loan Structure

Abstract

While uncertainty has become a hot research topic recently, the mechanism of how it works remains an open question. This paper investigates the relationship between uncertainty and lending behavior in the syndicated loan market in order to propose new micro mechanisms. The databases of Dealscan, Compustat, and CRSP are merged and firm-level uncertainty is calculated from stock return volatility. Through comprehensive empirical studies, it is found that when uncertainty is higher, the loan shares tend to concentrate in the hands of lead lenders. Along with empirical analysis, I propose a simple model to explain why higher uncertainty leaves the lead lenders with higher loan shares. The mechanism lies in that uncertainty raises the unobservable investigation effort, which makes the benefit of shirking for the lead lenders higher. As a result, the participant lenders have to pull back their investment in order to leave the lead lenders with a higher share, which reduces the incentive of lead lenders to shirk the responsibilities. This paper proposes new transmission channels of uncertainty's impact and sheds light on further theoretical and empirical works in this topic.

1.1 Introduction

A growing body of literature discusses the relationship between uncertainty and other key macroeconomic variables, but the mechanism of how this relationship works remains an open question. This paper investigates a micro-foundation of uncertainty's impact on the wider market through the analysis of one specific market — the syndicated loan market — and proposes new transmission channels on how uncertainty affects the economy.

In the existing literature, there are two main explanations on how uncertainty strikes the economy. In the first strand of literature, Bloom (2009) proposes the wait-and-see channel by introducing the adjustment cost and irreversible investment in the production sector. Since it is costly to adjust the investment, entrepreneurs have to reduce their investment when uncertainty is higher; thus, the output is adversely affected by higher uncertainty. In the second strand, which includes Gilchrist, Sim, and Zakrajšek (2014) and Christiano, Motto, and Rostagno (2014), uncertainty is incorporated along with financial frictions.

These authors introduce a model showing that the balance sheet is tighter and the default risk is higher with higher borrower uncertainty. Since default is costly with financial friction, the lenders reduce loans and raise the loan price to the borrowers; thus, the credit shrinks and output decreases with higher uncertainty.

This paper is closely aligned with the second strand of literature in that I also use financial friction as part of a mechanism and mainly focus on the financial sector instead of the production sector. I show that uncertainty weighs on the syndicated loan market by altering the loan structure of the syndicated loan and indirectly affects the overall loan size. Before we look into the details of the analysis, it is helpful to have some background information about the syndicated loan market. A syndicated loan is one of the crucial sources of corporate finance, especially for big deals. According to Sufi (2007), syndicated lending accounts for the origin of more than half of U.S. corporate finance and almost 90 percent of the largest non-financial firms in the Compustat universe that have obtained syndicated loans. The overall size of syndicated loans has grown dramatically from $137 million in 1987 to $4.2 trillion in 2013 (Thomson-Reuters, 2013).

A syndicated loan is a loan with multiple lenders. Those lenders take on different roles in the transaction and can be divided into two main groups: lead lenders and participant lenders. The lead lenders initiate the loan, collect information, and monitor the firm. The participant lenders, on the other hand, are basically passive followers. Although they retain a share of the loan, they usually do not take on any specific responsibilities. The motivation of lead lenders to syndicate out the loans includes the legal limit on the size of a single loan or a lack of capital. The motivation of participant lenders to join the venture includes a lack of origination capabilities or economies of scale in terms of transaction cost and monitoring cost.

A typical syndication process works as follows. The lead lenders collect the information, investigate, and negotiate with the borrowers to originate the loan. The deal can either be underwritten in that the lenders guarantee the commitment or a "best-effort" syndication in which the lead lenders commit to underwriting less than the total loan size. The lead lenders then come up with the "information memorandum," which is shown to potential participant lenders. This memorandum contains information and analysis about the financial situation, business model, credit-worthiness, and relevant information. The participant lenders might provide comments and suggestions and they decide the loan amount to which they will subscribe. The deal then closes and the lead lender is responsible for the administration and monitoring of the loan. The lead lender explicitly collects a fee ranging from 10 to 40 basis points for these responsibilities (Dennis and Mullineaux, 2000).

In this paper, I am mainly focused on the loan structure, or the loan shares, retained by the lenders. Naturally, different lenders take different shares in the same deal. What is interesting is that the lead lenders tend to take higher loan shares when uncertainty rises. Early banking literature sheds some light on explaining this phenomenon. Holmstrom and Tirole (1997) model that the monitoring agent needs to invest some capital in the joint venture due to moral hazard problems and the more invested by the monitoring agent, the less likely the agent shirks. Following this framework, I propose a model that better fits the syndicated loan market. The lead lender is taken as the monitoring agent and the participant lender as an uninformed investor. Given that the lead lenders and participant lenders are faced with the same loan contract in reality, I take the unobservable investigation cost as the cause of an agency problem. With higher uncertainty, the investigation effort of the lead lenders increases and they are more likely to shirk. As a result, participant lenders pull back their investment so that lead lenders are left with a higher loan share, thus providing them with more incentive to perform the due diligence. In addition, when the loan supplied by lead lenders is relatively stable, the total loan supply is driven by uncertainty through the loan structure. The model I develop has neat derivation and solution and can thus be further incorporated within a standard DSGE model to serve as a micro-foundation.

Empirical evidence is provided in order to confirm the model above. A merged database of Dealscan, Compustat, and CRSP is constructed in order to investigate the relationship between uncertainty and loan quantity. Uncertainty is measured by the idiosyncratic volatility of the stock return of the borrowers and the detailed specification is specified in Section 1.4.2. From the time series plot and sub-sample table, there is very strong evidence that uncertainty drives the loan share and, ultimately, the total loan supply. The panel regression models deliver robust results as well. After other determinants are controlled for, one standard deviation increase in uncertainty can cause a 7 percent rise above the average loan share. For the loan supply model, uncertainty has a significant coefficient when loan share is not included as a control, whereas uncertainty's coefficient turns insignificant when the loan share is included as a regressor. This finding provides some evidence that uncertainty affects the loan supply through the loan share.

The analysis in this paper on loan structure is inspired by Sufi (2007)'s work; however, it differs in the following aspects. First, the research questions are different. Sufi tried to examine how information availability affects the loan shares of lead lenders; to be specific, what he is interested in is whether or not the borrower has a stock ticker or credit rating affects the loan structure. To the contrary, our finding adds to the literature by explaining the impact of uncertainty, which is completely different from information availability. What I argue is that even though the agent has the same level of information, the loan structure varies when the firms' situation is more volatile or more uncertain. I am able to show that after controlling for the information availability variables used by Sufi, the uncertainty's impact on the loan structure is still statistically and economically significant. Second, Sufi's work is purely empirical and this work is a combination of theoretical and empirical results. In addition to documenting the relationship between uncertainty and loan structure, the model sheds further light on this relationship. I model the agency problem between lead lenders and participant lenders using the framework by Chen and Zha (2014) and demonstrate that this model has the potential to be incorporated in the standard DSGE model for future research. Third, the econometric model in our paper is better documented. I include the explanatory variables from the lender side as well, which is basically overlooked by Sufi.

The following parts of the paper are organized as follows. Section 1.2 reviews the literature, Section 1.3 proposes the theoretical model, Section 1.4 conducts the empirical evidence, and Section 1.5 concludes the paper.

1.2 Literature Review

Uncertainty has recently drawn extensive the attention of researchers. It is well-documented that uncertainty adversely affects the economy and there are several explanations focusing on different angles of the mechanism. Bloom (2007, 2009) incorporates the time-varying second moment of productivity to a standard firm-level production model with the labor and capital adjustment cost. This model yields an inactive region widened by higher uncertainty because firms "wait-and-see", which means that firms become more cautious and scale back their plans in hiring and investment. A model simulation delivers the similar results as witnessed in the empirical findings. Bachmann and Bayer (2013) follow this mechanism and further analyze the macro implication of the "wait-and-see" effect by incorporating it in the DSGE model framework. Through calibration using German firm data, which covers broader companies compared to U.S. data, they show that the uncertainty shocks are unlikely to be the major driver of macro fluctuations.

The other strand of literature involves financial friction. Christiano et al. (2014) come up with a monetary dynamic general equilibrium model incorporating the Bernanke-Gertler-Gilchrist financial accelerator mechanism. They model uncertainty as the cross-sectional volatility of entrepreneur productivity. When the uncertainty is higher, the credit spread will rise, and it will be harder for the entrepreneur to raise capital, therefore investment falls, as does output and consumption. The simulation of the data shows that uncertainty, or "risk shocks" as termed by the authors, are the most important shocks driving the business cycles. Gilchrist et al. (2014) propose a similar model. In their model, firms have a moral hazard problem when securing external financing. Higher uncertainty increases the downside risk of firms, thus raising the cost of funds. When the credit spread is widened, the economy moves in the opposite direction of the uncertainty. My paper is actually related to the financial friction literature, but the model in this paper is focused on the moral hazard problem and dynamics among lenders instead of the relationship between lenders and borrowers.

The model I propose in this paper is also inspired by Diamond (1984) and Holmstrom and Tirole (1997). They argue that although the introduction of monitoring agents can alleviate the agency problem between lenders and borrowers, there is another layer of the agency problem between the monitoring agent and its client. In my model, the lead lenders are monitoring agents and participant lenders are their clients. This paper introduces uncertainty into the model, and uses an investigation effort as the channel to connect uncertainty and the loan structure. In addition, my model differs from ones in the previous literature in order to better capture the syndicated loan practice. In Diamond (1984)'s paper, the monitoring agent retains a zero asset so that it cannot be used to explain the capital structure. In Holmstrom and Tirole (1997)'s paper, the moral hazard problem is due to the fact that the firms do not work hard. In my model, firms instead cheating by reporting lower productivity. In Holmstrom's paper, the capital retained by the monitoring agent is a substitute to capital from other investors, and it will not affect the total loan supply; however, in my model, loan size can be varying and a lower loan share can raise the capital supply.

The empirical evidence of my paper follows a rich literature covering the syndicated loan market. Earlier works include Simons (1993), Dennis and Mullineaux (2000), Lee and Mullineaux (2004), and Jones, Lang, and Nigro (2005). These authors focus primarily on the incentives of syndicates, and find that the size of the loan, reputation of the lender, default probability of the borrowers, and information availability will affect the occurrence and price of the loan.

Specifically, for the literature on the loan structure, Sufi (2007) empirically documents that the lead bank has to retain a larger share of the loan when borrowers do not have a credit rating and ticker. He further finds that this data pattern occurs mainly as the result of a moral hazard problem of the lead bank rather than an adverse selection problem, and that lead bank and borrower's reputation can mitigate the effect to some extent.

1.3 Theory

1.3.1 Environment Overview

There are three agents in the market: the entrepreneur, lead lender and participant lender. The lead lender and participant lender first pool their capital. Suppose the lead lender has the capital input of y_l and the participant lender's capital input is y_p; thus the total capital is $y = y_l + y_p$, which is the total loans made to the entrepreneur.

There is some uncertainty in the productivity function of the entrepreneur and I assume the entrepreneur's technology function as follows:

$$Y = Ay \text{ , and}$$

$$A = \begin{cases} A_1 = \overline{A}[1 - \sigma\sqrt{\frac{1-\pi}{\pi}}] & \text{with probability } \pi \\ A_2 = \overline{A}[1 + \sigma\sqrt{\frac{\pi}{1-\pi}}] & \text{with probability } 1 - \pi \end{cases} . \tag{1.1}$$

A is the random productivity level variable as specified above; $\overline{A}$ is the average productivity level, $A_1$and A_2 denote the productivity in bad and good states, respectively, in which $A_1 < A_2$; π is the probability of the bad state with $0 < \pi < 1$; and σ is the measure of uncertainty. It can be calculated that $mean(A) = \overline{A}$ and $sd(A) = \overline{A}\sigma$.

The lenders sign the optimal contract with the entrepreneur before A is realized, and the entrepreneur reports the productivity level after A is realized. Suppose p_1, p_2 is set by the contract as the payoff to the lenders when the entrepreneur reports A_1 and A_2, respectively. Chances are that the entrepreneur cheats by reporting A_1 when the true productivity level is A_2. To prevent this from happening, the lead lender can decide whether or not to pay the investigation cost before the entrepreneur reports on the states. If the lead lender does not investigate, then the probability that the entrepreneur is found cheating is 0. If the lead lender investigates the entrepreneur, the probability that the entrepreneur is found to be cheating rises to to α, where $0 < \alpha < 1$. When it is discovered that the entrepreneur is cheating, all of the return of the entrepreneur will be confiscated.

Here I refer to the investigation cost as $M(\sigma)y$ and this increases with the uncertainty level σ. The reason why the investigating cost increases with uncertainty is based on the observation that more uncertain firms tend to be those whose situations are harder to evaluate, for example, firms that have more research expenses, operate in a volatile industry, or have a complex business model. It costs more to investigate those more uncertain entrepreneurs. In order to simplify the derivation, I specify the investigation cost function as

$$M(\sigma) = M_0(A_2 - A_1)^{\gamma} \text{ with } \gamma > 1 \text{ and } M_0 > 0. \tag{1.2}$$

The contract and the reported states of the entrepreneur are public information to both the lead lender and the participant lender, but only the lead lender is responsible for investigating. The participant lender cannot observe whether or not the lead lender has paid the investigation cost.

After the productivity shock A is realized and the lead lender investigates (or does not investigate), the lenders receive the return from the entrepreneur according to the contract, and share the profit from the project based on their capital input; thus the lead lender receives the share $s = \frac{y_l}{y}$ and the participant receives $1 - s = \frac{y_p}{y}$ of the project return. There is a probability that the lead lender claims an investigation has been completed when it has not in fact been done. The participant lender has to adjust its capital input y_p to make sure that the lead lender is not reporting erroneously. I will outline their specific incentive problems in the following sections.

1.3.2 Timeline of Movement

- The capital input of the lead lender y_l and the distribution of the entrepreneur's productivity A are given and they are public information for all agents.
- The participant lender decides its capital input y_p.
- The lead lender and participant lender sign a contract with the entrepreneur thus determining p_1 and p_2.
- The lead lender decides whether or not to investigate. If investigating, the lead lender has to pay the fixed cost of $M(\sigma)$.
- The productivity shock A is realized, and the entrepreneur reports its states. There is probability that the entrepreneur lies by understating its productivity level. If the lead lender has paid the investigation cost, then the probability is α that cheating is found. If the lead lender does not pay the investigation cost, then the entrepreneur's cheating will never be found.
- The lead lender reports to the participant lender whether it has investigated or not. (It is important to note that the lead lender might lie by claiming to have investigated

when in fact no investigation has occurred) The two lenders will share the return based on the entrepreneur's reported states, the investigation report by the lead lender, and the share of capital input y_l and y_p.

1.3.3 Lead Lender's Problem

I begin with an analysis of the lead lender's problem - whether an investigation is done or not done - which are shown in Case I and Case II, respectively. Next, I look into the lead lender's decision on investigation report to participant lender.

Case I: Lead Lender Investigates Entrepreneur

The optimal contract problem for the lead lender when investigating is

$$\max_{p_1,p_2}\{[\pi p_1+(1-\pi)p_2-My-y]s\}, \tag{1.3}$$

subject to

$$p_1 \leq A_1 y, \tag{1.4}$$

$$p_2 \leq A_2 y, \tag{1.5}$$

$$[1-\alpha][A_2 y-p_1] \leq A_2 y-p_2. \tag{1.6}$$

Equation (1.3) is the expected return function for the lead lender. The total gains of both the lead and participant lenders are the payment of p_1 and p_2 under different circumstances, respectively, net of the cost of investigating and the cost of funds. Here I standardize the cost of unit funds as 1, so that the cost to y units of the loan is y. Among the total returns, the lead lender receives the share of s. Equation (1.4) and (1.5) are limited liability constraints and the intuition is straightforward: the entrepreneur cannot be asked to submit a benefit more than what is produced in each case. Equation (1.6) is an incentive constraint for the entrepreneur. LHS is the expected return if the entrepreneur reports a bad state while it is in fact in a good state. RHS is the expected return for the entrepreneur when it tells the truth and it is in a good state. This constraint ensures that the entrepreneur does not lie to the lenders.

Proposition 1. *The incentive constraint (1.6) is binding, the limited liability constraint for p_1 (1.4) is binding, and the limited liability constraint for p_2 (1.5) is not binding.*

Proof: See Appendix.

Applying Proposition I, I can solve out p_1, p_2 as

$$p_1 = A_1 y, \tag{1.7}$$

$$p_2 = [A_2 - (1-\alpha)(A_2 - A_1)]y. \tag{1.8}$$

The total return to lenders when A_1 is realized is

$$\Pi_1 = [A_1 - M - 1]y. \tag{1.9}$$

The total return to lenders when A_2 is realized is

$$\Pi_2 = [A_2 - (1-\pi)(1-\alpha)(A_2 - A_1) - M - 1]y. \tag{1.10}$$

The total expected return to lenders is

$$E\Pi = \left[\overline{A} - (1-\pi)(1-\alpha)(A_2 - A_1) - M - 1\right] y. \tag{1.11}$$

Let $B = \overline{A} - (1-\pi)(1-\alpha)(A_2 - A_1) - M - 1$, then $E\Pi = By$. It can be seen that the expected return is proportional to y and return rate is not related to y.

Given that the lenders will share the return based on their capital input share, the expected return for the lead lender and participant lender is $E\Pi_l = By_l$ and $E\Pi_p = By_p$, respectively.

Case II: Lead Lender Does Not to Investigate Entrepreneur

In this case, the lead lender does not investigate the entrepreneur, which is unobservable to the participant lender. The probability that the entrepreneur is found to be cheating is 0; thus, the entrepreneur will report the bad productivity A_1 every time. In this case, $p_1 = A_1 y$ and p_2 satisfies that $A_1 y \leq p_2 \leq A_2 y$. In this case, the lead lender saves the investigation cost My. Therefore, the return for the lenders is

$$\Pi_1^U = \Pi_2^U = E\Pi^U = [A_1 - 1]\, y. \tag{1.12}$$

Lead Lender's Decision on Investigation Report

Now I consider the agency problem between lenders. When the entrepreneur reports A_2, one can reason with certainty that both the lead lender and entrepreneur are not cheating.

First, the entrepreneur is not lying because he or she has no incentive to report higher productivity while achieving a lower productivity state. The lead lender is unable to lie to the participant lender because the entrepreneur's report is publicly available.

The only possible case wherein the lead lender may lie to the participant lender is if the entrepreneur reports A_1. There are two possible action sets that may lead to this result. In the first case, the lead lender pays the investigating cost and the entrepreneur achieves and then reports the lower productivity A_1. In this case, the lead lender will never lie to the participant lender by claiming not investigating the entrepreneur. The reason is straightforward: the lead lender's return is $[A_1 - M - 1]y_l$ when claiming an investigation was completed. If the lead lender decides otherwise to claim that the entrepreneur was not investigated, then the return is $[A_1 - M - 1]\,y - [A_1 - 1]\,y_p = [A_1 - M - 1]y_l - My_p$, which is smaller compared to the case when the lead lender tells the truth.

In the second case, the lead lender does not investigate the entrepreneur and the entrepreneur achieves A_2 while reporting A_1. In this case, the lead lender lies to the participant lender by claiming an investigation completed. The rationale is as follows. If the lead lender tells the truth, its return is $[A_1 - 1]\,y_l$. If it lies by claiming to have investigated, the return is $[A_1 - 1]\,y - [A_1 - M - 1]\,y_p = [A_1 - 1]\,y_l + My_p$, which brings the lead lender a greater return than if it tells the truth.

From the analysis above, it can be seen that the lead lender will always claim to have investigated the entrepreneur and is cheating when it does not in fact conduct the investigation. Unfortunately, the participant lender cannot distinguish between the two cases identified in the preceding paragraphs.

1.3.4 Participant Lender's Problem

The participant lender has access to the contract signed between lead lender and entrepreneur (p_1, and p_2) as well as the reported states by the entrepreneur (A_1 or A_2). However, the participant lender cannot observe whether or not the lead lender investigates. Therefore, there is a probability that the lead lender informs the participant lender that an investigation was completed when that is not the case. Under this circumstance, the entrepreneur will always report to the lead lender that the productivity level is A_1. The lead lender will share the return with the participant lender following Case I, so the participant lender gets the return

$$E\Pi_p^{lie} = [A_1 - M - 1)]\, y_p < E\Pi_p, \tag{1.13}$$

$$E\Pi_p - E\Pi_p^{lie} = \alpha(1-\pi)(A_2 - A_1)y_p. \tag{1.14}$$

I further assume that the required return for the participant lender Π_p^R is between $E\Pi_p^{lie}$ and $E\Pi_p$; thus, the participant lender does not want the lead lender to cheat. The participant lender has to adjust its capital input in order to induce the lead lender to investigate the entrepreneur.

When the lead lender lies to the participant lender, the expected return for the lead lender is

$$E\Pi_l^{lie} = [A_1 - 1]\, y - [A_1 - M - 1)]\, y_p. \tag{1.15}$$

The lead lender will not lie if the return when being honest is greater or equal to $E\Pi_l^{lie}$:

$$E\Pi_l = By_l \geq E\Pi_l^{lie}. \tag{1.16}$$

If rearranging the incentive constraint above, I obtain:

$$My_p \leq [B - (A_1 - 1)]\, y_l. \tag{1.17}$$

RHS is actually the cost of lying for lead lender's loan share y_l. Remember that By_l is the return for the lead lender when an investigation is actually completed. But if the lead lender does not investigate, the return decreases to $(A_1 - 1)y_l$. Therefore, the cost of lying is $[B - (A_1 - 1)]\, y_l$. To make sure that the cost is greater than zero, here I assume that $\sigma < \left[\frac{\alpha(1-\pi)}{M_0}\right]^{\frac{1}{\gamma-1}} \frac{1}{\bar{A}(\sqrt{\frac{\pi}{1-\pi}}+\sqrt{\frac{1-\pi}{\pi}})}$.

LHS is the gain of lying for the lead lender. By telling the truth, the participant lender is supposed to receive the return of $[A_1 - 1]\, y_p$ from the joint venture. However, if the lead lender deceitfully claims to have investigated, then the return to the participant lender reduces to $[A_1 - M - 1)]y_p$. By lying, the lead lender saves $[A_1 - 1]\, y_p - [A_1 - M - 1)]y_p = My_p$. Another way to think of this is that the lead lender retains an investigation fee of My_p, which is collected from the participant lender. Instead of using this fee for an investigation, the lead lender just pockets the gain.

At the same time, I must consider the possibility that the participant lender may cooperate with the entrepreneur. To be more specific, it is possible that the participant lender can induce the entrepreneur to reveal the true situation by providing compensation to the entrepreneurs. If this cooperation is feasible, the agency problem between lead lenders and participant lenders can be avoided and Equation (1.17) is thus an invalid constraint. Unfortunately, this scenario is not feasible. The participant lender is unable to provide enough incentives to the entrepreneur because the entrepreneur's gains increase when the lead lender shirks. The detailed proof is provided in the Appendix.

Therefore, the agency problem between the lead lender and the participant lender continues to exist and Equation (1.17) is a valid constraint. If the participant lender wants the lead lender to investigate and tell the truth, the input y_p shall not be greater than a certain level. Since the return to the participant lender is proportional to the capital input, eventually this constraint is binding.

$$y_p = \frac{B-(A_1-1)}{M} y_l. \tag{1.18}$$

By plugging s, B, A_1, M into the equation, I can get

$$s = \frac{y_l}{y_l + y_p} = \frac{M_0 \left[\overline{A} \left(\sqrt{\frac{\pi}{1-\pi}} + \sqrt{\frac{1-\pi}{\pi}} \right) \sigma \right]^{\gamma-1}}{\alpha(1-\pi)}. \tag{1.19}$$

With the reasonable parameter setting, the loan share retained by the lead lender, s, increases when uncertainty rises. Since y_l is exogenously given, the total loan supply falls when the loan share retained by the lead lender rises.

1.4 Empirical Evidences

1.4.1 Data

The loan data comes from the Loan Pricing Corporation's (LPC) Dealscan Database. The total database includes 277,949 loans from 1987 to 2014. The loan details are collected from the staff reports of lead arrangers and SEC filings. I calculate the following variables from Dealscan: the loan share of the lead lenders, number of the banks involved, loan size, maturity, and credit spread. Thanks to the link table provided by Chava and Roberts (2008), Dealscan can be merged with Compustat through the gvkey of the borrowers. There is a an well-established link between CRSP and Compustat provided by WRDS; thus I am able to merge together the loan variables, balance sheet variables of the borrowers, and the stock return data.

As argued by the previous literature, the characteristics of the lenders also affect the lending behavior. Unfortunately, Dealscan does not provide information about the lenders. Although I can use some links from Chava & Roberts (2008), there are not enough data points since their links are mainly for borrowers. I can only identify the lenders in Chava and Roberts's link only if the lenders also show up as borrowers in Dealscan. Therefore, I have to hand-match the lenders in Dealscan with the firms in Compustat by names and locations. The identities of the lenders are aggregated to the parent firm level if they are recorded as departments or subsidiaries in Dealscan. When handling M&A, in which some lenders disappear or some lenders change names, I follow these simple rules: If one lender acquires another lender and does not change its name, I will use the acquirer's identity for the firm after the acquisition date. If the two lenders merge and establish a new firm, I regard the newly constituted firm as a completely different entity. In addition, I use information from the lender's official website, Chicago Fed's BHC merging data, and merge table generously provided by Mora (2014) and Ivashina (2009) as the cross-reference. Since the loan deals are concentrated under the big lenders, hand-matching the top 100 lenders' loans along with the links provided by Chava & Roberts covers around half of the original sample size.

To ensure a reliable result, I discard the observation if the borrowers are in the utility or financial sector, or the borrowers with irregular values of balance sheet variables, like the negative asset or negative net worth. For the stock return data, I only include the firms listed on NYSE, AMEX and Nasdaq.

1.4.2 Measure of Uncertainty

Existing literature has different views on the proxy of uncertainty. While some literature measures uncertainty as the cross-section variance, I follow Bloom (2007) and Gilchrist et al. (2014) to measure the uncertainty as the idiosyncratic stock return volatility of the borrowers. The reason to use this measure is that it better fits the uncertainty modeled in Section 1.3. The uncertainty of the productivity is reflected on the stock return volatility of the firm because stock return is a good indicator that takes into account all types of information provided by the relevant institution. Some might argue that this measure is not a direct variable in the loan market; but in reality, when the lenders are making a decision on whether or not to make loans to the firms, they actually face the similar information available to stock investors. Both lenders and stock investors have to look into information including the financial statement, industry characteristics, management, etc., which are all reflected by stock return. The other concern is the irrelevant information contained in the stock return, like some well-established factors in the market. In order to address this problem, I calculate the volatility from the residuals of the factor model instead of using the simple stock return standard deviation.

To be specific, I collect the stock return data from the CRSP database and run the standard Fama-French factor model for each firm in each quarter in the first stage. The model is shown in (1.20); i denotes the firms, t_d denotes each day in a specific quarter, $R_{it_d}$is the stock return for firm i, $r^f_{t_d}$ is the risk free rate, f_{t_d} denotes the factors(Farmer-French 3 factors, momentum factor, and VXO), and u_{it_d} is the residual at day t_d. Then I use residual $\hat{u}_{it_d}$ as the estimate for u_{it_d}, and $\tilde{u}_{it}$ to denote the means of $\hat{u}_{it_d}$ for each firm in each quarter. so,

$$R_{it_d} - r^f_{t_d} = \alpha_i + \beta_i' f_{t_d} + u_{it_d}. \tag{1.20}$$

In the second stage, I calculate the standard deviation of the residuals from the first stage for each firm in each quarter and denote it by σ_{it}, which is the measure of the uncertainty.

$$\sigma_{it} = \sqrt{\left[\frac{1}{D_t}\sum_{d=1}^{D_t}(\hat{u}_{it_d} - \tilde{u}_{it})\right]}. \tag{1.21}$$

Since the stock return reflects the overall information about the firm, any change in the commercial policy, financial statement, industry characteristics, management, or other things will cause the return to change. After I control for the common component in the stock return, the standard deviation will better reflect the uncertainty of the firm itself.

1.4.3 Measure of Loan Structure

In this paper, I follow Ivashina (2009) for the classification of lead lenders and participant lenders. In the Dealscan database, there is a variable called "lender role" which indicates the roles of various lenders. The lead lenders include the Administrative Agent, Bookrunner, Lead arranger, Lead bank, Lead Manager, Arranger, and Agent. The participant lenders include the Participant, Lender, and other titles. The loan structure is indexed by the share of the loan amount supplied by the lead lenders $= \frac{\text{Loan Amount of Lead Lenders}}{\text{Loan Amount of All Lenders}}$. [1]

Based on the classification of lenders, I plot the loan structure over time in Figure 1.1. The red region represents the loan amount supplied by the participant lenders and the blue region shows the loan amount supplied by the lead lenders. The loan amount refers to the total loan size of all the deals in the same quarter. It is very striking that most of the variability of the total loan supply comes from the variability of the participant lender's loan supply; the lead lenders' contribution remains relatively stable. Therefore, the loan share retained by the lead lenders is an important driver for the credit supply, which motivates this paper. If uncertainty affects the loan share retained by the lead lenders, it will ultimately translate into an impact on the total loan supply.

[1] I also try another index, the Number Share of the Lead Lenders=$\frac{\text{\# of Lead Lenders}}{\text{\# of All Lenders}}$, as the robustness check. It delivers a similar result, which is reported in the Appendix.

1.4.4 Descriptive Statistics

For the controls in regressions, I include the following variables: maturity as a loan variable; sales, operating income margin, debt equity ratio, current ratio, net worth, leverage, and transparency as borrower-side variables; asset and capitalization as lender-side variables. The definitions of these variables are listed in the appendix. What is worth noting is the dummy variable of transparency, which corresponds to the opaqueness in Sufi (2007)'s analysis and captures the information availability. After the merging and data cleaning procedure is completed, I have 5764 observations from 1987 to 2012.

According to the methodology specified in Section 1.4.2, I can calculate the uncertainty level with respect to each loan. Then I can rank the whole sample based on the uncertainty level: a low uncertainty level refers to the first tertile, a medium uncertainty level refers to the second tertile, and a high uncertainty level refers to the third tertile. Means of the variables are reported in Table 1.1. First, take a look at the upper panel, which records the loan variables. It is obvious that uncertainty matters for almost all of the variables. The loan shares[2] retained by the lead lender rises from 23% to 39%, almost doubles when uncertainty migrates from the low to high level sub-sample. Besides, it is clear that with a higher borrower uncertainty, there will be more lenders, a smaller loan size, shorter maturity, and a higher credit spread.

The middle panel of Table 1.1 summarizes the borrower's balance sheet variables. It is seen that higher uncertainty is related to smaller sales, a lower income margin, a higher debt equity ratio, current ratio, leverage, and lower net worth. The lower panel denotes the lender's variables, which indicates that the lenders have less assets and a weaker balance sheet when uncertainty increases. All of these findings are consistent with both the conventional wisdom and findings in previous literature.

[2]In the following analysis, it by default refers to the loan shares by loan size.

The relationship between uncertainty and loan structure is even more clearly demonstrated when a graph is plotted. First, derive the the average of the loan share and uncertainty across different loans in the same quarter and plot them over time, as shown in Figure 1.2. The loan share of the lead lenders and uncertainty move together most of the time, and they are both counter-cyclical. Particularly, they both peak during the last financial crisis. The correlation of these two series is as high as 0.36. Besides, I build different uncertainty measures to test robustness of the result. To be specific, I calculate uncertainty using different factor structure and length of moving-window, and the details are discussed in Appendix A4. The corresponding average uncertainty measure over time are plotted in Figure 1.3 and Figure 1.4. It is obvious that those uncertainty measures are very similar, which implies that the data pattern that is found with our benchmark measure is robust across different uncertainty definitions.

1.4.5 Uncertainty and Loan Shares

Here I use the panel data regression model to investigate the relationship between uncertainty and loan structure. The dependent variable is the loan share retained by the lead lenders. The key independent variable is the uncertainty measure, which is calculated following Section 1.4.3. In the meantime, the results are quite similar if I adjust the uncertainty calculation method by altering the window to collect the residuals in the second stage of the calculation. In the Appendix A4, the horizon is changed to collect the residuals for up to one year and half year rather than one quarter (A one-quarter-window is used in Section 1.4.2). This leads to even better results.

In this model, the loan share of the lead lender is the dependent variable, whereas uncertainty is the key explanatory variable. The first lag is used for uncertainty in order to alleviate the problem of reverse causality. The results do not change a lot when the current value of uncertainty is used. Actually, the reverse causality is not a big issue for this specific model: Since investors in the stock market have no information about the loan share until the deal is closed, the stock return data reflects no information on the loan share. It is the uncertainty that drives the change in the loan share rather than vice versa.

The left panel of Table 1.2 reports the simple OLS result. The coefficient before uncertainty is positive at the 1% significance level. The coefficients for the controls are significant and their signs are consistent with our expectations. For the middle panel, the time fixed effect is active. The coefficient of uncertainty rises from 3.9 to 5.3 and is also significant under the 1% level. The right panel activates both the time fixed effect and the industry fixed effect. The R-squared is the highest in this scenario, which indicates an improvement of the fitness. The coefficient for the uncertainty is comparable with the middle panel and the coefficients of the controls have the correct signs and most are significant. Therefore, statistically speaking, I find some evidences that a higher uncertainty implies a higher loan share retained by the lead lenders.

I further evaluate the magnitude of the coefficient to see whether this relationship is economically important. It can be calculated that the standard deviation of the uncertainty is 0.55, thus one standard deviation change in uncertainty will lead to a 2% change in the loan share retained by the lead lenders. Note that the average of the loan share is 31%, this magnitude is equivalent to 7% percent increase of the average loan share, which is economically important. The results are robust when I use different measures of uncertainty and loan structure as reported in the Appendix A4-A5. The other concern with the analysis is that the regression results I get is too good to be true, and the statistical significance is spurious as the result of multicollinearity. I elaborate on this point in Appendix A6.

1.4.6 Uncertainty and Loan Size

Given that there is a relationship between uncertainty and loan share, what does it imply ultimately? In this subsection, I map this relation into the total loan amount. Actually,

Figure 1.1 already implies the existence of a potential channel in which the loan size is substantially driven by the loan share retained by the lead lenders. Combined with the relationship between uncertainty and loan share, it is natural to reason that uncertainty affects the total loan supply through its loan shares. To establish the relationship in a more rigorous way, panel regression models are run (see Table 1.3). The loan size, which is the dependent variable, is measured by the ratio of the loan amount over the net worth of the borrowers.

The explanatory variables include the uncertainty measure along with the other controls in Model I. The coefficient of uncertainty is negative, which is also statistically and economically significant. When uncertainty increases by one standard deviation, the loan size/net worth decreases by 6%. In Model II, loan share retained by lead lenders is also included as explanatory variable, which has a significant negative coefficient. When the loan share increases by 1%, the loan size ratio falls by 0.7%. However, the coefficient for the uncertainty measure is no longer significant, and even has a wrong sign. Therefore, I discover the empirical evidence that uncertainty affects the total loan supply through the loan share variable.

1.5 Conclusion

This paper adds to the literature on the mechanism of how uncertainty affects the economy. I propose a simple model to explain the positive relationship between uncertainty and the loan share retained by the lead lender. The core of this model is the unobservable investigation effort taken by the lead lender. When higher uncertainty results in a higher investigation effort, the loan share retained by the lead lender increases because the higher share will raise the potential loss if the lead lender shirks. The empirical evidence comes from the merged databases of Dealscan, Compustat, and CRSP. The descriptive statistics, time series plot, and different econometric models demonstrate the robust relationship between uncertainty and loan structure, which confirms our theoretical model. This paper sheds light on explaining the impact of uncertainty in the loan market and shall inspire further theoretical and empirical works in this field.

References

Bachmann R., and C. Bayer. 2013. ”Wait-and-See” business cycles? Journal of Monetary Economics, 60 (6), 704-719.

Bloom N., S. Bond, and J. V. Reenen . 2007. Uncertainty and Investment Dynamics. Review of Economic Studies, 74 (2), 391-415.

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

Burkart M., D. Gromb, and F. Panunzi. 1997. Large Shareholders, Monitoring, and the Value of the Firm. The Quarterly Journal of Economics, 693-728.

Chava S., and M. R. Roberts. 2008. How Does Financing Impact Investment? The Role of Debt Covenants. The Journal of Finance, 63 (5), 2085-2121.

Chava S., and A. Purnanandam. 2011. The Effect of Banking Crisis on Bank-dependent Borrowers. Journal of Financial Economics, 99 (1), 116-135.

Chen K., and T. Zha. 2014. Lending Efficiency Shocks. Working Paper.

Chodorow-Reich G. 2014. The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008–9 Financial Crisis. The Quarterly Journal of Economics, 129 (1), 1-59.

Christiano L., R. Motto, and M. Rostagno. 2014. Risk shocks. American Economic Review, 104 (1), 27-65.

Dennis S. A., and D. J. Mullineaux. 2000. Syndicated Loans. Journal of Financial Intermediation, 9, 404-426.

Diamond D. 1984. Financial Intermediation and Delegated Monitoring. Review of Economic Studies, 51 (3), 393-414.

Gilchrist S., J. W. Sim, and E. Zakrajsek. 2014. Uncertainty, Financial Frictions, and Investment Dynamics. NBER Working Papers 20038.

Gorton G., and G. Pennacchic. 1995. Banks and Loan Sales Marketing Nonmarketable Assets. Journal of Monetary Economics, 35 (3), 389-411.

Holmstrom B., and J. Tirole. 1997. Financial Intermediation, Loanable Funds, and the Real Sector. The Quarterly Journal of Economics, 663-691.

Ivashina V., and D. Scharfstein. 2010. Bank Lending during the Financial Crisis of 2008. Journal of Financial Economics, 97, 319-338.

Jones J., W. Lang, and P. Nigro. 2005. Agent Bank Behavior in Bank Loan Syndication. Journal of Financial Research, 28, 385–402.

Lee S. W., and D. J. Mullineaux. 2004. Monitoring, Financial Distress, and the Structure of Commercial Lending Syndicates. Financial Management, 33, 107-130.

Leland H., and D. Pyle. 1977. Informational Asymmetries, Financial Structure, and Financial Intermediation. Journal of Finance, 32 (2), 371-387.

Mora N. 2014. Lender Exposure and Effort in the Syndicated Loan Market. Journal of Risk and Insurance, 82 (1), 205-252.

Pichler P., and W. Wilhelm. 2001. A Theory of the Syndicate: Form Follows Function. The Journal of Finance, 56 (6), 2237-2264.

Simons K. 1993. Why Do Banks Syndicate Loans? New England Economic Review of the Federal Reserve Bank of Boston, (Jan), 45-52.

Sufi A. 2007. Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans. The Journal of Finance, 62 (2), 629-668.

Standard and Poor. 2006. A Guide to the Loan Market. McGraw-Hill Companies, Inc.

Thomson-Reuters Corporation. 2013. Global Syndicated Loans Review Reuters, 2013 full year.

Appendix

A1: Proof of Proposition 1

Suppose the limited liability constraint for p_1 is not binding, then p_1 can always increase by a small amount without violating the incentive constraint and limited constraint for p_2. With this small change, the lenders can raise their expected utility or expected return. Therefore, p_1 will keep increasing until it reaches the upper bound, so that $p_1 = A_1 y$. Therefore, the limited liability constraint for p_1 (1.4) is binding.

Suppose the limited liability constraint for p_2 is binding, thus $p_2 = A_2 y$. Plugged p_1 and p_2 into (1.5), we have:

$$[1-\alpha][A_2 y - A_1 y] \leq A_2 y - A_2 y.$$

Since $\alpha > 0$ and $A_1 < A_2$, the LHS of the inequality is a positive value and the RHS is equal to zero, which violates the above inequality. Therefore, $p_2 < A_2 y$ and the limited liability constraint for p_2 is not binding.

Given that the limited liability constraint for p_2 is not binding and $p_1 = A_1 y$, I plug them into the incentive constraint:

$$p_2 \leq A_2 y - [1-\alpha][A_2 y - A_1 y].$$

Since the lenders will always want higher p_2, it will continue increasing until it reaches the upper limit, which is under the circumstance that $p_2 = A_2 y - [1-\alpha][A_2 y - A_1 y]$. This is how the binding incentive constraint is derived.

A2: Proof that Participant Lenders Cannot Cooperate with Entrepreneurs Against Lead Lenders

I start with the gains of the entrepreneurs after introducing the agency problem between the lead lender and participant lender. If the lead lender always tells the truth to the participant lender, the expected return for the entrepreneur is:

$$\Pi_{entre} = \pi(A_1 y - p_1) + (1-\pi)(A_2 y - p_2) = (1-\pi)(1-\alpha)(A_2 - A_1)y.$$

With the agency problem between lenders, the entrepreneur benefits since it can always claim to have a low productivity without getting caught:

$$\Pi'_{entre} = \pi(A_1 y - p_1) + (1-\pi)(A_2 y - p_1) = (1-\pi)(A_2 - A_1)y.$$

Hence, the entrepreneur's gain from the lender's agency problem is:

$$\Delta\Pi_{entre} = \Pi'_{entre} - \Pi_{entre} = (1-\pi)\alpha(A_2 - A_1)y.$$

Now the participant lender's loss is derived from the probability that the lead lender might lie to the participant lender, which is exactly shown by Equation (1.14):

$$\Delta\Pi_{par} = -\alpha(1-\pi)(A_2 - A_1)y_p.$$

Since $-\Delta\Pi_{par} < \Delta\Pi_{firm}$, the loss of the participant lender is smaller than the gain of the entrepreneur when the lead lender lies to the participant lender. The participant lender is unable to provide enough incentive to ask the entrepreneur to reveal the real situation. As a result, the participant lender can only rely on adjusting the capital input to satisfy the incentive constraint for the lead lender.

A3: Definition of Variables

Loan Share (by loan size): Loan amount supplied by lead lenders/loan amount supplied by all lenders.

Loan Share (by number of lenders): Number of lead lenders/number of all lenders.

Loan Size (log): Natural Logarithm of value of the loan.

Sales (log): Natural Logarithm of revenue.

EBITDA: Operating income before depreciation.

Operation Income Margin: EBITDA/asset.

Debt/Equity: (Long-term debt + debt in current liabilities)/shareholder equity.

Current Ratio: Total current assets/total current liabilities.

Net Worth (log): Natural Logarithm of (asset-liability).

Leverage: Total asset/shareholder equity.

Transparency: Dummy, which is equal to 1 if the borrower has both S&P senior long-term debt rating and ticker; otherwise, the dummy is equal to 0.

Credit Rating: S&P senior long-term debt rating.

Lender Asset (log): Natural Logarithm of lender's asset.

Lender Capitalization Rate: Lender shareholder equity/total assets.

A4: Different Measures of Uncertainty

The uncertainty measure can be calculated with a different time horizon from the residuals. The residuals are collocated up to 1 quarter prior to the origination date in our previous analysis. In Table 1.8 and Table 1.9 I redo the analysis between uncertainty and loan shares for uncertainty calculated from half-year window and 1-year window, respectively. Other variables and econometric models remain the same. It is important to note that the relationship in the previous model persists when I use a different uncertainty measure.

The uncertainty measure can also be calculated using different factors. In one scenario, I calculate the uncertainty strictly following Gilchrist et al. (2014)'s approach and do not include VXO as a factor. The result of the regression between uncertainty and loan shares are reported in Table1.10. In another scenario, I don't include any factor structure at all, and uncertainty is measured by the simple volatility of stock return. The results of the regression between uncertainty and loan shares are reported In Table 1.11. Note that the uncertainty is calculated from the 1-quarter window and econometric models remain the same. It is important to note that the relationship in the previous model persists when I change the factor structure.

A5: Different Measure of the Loan Structure

The loan share by the number of the lead lenders is now used, which is defined as the $share = \frac{\text{\# of Lead Lenders}}{\text{\# of All Lenders}}$. The other variables and econometric models remain the same. As found in Table 1.12The relationship that I found in the benchmark model prevails when I use different loan structure measure.

A6: Multicollinearity's Concern

Theoretically speaking, as long as I don't have perfect multicollinearity, the estimates of the parameters are consistent and unbiased. However, the variance of the estimates would be inflated, thus the typical syndrome of multicollinearity is significance for the whole model whereas insignificance for individual coefficients. As reported in Table 1.2 and Table 1.3, I have significant estimates for most of the coefficients, which implies little evidence for multicollinearity. Nevertheless, I still address this concern to stay safe. In Table 1.4 and Table 1.5, the correlation matrix of the explanatory variables in my models are reported. It is found that the correlation among the variables are generally not very high, which leaves little room for multicollinearity. The only exception is the net worth, which is highly correlated with the sales. In the literature, the net worth is thought to be useful in explaining the lending behavior, and that's exactly the reason why I include it in the benchmark model. In Table 1.6 and Table 1.7, I redo the regression between the loan share and uncertainty when I exclude the net worth. In the meantime, I also report the results of the models using uncertainty alone as the explanatory variable for reference. It can be seen that the results excluding net worth as explanatory variable has similar coefficients as the benchmark model, and results with uncertainty alone give us greater coefficients for uncertainty, which is basically the results of the missing variable bias. To sum up, I don't find any evidence that multicollinearity is affecting our analysis.

Tables and Figures

Figure 1.1: Loan Structure Over Time

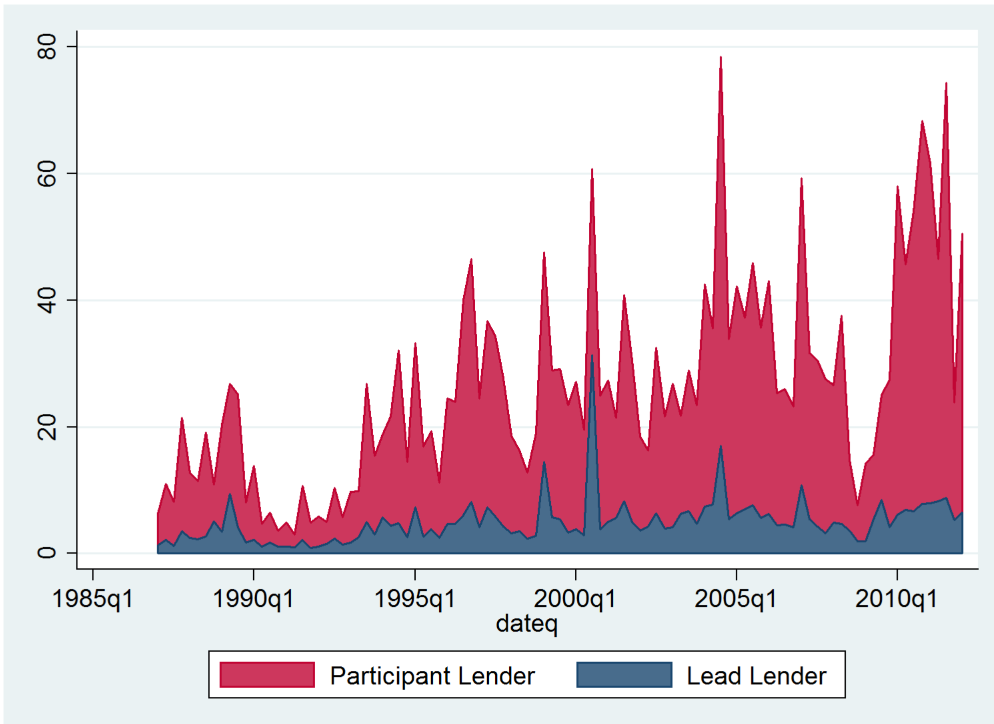

Figure 1.2: Time Series of Loan Structure and Uncertainty

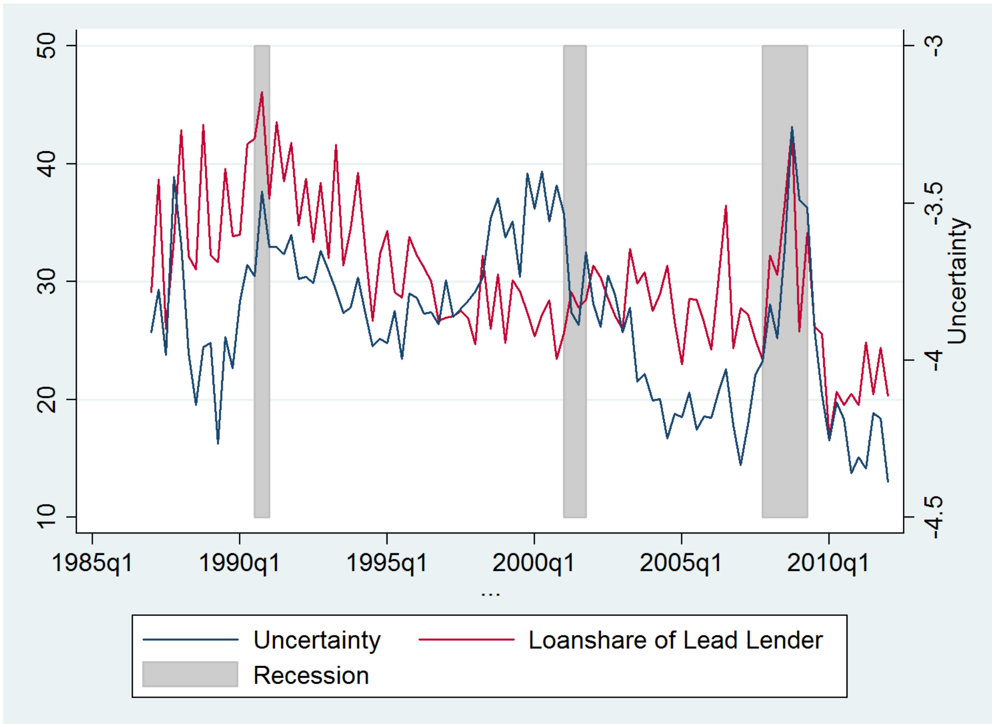

Figure 1.3: Uncertainty with Different Windows

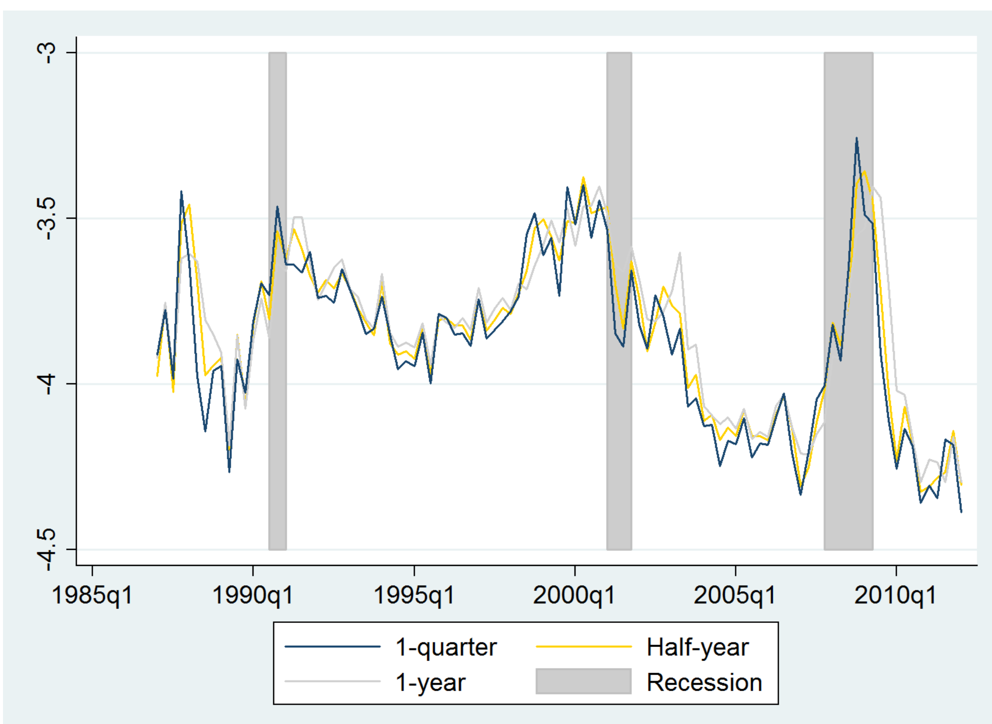

Figure 1.4: Uncertainty with Different Factor Structure

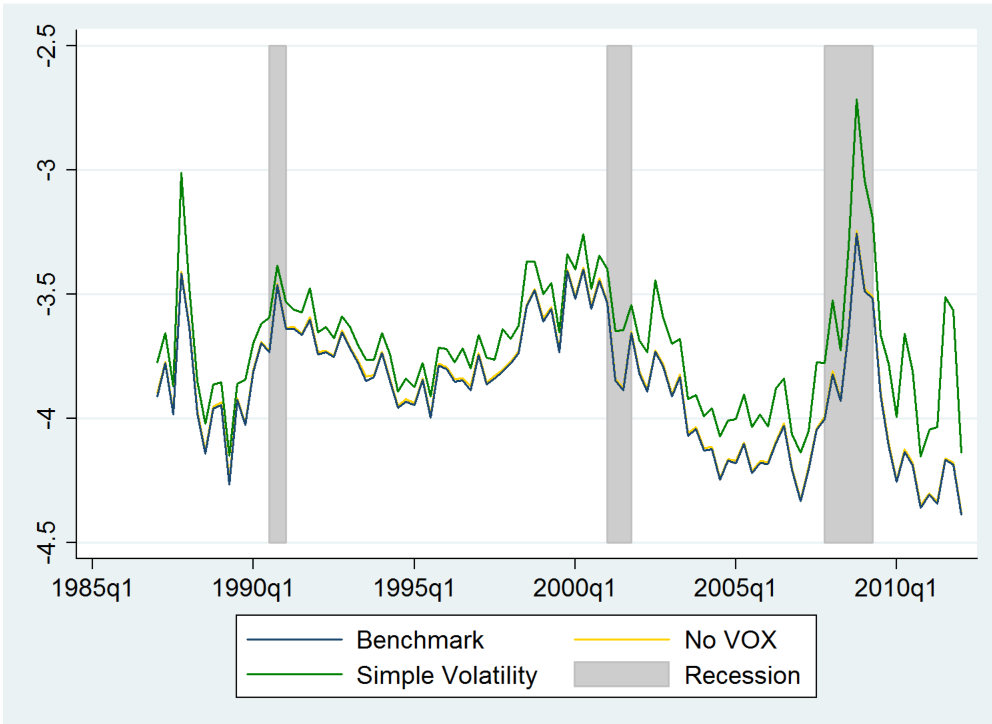

Table 1.1: Sub-sample Means by Uncertainty Level

	Uncertainty Level		
Variables	Low	Medium	High
Loan Share (by loan size)	23.11	30.62	39.31
Loan Share (by # of lenders)	0.19	0.24	0.29
# of banks	11.17	8.88	6.85
Loan Size	514M	246M	103M
Maturity	52.03	50.76	43.22
Credit Spread	133.79	182.79	253.65
Sales (log)	6.21	5.12	3.96
Operation Income Margin	0.04	0.04	0.03
Debt Equity Ratio	0.99	1.16	1.28
Current Ratio	1.79	2.02	2.15
Net Worth (log)	6.68	5.59	4.37
Leverage	2.98	2.86	3.11
Lender Asset (log)	13.14	12.76	12.56
Lender Capitalization Rate	0.08	0.07	0.07

Table 1.2: Uncertainty and Loan Shares

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	3.930***	5.375***	4.176***
	(6.158)	(7.508)	(5.896)
Maturity (log)	-2.503***	-3.042***	-3.133***
	(-5.930)	(-6.955)	(-7.730)
Sales (log)	-1.240***	-1.279***	-2.022***
	(-3.564)	(-3.661)	(-3.476)
Transparency	-2.628***	-3.248***	-2.968***
	(-4.323)	(-5.389)	(-4.600)
Operation Margin	-29.75***	-27.95***	-22.99**
	(-3.137)	(-2.941)	(-2.199)
Current Ratio	1.631***	1.591***	1.262***
	(5.812)	(5.678)	(3.850)
Net Worth (log)	-3.529***	-3.180***	-3.060***
	(-8.311)	(-7.387)	(-5.249)
Leverage	-0.590***	-0.561***	-0.569***
	(-4.207)	(-3.885)	(-3.701)
Lender Asset(log)	-0.496**	-1.332***	-1.683***
	(-2.400)	(-3.433)	(-4.474)
Lender Capitalization Rate	38.09***	3.656	-9.274
	(3.224)	(0.242)	(-0.624)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.262	0.304	0.402

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.3: Uncertainty and Loan Size

Variables	Model I	Model II
Loan Share		-0.742***
		(-8.520)
Uncertainty	-6.031**	1.325
	(-2.553)	(0.322)
Maturity (log)	6.100***	6.459***
	(4.302)	(2.727)
Sales (log)	-2.443	-14.15***
	(-1.348)	(-4.184)
Transparency	-3.360	-13.96***
	(-1.548)	(-3.717)
Operation Margin	231.6***	449.9***
	(6.384)	(7.409)
Current Ratio	1.221	0.315
	(1.108)	(0.165)
Net Worth (log)	-21.20***	-9.446***
	(-11.53)	(-2.783)
Leverage	18.20***	23.81***
	(42.40)	(26.62)
Lender Asset (log)	3.315**	4.618**
	(2.483)	(2.110)
Lender Capitalization Rate	-35.12	258.5***
	(-0.759)	(2.997)
Time FE	Yes	Yes
Industry FE	Yes	Yes
R-squared	0.359	0.416

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.4: Correlation Matrix of Variables Panel A

Variables					
Uncertainty	1.00				
Maturity (log)	-0.02	1.00			
Sales (log)	-0.41	-0.22	1.00		
Transparency	-0.25	-0.09	0.47	1.00	
Operation Margin	-0.10	0.00	0.08	-0.03	1.00
Current Ratio	0.10	0.08	-0.23	-0.18	-0.01
Net Worth (log)	-0.46	-0.20	0.84	0.47	0.00
Leverage	0.07	0.03	0.10	0.13	-0.07
Lender Asset (log)	-0.21	0.01	0.28	0.22	-0.02
Lender Capitalization Rate	-0.09	0.03	0.03	-0.01	0.01

Table 1.5: Correlation Matrix of Variables Panel B

Variables					
Current Ratio	1.00				
Net Worth (log)	-0.14	1.00			
Leverage	-0.21	-0.18	1.00		
Lender Asset (log)	-0.03	0.33	0.03	1.00	
Lender Capitalization Rate	0.08	0.03	-0.07	0.09	1.00

Table 1.6: Uncertainty and Loan Shares: Smaller Model Panel A

Variables	OLS		Fixed Effect I	
Uncertainty	13.72***	5.240***	14.88***	6.703***
	(28.57)	(8.591)	(26.81)	(9.903)
Maturity (log)		-2.261***		-2.873***
		(-5.382)		(-6.585)
Sales (log)		-3.829***		-3.600***
		(-18.07)		(-16.82)
Transparency		-3.759***		-4.288***
		(-6.388)		(-7.355)
Operation Margin		-13.13		-13.09
		(-1.378)		(-1.373)
Current Ratio		1.666***		1.621***
		(5.819)		(5.667)
Leverage		-0.0181		-0.0432
		(-0.145)		(-0.341)
Lender Asset (log)		-0.876***		-1.555***
		(-4.299)		(-4.032)
Lender Capitalization Rate		40.98***		8.185
		(3.441)		(0.539)
Time FE	No	No	Yes	Yes
Industry FE	No	No	No	No
R-squared	0.106	0.249	0.164	0.294

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.7: Uncertainty and Loan Shares: Smaller Model Panel B

Variables	Fixed Effect II	
Uncertainty	13.11***	4.977***
	(22.58)	(7.178)
Maturity (log)		-3.050***
		(-7.508)
Sales (log)		-4.736***
		(-17.84)
Transparency		-3.264***
		(-5.064)
Operation Margin		-4.187
		(-0.425)
Current Ratio		0.903***
		(2.811)
Leverage		-0.0760
		(-0.623)
Lender Asset (log)		-1.762***
		(-4.674)
Lender Capitalization Rate		-10.12
		(-0.679)
Time FE	Yes	Yes
Industry FE	Yes	Yes
R-squared	0.273	0.398

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Loan Shares and Uncertainty with Half-year Window

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	4.936***	6.971***	5.551***
	(7.388)	(9.211)	(7.072)
Maturity (log)	-2.404***	-3.040***	-3.108***
	(-5.716)	(-7.011)	(-7.680)
Sales (log)	-1.222***	-1.288***	-1.941***
	(-3.513)	(-3.696)	(-3.343)
Transparency	-2.507***	-3.100***	-2.872***
	(-4.129)	(-5.160)	(-4.456)
Operation Margin	-26.04***	-23.37**	-20.35*
	(-2.763)	(-2.482)	(-1.945)
Current Ratio	1.632***	1.567***	1.295***
	(5.847)	(5.647)	(3.962)
Net Worth (log)	-3.404***	-2.960***	-2.924***
	(-8.000)	(-6.853)	(-5.025)
Leverage	-0.596***	-0.547***	-0.571***
	(-4.278)	(-3.852)	(-3.717)
Lender Asset (log)	-0.457**	-1.308***	-1.656***
	(-2.209)	(-3.377)	(-4.408)
Lender Capitalization Rate	37.70***	2.364	-9.606
	(3.198)	(0.157)	(-0.648)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.264	0.308	0.403

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Loan Shares and Uncertainty with 1-year Window

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	5.717***	8.024***	6.671***
	(7.837)	(9.791)	(7.849)
Maturity (log)	-2.334***	-3.075***	-3.135***
	(-5.569)	(-7.147)	(-7.763)
Sales (log)	-1.179***	-1.256***	-1.878***
	(-3.391)	(-3.608)	(-3.239)
Transparency	-2.458***	-3.036***	-2.805***
	(-4.059)	(-5.069)	(-4.356)
Operation Margin	-24.58***	-21.57**	-19.01*
	(-2.615)	(-2.291)	(-1.818)
Current Ratio	1.599***	1.531***	1.279***
	(5.732)	(5.519)	(3.918)
Net Worth (log)	-3.327***	-2.846***	-2.797***
	(-7.799)	(-6.566)	(-4.806)
Leverage	-0.609***	-0.547***	-0.570***
	(-4.453)	(-3.937)	(-3.720)
Lender Asset (log)	-0.493**	-1.283***	-1.629***
	(-2.387)	(-3.311)	(-4.342)
Lender Capitalization Rate	38.18***	4.681	-6.427
	(3.242)	(0.310)	(-0.434)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.266	0.311	0.405

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.10: Loan Shares and Uncertainty w/o VXO factor

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	3.931***	5.375***	4.177***
	(6.159)	(7.509)	(5.897)
Maturity (log)	-2.503***	-3.042***	-3.133***
	(-5.930)	(-6.955)	(-7.730)
Sales (log)	-1.240***	-1.279***	-2.022***
	(-3.564)	(-3.661)	(-3.476)
Transparency	-2.628***	-3.248***	-2.968***
	(-4.323)	(-5.389)	(-4.601)
Operation Margin	-29.74***	-27.95***	-22.99**
	(-3.136)	(-2.940)	(-2.198)
Current Ratio	1.631***	1.591***	1.262***
	(5.812)	(5.678)	(3.851)
Net Worth (log)	-3.529***	-3.180***	-3.060***
	(-8.310)	(-7.387)	(-5.249)
Leverage	-0.590***	-0.561***	-0.569***
	(-4.207)	(-3.885)	(-3.701)
Lender Asset (log)	-0.496**	-1.331***	-1.683***
	(-2.400)	(-3.433)	(-4.474)
Lender Capitalization Rate	38.09***	3.659	-9.271
	(3.224)	(0.242)	(-0.624)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.261	0.304	0.401

Robust t-statistics in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.11: Loan Shares and Simple Stock Return Volatility

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	3.375***	4.724***	3.665***
	(5.546)	(6.516)	(5.018)
Maturity (log)	-2.579***	-3.078***	-3.166***
	(-6.140)	(-7.036)	(-7.806)
Sales (log)	-1.242***	-1.277***	-2.007***
	(-3.569)	(-3.653)	(-3.447)
Transparency	-2.686***	-3.314***	-2.998***
	(-4.413)	(-5.484)	(-4.642)
Operation Margin	-31.46***	-29.69***	-24.28**
	(-3.309)	(-3.114)	(-2.321)
Current Ratio	1.604***	1.582***	1.264***
	(5.711)	(5.637)	(3.854)
Net Worth (log)	-3.705***	-3.389***	-3.247***
	(-8.856)	(-7.981)	(-5.595)
Leverage	-0.602***	-0.573***	-0.581***
	(-4.286)	(-3.963)	(-3.777)
Lender Asset (log)	-0.605***	-1.340***	-1.687***
	(-2.912)	(-3.452)	(-4.480)
Lender Capitalization Rate	32.22***	1.960	-10.76
	(2.735)	(0.129)	(-0.724)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.260	0.301	0.400

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 1.12: Loan Share(by # of Lenders) and Uncertainty

Variables	OLS	Fixed Effect I	Fixed Effect II
Uncertainty	0.0214***	0.0233***	0.0152***
	(5.837)	(5.811)	(3.816)
Maturity (log)	-0.0270***	-0.0320***	-0.0329***
	(-10.21)	(-11.49)	(-13.71)
Sales (log)	-0.0104***	-0.00976***	-0.00526*
	(-5.212)	(-4.928)	(-1.716)
Transparency	-0.0231***	-0.0223***	-0.0169***
	(-6.129)	(-5.986)	(-4.593)
Operation Margin	-0.299***	-0.342***	-0.413***
	(-5.091)	(-5.617)	(-6.729)
Current Ratio	0.00732***	0.00687***	0.00664***
	(4.782)	(4.467)	(3.565)
Net Worth (log)	-0.0173***	-0.0174***	-0.0272***
	(-7.268)	(-7.321)	(-8.739)
Leverage	-0.00159**	-0.00168**	-0.00303***
	(-2.257)	(-2.354)	(-4.170)
Lender Asset (log)	-0.000475	-0.0127***	-0.0158***
	(-0.365)	(-5.682)	(-6.980)
Lender Capitalization Rate	-0.195***	-0.487***	-0.528***
	(-2.618)	(-5.105)	(-6.751)
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.104	0.131	0.196

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

2 Uncertainty and Credit Spread

Abstract

The relationship between uncertainty and credit spread is explored using the loan-level data from the merger of Compustat, CRSP, and Dealscan. It is found that credit spread rises by 13 percent with a standard deviation increase in borrower's uncertainty and this impact is even larger in an economic downturn. The interaction between uncertainty and other determinants of credit spread is also studied and a significant size effect is found; namely, the impact of uncertainty is weakened when a borrower has greater revenue and more net worth. This paper also looks into the uncertainty from the side of the lender which has been generally ignored in the previous literature. The lender's uncertainty proxy is built from lender's loan history and findings show that the lender's uncertainty can also affect the credit spread, although the impact is smaller compared with the uncertainty from the borrower. The empirical results in the paper may inspire further development in the theoretical model on the relationship between uncertainty and credit spread.

2.1 Introduction

This paper looks into the relationship between uncertainty and credit spread. Using new data, I will confirm the theoretical prediction by demonstrating the positive relationship between uncertainty and credit spread. In addition, I also present new results that are not accounted for in current theories: 1) The effect of uncertainty interacts with other determinants, especially size variables; 2) Lender uncertainty plays a role as well as borrower uncertainty while the impact is less substantial compared to the more conventional borrower uncertainty; 3) The impact of uncertainty is cyclical and a more uncertain firm is stricken more severely during an economic downturn.

Traditionally, researchers analyze the adverse impact of uncertainty via a the "wait-and-see" channel. According to Bloom (2007, 2009), when uncertainty is higher, firms cut their investment because it is costly to adjust investment as the transaction is irreversible. Inspired by the financial crisis of 2007-2009, more people relate to uncertainty through financial frictions. Both Gilchrist, Sim, and Zakrajšek (2014) and Christiano, Motto, and Rostagno (2014) model that when the uncertainty with a borrower is high, the balance sheet is tighter and the default risk is higher. When financial friction is present, it is costly to default; thus lenders reduce the loans to borrowers, resulting in shrinking credit and decreasing output in this period of higher uncertainty.

It is important to note that there is limited direct evidence from micro data studying how credit spread changes with uncertainty, let alone further investigation of how this relationship interacts with other factors. This paper will try to fill this gap in the literature by conducting a thorough empirical analysis using the loan-level data. More specifically, the Dealscan database, which contains rich information on individual loan deals in the United States, is studied in depth to highlight the relationship between uncertainty and credit spread and how other factors may impact this relationship.

This paper is different from previous literature in the following respects. First, the merged data of the Dealscan, Compustat, and CRSP database is used, which provides access to direct loan credit data for each individual loan, instead of aggregate credit spread data. Gilchrist, Sim, and Zakrajšek (2014) also conduct their empirical analysis on firm-level data; however, the credit spread data in this analysis measures the cost of the loan more precisely. The data that Gilchrist et al. use is derived from the bond prices in the second market, thus the credit spread in their data might not reflect the real cost of the loan for the firms. For example, when the credit spread in the bond market is higher, borrowers decide not to issue bonds at that moment and the credit spread does not affect the cost of financing. To the contrary, credit spread data in this analysis is exactly the price the borrowers pay when they get external financing.

Second, this is the first paper to study uncertainty from the side of the lender. As argued in the previous literature, the credit spread is higher when the borrower is more uncertain. What is usually overlooked is whether or not uncertainty from the lender side also matters.

After all, a loan is a transaction that involves two parties. I find that the credit spread is also higher with a higher level of lender uncertainty; however, the magnitude of the impact is smaller than that of the borrower's.

Third, the impact of the borrower's uncertainty is cyclical. There are at least two conflicting hypotheses on this issue and they are both supported well by economic theory. This paper provides some evidence to help validate these theories. It is found that when the GDP growth rate is lower, a more uncertain borrower will secure a loan at higher prices; this finding confirms the credit ration theory.

Finally, I also study the interaction between uncertainty and other variables. It is found that the impact of the solvency ratio and liquidity ratio are not significant; however, the size variables, including net worth, sales, and loan size, help dampen the relationship between uncertainty and credit spread.

With regards to the definition of uncertainty, there have been different measures employed in the literature. In this paper, we will proxy uncertainty as the idiosyncratic volatility of the stock return of the borrowers. The detailed specification is outlined in Section 2.4.1.

The following parts of the paper are organized as follows. Section 2.2 will review the literature, Section 2.3 provides the theoretical explanation, Section 2.4 specifies the methodology, Section 2.5 conducts fixed effect regressions, and Section 2.6 concludes the paper.

2.2 Literature Review

There are roughly two strands of literature related to our work. In the first strand, the literature concerning uncertainty from the producers aspect has drawn extensive attention of researchers recently. It is well-documented that uncertainty demonstrates counter-cyclical behaviors but the explanation for it remains an open end. Bloom (2007, 2009) proposes a standard firm-level production model with a time-varying second moment of the productivity along with the labor and capital adjustment cost. This model yields an inactive region widened with higher uncertainty in which firms “wait-and-see”, meaning that firms become more cautious and scale back their plans in hiring and investment. Boom’s simulation delivers similar results as the empirical findings. Bachmann and Bayer (2013) follow this explanation and further explore to the macro implication of “wait-and-see” by incorporating it into a DSGE model framework. Through calibration using German firm data, which covers a broader scope compared to U. S. data, they find that uncertainty shocks are unlikely to be the major driver of macro fluctuations.

The second strand of literature emphasize the interaction between uncertainty and financial friction. Chirstiano et al. (2014) come up with a monetary dynamic general equilibrium model incorporating a Bernanke-Gertler-Gilchrist financial accelerator mechanism. They model uncertainty as the cross-sectional volatility of entrepreneur productivity. When uncertainty is higher, the credit spread will rise and it will be harder for a entrepreneur to raise capital; therefore investment falls so as does output and consumption. The simulation of the data shows that uncertainty, or “risk shocks” as termed by the authors, is the most important shock driving business cycles. Gilchrist et al. (2014) uses a the similar model and incorporate the adjustment cost and financial friction together. The firms then have a moral hazard problem when seeking external financing. The uncertainty will raise the downside risk of firms and raise the cost of funds. When the credit spread widens, the economy moves in the opposite direction of the uncertainty.

There is yet another strand of literature that studies the determinant of credit spread. In Gilchrist et al. (2014), there are some primitive empirical analysis on which this paper follows and extends. Gilchrist et al. use the credit spread data from the bond market and find that the impact of uncertainty on investment is dampened when the credit spread is also included as the regressor. Based on this finding, Gilchrist et al. argue that the credit spread is the transmission channel of uncertainty's impact.

The empirical works on credit spread is huge in the literature. In this paper, I only cover the research conducted using the Dealscan database, which is closer to my work and sheds light on the understanding of this specific database. Ivashina (2009) documents the relationship between loan shares retained by the lead lenders and the credit spread. She further breaks the impact of a loan structure into two part. The asymmetric information effect reduces the credit spread whereas the diversification effect raises the spread. Chodorow-Reich (2014) finds that the credit spread is higher when lenders are less healthy and there is less of a relationship between lenders and borrowers. By linking the Dealscan database to employment data from the Bureau of Labor Statistics, Chodorow-Reich further points out that the borrowers have to reduce their employment during a credit market disruption. Hubbard et al. (2002) also investigate the bank-side determinant on the cost of the loan. It is found that after controlling for borrower risk and information cost, the credit spread is significantly lower for well-capitalized banks. Similar to the findings of Gilchrist et al. (2014), the authors find that the borrowers hold more precautionary savings when the lender's balance sheets are weaker. Güner (2006) documents the relationship between the loan price and the loan sales. This author finds that when banks sell the loan, which is against the interest of the borrowers, the credit is lower as is the compensation to borrowers.

2.3 Theoretical Explanation

In this section, I explain the relationship between uncertainty and credit spread using the simplified model developed by Christiano et al. (2014). I only focus on the financial sector; all other structures are abstracted. It is assumed that there are many identical entrepreneurs and banks in the economy and they all want to maximize their utility. To simplify the analysis, the utility function of both entrepreneurs and banks are linear, thus the utility maximization is equivalent to the return maximization.

2.3.1 Characteristic Functions

An entrepreneur has a net wealth N and receives a loan of B from banks, thus the entrepreneur has a total asset of $A = N + B$, which implies the leverage of $L = \frac{A}{N}$. The entrepreneur has a productivity function $(1+R_k)\omega$, in which R_k is the average return rate. ω is the idiosyncratic component of productivity with a log normal distribution and Christiano et al. assume that $\int_0^\infty \omega dF(\omega)$=1, which standardizes the average return of the product. The entrepreneur signs a contract with the bank that specifies both the loan amount, B, and interest rate, Z.

When the entrepreneur gets a low enough idiosyncratic shock of ω, he or she is unable to pay off the bull amount of the loan. The bank takes everything away from the entrepreneur after paying for the monitoring cost. Based on this, we can calculate the default cutoff ϖ from $(1+R_k)\varpi A = ZB$, which implies $\varpi = \frac{Z}{1+R_k}\frac{L-1}{L}$.

The return to the entrepreneur is calculated as follows:

$$\Pi_e = \frac{\int_{\varpi}^{\infty}[(1+R_k)\omega A - ZB]dF(\omega)}{N(1+R)} = \int_{\varpi}^{\infty}[\omega - \varpi]dF(\omega) \times \left(\frac{1+R_k}{1+R}\right)L. \tag{2.1}$$

Now introduce a few characteristic functions to simplify the notation; their CDFs are shown in Figure 2.1:

$$F(\varpi) \equiv \int_0^{\varpi} dF(\omega), \tag{2.2}$$

$$G(\varpi) \equiv \int_0^{\varpi} \omega dF(\omega), \tag{2.3}$$

$$\Gamma(\varpi) \equiv \varpi(1 - F(\varpi)) + G(\varpi). \tag{2.4}$$

It can be derived that the return to the entrepreneur is:

$$\Pi_e = [1-\Gamma(\varpi)]\left(\frac{1+R_k}{1+R}\right)L. \tag{2.5}$$

.

Mathematically, $\Gamma(\varpi)$ has some nice properties: $0 \leq \Gamma(\varpi) \leq 1$ and $\Gamma'(\varpi) = 1 - F(\varpi) > 0$. Based on that, $1-\Gamma(\varpi)$ decreases with ϖ, in which $\varpi = \frac{Z}{1+R_k}\frac{L-1}{L}$ and $\varpi = \frac{Z}{1+R_k}$ as L goes to infinity. Therefore, the entrepreneur's return is unbounded as leverage L increases, which means that the entrepreneur wants to borrow an unlimited amount of money in the absence of constraint.

2.3.2 Bank's Problem

The bank borrows funds from a household at a constant rate of R and makes loans to the entrepreneur at the rate of Z. Assume that banks are in the competitive market and each bank earns zero profit. When the entrepreneur is bankrupt, or $\omega < \varpi$, the bank has to pay the monitoring cost as μ share of the return. The zero profit condition can be written down as follows:

$$(1-F(\varpi))ZB + (1-\mu)\int_0^{\varpi} \omega dF(\omega)(1+R_k)A = (1+R)B. \tag{2.6}$$

This condition builds up the relation between cutoff and leverage as follows:

$$L = \frac{1}{1-\frac{1+R_k}{1+R}[\Gamma(\varpi)-\mu G(\varpi)]}. \tag{2.7}$$

The simulation using reasonable parameter values demonstrates the positive relationship between ϖ and L, which further translates into the upward sloping curve of a zero-profit constraint shown in Figure 2.2.

2.3.3 Entrepreneur's Problem

The entrepreneur wants to maximize his or her return given the zero profit constraint:

$$\underset{L,\varpi}{Max}\{\Pi_e = [1-\Gamma(\varpi)]\left(\frac{1+R_k}{1+R}\right)L\}, \tag{2.8}$$

subject to

$$L = \frac{1}{1-\frac{1+R_k}{1+R}[\Gamma(\varpi)-\mu G(\varpi)]}. \tag{2.9}$$

The first-order condition is

$$\frac{1-F(\varpi)}{1-\Gamma(\varpi)} = \frac{\frac{1+R_k}{1+R}[1-F(\varpi)-\mu\varpi F'(\varpi)]}{1-\frac{1+R_k}{1+R}[\Gamma(\varpi)-\mu G(\varpi)]}. \tag{2.10}$$

The first-order condition pins down ϖ. Based on that, we can further solve out leverage as $L = \frac{1}{1-\frac{1+R_k}{1+R}[\Gamma(\varpi)-\mu G(\varpi)]}$ and credit spread as $\frac{Z}{1+R} = \frac{1+R_k}{1+R}\varpi\frac{L}{L-1}$.

2.3.4 Impact of Uncertainty

In the model, a higher uncertainty means that the standard deviation of ω is greater and the distribution of ω is more fat-tailed. The mechanism of this channel is summarized by the following Figures. Figure 2.2 illustrates the entrepreneur's decision-making when uncertainty changes. The red curve is the bank's zero profit constraint and the blue curve is the entrepreneur's indifference curve. The entrepreneur optimizes his or her profit by selecting the tangent point of the two curves, which specifies the credit spread and leverage.

When uncertainty is higher, the zero profit constraint shifts to the upper left because the default risk of the entrepreneur is higher for the same amount of the loan, and the bank needs to charge a higher rate to those that are not bankrupt in order to retain the same profit. With this less favorable constraint, the entrepreneur has to pick a new tangent point with a higher credit spread and lower leverage. Therefore, there exists a positive relationship of uncertainty and credit spread, which is shown in Figure 2.3.

In the following analysis, I try to validate the prediction of the theory through the data. Apart from the evidence about the positive relationship between credit spreads and uncertainty, I also find some other interesting data patterns that cannot be well explained by the current theory, thus implying a more sophisticated model to take into account the new empirical findings.

2.4 Statistical Methods

2.4.1 Measure of Uncertainty

The existing literature has different views on the proxy of uncertainty. While some literature measures uncertainty as the cross-section variance, I follow Bloom (2007) and Gilchrist et al. (2014) to measure the uncertainty as the idiosyncratic stock return volatility of the borrowers. This measure fits the uncertainty modeled in Section 2.3 and is therefore a good one to use. The uncertainty of the productivity is reflected on the stock return volatility of the firm because the stock return is a good indicator that takes into account all the types of the information of the relevant institution. Some might argue that this measure is not a direct variable in the loan market; however, in reality, when the lenders are making decisions on whether or not to approve loans to the firms, they actually face information similar to that of as stock investors. Both lenders and stock investors must look into the financial statement, industry characteristics, management, etc. of firms and the stock return contains all this information. Another concern is the presence of irrelevant information contained in the stock return, including some well-established factors in the market. In order to address that problem, I calculate the volatility from the residuals of the factor model instead of using a simple stock return standard deviation.

To be specific, I collect the stock return data from the CRSP database and run a standard Fama-French factor model for each quarter in the first stage. The model is shown in Equation (2.11). i indexes firms; t_d indexes each day in a specific quarter; R_{it_d}, $r^f_{t_d}$, f_{t_d}, and $u_{it_d}$are the stock return for firm i, the risk-free rate, a group of factors (Fama-French 3 factors, momentum factor, and VXO), and residual at day t_d. I then estimate $\hat{u}_{it_d}$ for u_{it_d} and take the average for each firm in each quarter, denoted by $\tilde{u}_{it}$.

$$R_{it_d} - r^f_{t_d} = \alpha_i + \beta_i' f_{t_d} + u_{it_d}. \tag{2.11}$$

In the second stage, I calculate the standard deviation of the residuals from the first stage for each firm in each quarter and denote it by σ_{it}, which is the measure of the uncertainty.

$$\sigma_{it} = \sqrt{\left[\frac{1}{D_t}\sum_{d=1}^{D_t}\left(\hat{u}_{it_d} - \tilde{u}_{it}\right)\right]}. \tag{2.12}$$

Since the stock return reflects the overall information about the firm, any change in the commercial policy, financial statement, industry characteristics, management, or other facet will cause the return to change. After controlling for the common component in the standard deviation, I can focus on the idiosyncratic volatility of the firms.

2.4.2 Data

The loan data is from the Loan Pricing Corporation's (LPC) Dealscan Database. The total database includes 277,949 loans from 1987 to 2014. The loan details are collected from staff reporters of lead arrangers and SEC filings. I get the following variables from the Dealscan database: the loan share of the lead lenders, number of banks, loan size, loan purpose, maturity, and credit spread. Thanks to the link table provided by Chava and Roberts (2008), Dealscan can be merged with Compustat through the gvkey of the borrowers. In addition, there is a well-established link between CRSP and Compustat provided by WRDS, thus I am able to merge together the loan variables, balance sheet variables of the borrowers, and the stock return data. With Compustat, I include the following variables: sales, operating income margin, debt equity ratio, current ratio, net worth, credit rating, and leverage.

As argued in previous literature, the characteristics of the lenders also affect lending behavior. Unfortunately, Dealscan does not provide the information for the lender side of the loan transactions. Although some of the links from Chava and Roberts can be used for this analysis, there is not enough data because the link is mainly for borrowers. I can only identify the lenders in Chava and Roberts's link when the lender also borrows in Dealscan. Therefore, I have to hand-match the lenders in Dealscan with the firm in Compustat by names and locations. The identity of the lenders is aggregated to the parent firms if they are recorded as departments or subsidiaries in Dealscan. When handling M&A in which some lenders disappear or some lenders change names, I follow these rules: If one lender acquires another lender and does not change name after that, I use the acquirer's identity after the merger. If the two lenders merge and they found a new firm, I regard the combined firm after the merge as a completely different new firm. In addition, I use information from the lender's official website, Chicago Fed's BHC merging data, and the merge table generously provided by Mora (2014) and Ivashina (2009) as the reference. Since the loan deals concentrate on big lenders, hand-matching the top 100 lenders' loans along with the links provided by Chava and Roberts covers about half of the original sample size.

To ensure a reliable result, I drop the observation if the borrowers are in the utility or financial sector or if the borrower has an irregular value for its the balance sheet variable, such as a negative asset or negative net worth. For the stock return data, I only include the firms listed on NYSE, AMEX, and Nasdaq. After this merging and data-cleaning procedure, I have 15,168 loan contracts from 1987 to 2012.

2.4.3 Variables of Interest

As mentioned in the previous section, the uncertainty level is measured using stock return data. Here the stock return data is used in CRSP and is merged to Compustat by the link provided by WRDS. The dividend is included in the stock return. If a firm has more than one stock at the same time, then only the stock ID that is ranked first is used in the analysis. The credit spread is the difference between the charged interest rate and the base rate. In the Dealscan data, I use Allindrawn as the index of the credit spread. It sums up all of the fees related to a certain loan if the borrowers do in fact use the loan.[3]

Below, I define and describe the variables that potentially affect the credit spread. They are divided into three groups: loan variables, borrower variables, and lender variables. I use a braced "+" to indicate the positive impact of control variables, and a "-" to indicate a negative relationship. These relationships reflect the prediction of the theory and experiences.

Loan Variables:

Credit Spread (*): the amount the borrower pays in basis points over LIBOR for each dollar drawn down. It adds the spread of the loan with any annual (or facility) fee paid to the bank group.

Loan Size (-): Natural logarithm of the dollar value of the loan. If the loan includes different tranches with a different credit spread, then the loan size is the size of the individual tranche.

[3]The borrowers also need to pay an interest rate to lenders if they do not use the loan and they have a the credit line arrangement. In this scenario, the interest rate will be lower than if they use the loan from the lenders.

Maturity (+): Natural logarithm of the maturity measured by months.

Financial Purposes (+): Dummy that is equal to 1 if the loan is used for financial purposes, which includes an acquisition, LBO, IPO, merger, etc.

Borrower Variables:

Sales (-): Natural logarithm of the revenue of the borrowers.

Capital Return (+): Quarterly stock return including dividends.

Debt/Equity (+): (Long-term debt + debt in current liabilities)/shareholder equity.

Current Ratio (-): Total current assets/total current liabilities.

Net Worth (-): Natural logarithm of (asset-liability).

Leverage (+): Total asset/shareholder equity.

Credit Rating : S&P senior long-term debt rating.

Lender Variables:

Lender Capitalization (-): Lender shareholder equity/total assets. This is used to measure the abundance of the capital of banks.

Lender Total Asset (-): Natural logarithm of the total assets of the lender.

According to the methodology in Section 2.4.1, I can calculate the uncertainty level with respect to each loan. I can then rank the whole sample based on the uncertainty level, where a low uncertainty level refers to the first tertile, a medium uncertainty level refers to the second tertile, and a high uncertainty level refers to the third tertile. I calculate the means of the variables of the interest and report them in Table 2.2. Let us firstly take a look at the upper panel, which records the loan variables. It is obvious that uncertainty does matter for almost all of the variables. The credit spread rises from 129 bpt to 220 bpt, almost doubling, when uncertainty migrates from a low level to a high level. In addition, it is clear that with higher borrower uncertainty, the loan size will be smaller and maturity shorter.

The middle panel of Table 2.2 summarizes the borrower balance sheet variables. A higher uncertainty is related to smaller sales, a higher capital return, a higher debt equity ratio, current ratio, leverage, and lower net worth. The lower panel shows the lender's variables. It indicates that the lenders have less assets and a weaker balance sheet with higher uncertainty. All of these findings are consistent with conventional wisdom and the previous literature. If I plot the uncertainty level with the credit spread averaged over the deals within a quarter, I can find that the two time series move together, which is indicated in Figure 2.4.

2.5 Regression Evidence

2.5.1 Uncertainty and Credit Spread

Here we use a panel data regression model to investigate the relationship between uncertainty and credit spread. The dependent variable is the credit spread. The key independent variable is uncertainty built in Section 2.4.1. The controls include loan size, maturity, and loan purposes as the loan variables; sales, stock return, current ratio, net worth, leverage, and credit rating as the borrower side variables; lender asset and capitalization rate as the lender side variables. The econometric model follows the literature of the syndicated loan market and the results are reported in Table 2.4.

I use the uncertainty that is one period prior to the loan share to alleviate the problem of reverse causality. Since investors in the stock market have no public information on the credit spread until the deal is closed, it is mainly the change in uncertainty that causes the change in the credit spread rather than the opposite. In this setting, the problem of inconsistent estimation due to the mutual causality is alleviated.

I include the credit rating dummies for all of the econometric models in this paper since the cost of the loans in a real business environment is substantially affected by the credit rating . Here I use the S&P long-term credit rating as the proxy. The left panel of Table 2.4 reports the results of the model with only a credit rating fixed effect. It is important to note that the coefficient before uncertainty is positive under the 1% significance level. The coefficients for the controls are significant and the signs are consistent with my expectation. In this scenario, the coefficient for uncertainty is as high as 54. The middle panel turns on the time fixed effect as well. The coefficient of uncertainty has a similar magnitude and is also significant under the 1% level. The right panel turns on both the time fixed effect and the industry fixed effect. It can be seen that the R-squared is the highest in this case, which indicates an improvement of the fitness. The coefficient for the uncertainty is smaller compared to the middle panel and the coefficients of the controls have correct signs and are significant for the most part. Therefore, statistically, I find some evidence that higher uncertainty comes with a higher credit spread, which proves the validation of the model in the theory section.

I further check the magnitude of the coefficient to see whether this relationship is economically important. It can be calculated that the standard deviation of the uncertainty is 0.51, thus a standard deviation change of uncertainty will lead to a 22.6 bpt increase for the credit spread. Noting that the average credit spread is 170 bpt, this increase is equivalent to a 13% change of the average credit spread, which is economically important.

2.5.2 Size Effect

After knowing that uncertainty will negatively affect the credit spread, a matter of interest is in how this channel interacts with other characteristics of the borrowers. This leads to reasonable hypothesis; basically, when a firm has a very strong balance sheet, abundant capital, and low leverage, it is more likely that the lenders will not be too concerned about the uncertainty and will be willing to offer the loans at a lower rate. However, if a firm is in a terrible financial situation(i.e., they have low capital, high leverage, or low profitability), it makes sense that the lenders will have to charge more in case the borrower defaults. Although this sounds reasonable, the interaction effect must be confirmed by the data.

Based on my analysis, it is found that the size variables matter for the impact of uncertainty. Table 2.5 reports the interaction of uncertainty with different size measures of the borrowers. In the left panel, I add interaction term between net worth and uncertainty to the benchmark model. The middle panel adds the interaction with the borrower's sales; whereas the right panel adds the interaction with the loan size. The coefficients are all significantly negative and the magnitudes are economically important. The implication is straightforward: when net worth is higher, the loan size is larger, and revenue (sales) is higher, the uncertainty's impact on the credit spread is weakened. The current model fails to take into account this data patter and the fact that it can be important. It is well known that there is substantial heterogeneity among institutions in the market and that smaller institutions are under more pressure to raise capital. Once the heterogeneity of uncertainty's impact is shown and small borrowers, which are more vulnerable per se and are particularly more impacted by uncertainty, it is likely that the current theoretical model underestimates the macro impact of uncertainty and there is some potential to extend the existing model to capture this mechanism.

2.5.3 Uncertainty from the Lender Side

Until now I only focus on borrower uncertainty; however, it is natural to ask whether the uncertainty from the lender's side will also play a role. The question immediately following is how to measure uncertainty from the lender side. The most straightforward measure is to calculate the lender's uncertainty similar to the way we measure the borrower's: I run a Fama-French factor model with respect to the lender's stock return and extract the uncertainty from the standard deviation of the residuals. But there is a potential problem when I factor in the economic story behind the calculation. Even though I am able to find some relationship between the credit spread and lender's uncertainty, how do I explain this relationship? When explaining the relationship between the borrower's uncertainty and credit spread, I can simply reason that lenders are more cautious and worry about the variability of the borrower's project when a borrower is more uncertain, thus the lenders charge higher prices to compensate for the risk. When lenders' stock returns are more uncertain, it is not clear about how to map it into lending behavior. Actually, I also conduct an analysis using this type of uncertainty measure, but the results are not very robust.

Another way to measure the lender uncertainty is to use the uncertainty level of previous loans made by the lenders. The loan uncertainty is calculated as follows: For each lender at each time point, I record all of the loans it already made and still outstanding. Next, I calculate the uncertainty level for each borrower that receives the loan. Finally, I calculate the average of the uncertainty weighted by the size of the loan. The logic behind calculating the loan uncertainty in this way is as follows: The total loan amount that the lender is able to give out is limited by regulation or the lender's capital, thus the performance of previous loans will affect new loans. If the lenders realize that the previous loans turn out to be more uncertain, they may worry about the potential default and will have to behave more cautiously in order to reduce the credit risk. As a result, when some new borrowers reach out to ask for money, lenders tend to charge a higher interest rate even though the borrowers have good credit. Table 2.3 shows the sub-sample evidence of loan uncertainty. While the difference of the variables across uncertainty level groups are smaller than that of the borrower's uncertainty, there are still substantial changes in variables of interest when loan uncertainty varies. For instance, the credit spread rises by 18% above average when observations migrate from the low uncertainty group to the high uncertainty group.

The more convincing results comes from panel data analysis, which are summed in Table 2.6. The left panel, as before, is the benchmark model from Table 2.4. In the right panel, I add the loan uncertainty to the benchmark model. The coefficient of the loan uncertainty is significantly positive, whereas the magnitudes of the control variables do not change a lot, which confirms our hypothesis.

In the meantime, it is worth noting that the coefficient of loan uncertainty is around half of the borrower's uncertainty. It is not a surprising result in terms of economic intuition: When a loan deal is made, the primary factor determining the credit spread is naturally the more direct variable, such as the borrower's own uncertainty. Apart from that, indirect variables like the loan uncertainty or uncertainty from previous deals, plays a relatively less prominent role.

This empirical finding is important as it provides a potential new channel of uncertainty's impact, which is basically a contagious effect. In previous literature, researchers only focus on the borrower's uncertainty, but no study has been conducted on the uncertainty from other sources. This paper fills this gap by bringing to light the impact of the loan uncertainty. According to this channel, one firm's uncertainty can affect the otherwise irrelevant firm's credit spread through the lender's balance sheet; thus, uncertainty's adverse impact spreads from one firm to another. It gives us a new angle to understand why even good firms with low uncertainty still have to pay a high credit spread when the market is generally uncertain. Unfortunately, there is no model on this issue yet. I hope to further extend the CRM model to take this channel into account in a future study.

2.5.4 Cyclical Impact

Also of interest in this study is the cyclical property of uncertainty's impact. On the theoretical side, there are at least two hypotheses. One is an Uncertainty-aversion hypothesis. According to this hypothesis, it is more difficult for lenders to distinguish good firms from bad firms during an economic bust, thus the lenders will avoid signing a diversified contract with different borrowers and will instead provide deals with tighter restrictions for everyone. The implication of this theory is that uncertainty has more of an impact in good economic times. The other hypothesis is a credit ration theory. During a strong economy, banks do not need to care about uncertainty very much because they are more confident that they will get their money back; thus, borrowers with different levels of uncertainty will get similar types of loans in terms of credit spread, maturity, and covenants. On the other hand, lenders will have serious concerns about the credit worthiness of borrowers in a bad economy, and they worry that the borrowers with higher uncertainty will be more likely to default on their loans, which brings about a higher credit spread, shorter maturity, and more strict covenant.

In the following analysis, I validate the two hypotheses by introducing the interaction term between the GDP growth rate and uncertainty. For robustness, I explore both the borrower uncertainty and loan (lender) uncertainty. If the coefficient of the interaction term is positive, it implies that the gap of the credit spread charged to borrowers with different uncertainty levels widens in a good economy, which is consistent with the first hypothesis. If the coefficient of this term is negative, it means that the gap of the credit spread charged to borrowers with different uncertainty levels widens in a bad economy, which is consistent with the second hypothesis.

The results are summed in Table 2.7. Note that these models differ from a regular interaction term analysis in that the GDP growth rate is not included as the control variable. This is because I have already turned on the time fixed effect and the GDP growth only varies with time and is a redundant variable. In my robustness check, the GDP growth is included as a control variable as well, but the results only marginally change. The left panel of Table 2.7 is the benchmark model with both borrower and loan uncertainty. I add an interaction term between the borrower uncertainty and GDP growth in the middle panel and an interaction term between the loan uncertainty and growth rate in the right panel. Both of the regression results support the second theory: When the economy is in a downturn, the borrowers with higher uncertainty will run into more difficulty when applying for loans compared to their less uncertain competitors.

2.5.5 Different Measures of Uncertainty

In this section, I investigate the robustness of the results. The focus is the uncertainty measure. First, uncertainty can be calculated on different time horizons from the residuals. I try different length of the window, and the data patterns found in previous analysis persist. Table 2.8 and Table 2.10 report the regression results between uncertainty and credit spread for uncertainty calculated from a 1-year window. Other variables and econometric models remain the same. It can be seen that the magnitude of relationship I find is even greater than that of the benchmark model.

Second, uncertainty can be calculated from different factor models. In one scenario, I calculate the uncertainty strictly following Gilchrist et al. (2014)'s approach and do not include VXO as a factor. The result of the regressions are almost the same as the benchmark models. In another scenario, I don't include any factor structure at all, and uncertainty is measured by the simple volatility of stock return. The results of the regressions between uncertainty and credit spread are reported In Table 2.9 and Table 2.11. Note that the uncertainty is calculated from a 1-quarter window, and econometric models remain the same. It is important to note that the relationship found in previous models persist while the magnitude is smaller.

2.6 Conclusion

This paper adds to the literature on the mechanism of how uncertainty affects the economy. While there has been a large amount of literature analyzing this issue through macro level data, a thorough empirical analysis on uncertainty's impact using loan-level data is limited. I provide empirical evidence using the merged database of Dealscan, Compustat, and CRSP. It is found that the credit spread is higher for more uncertain borrowers, which confirms the empirical results from the bond data. This paper also looks at the interaction between uncertainty and other determinants of credit spread and finds a significant size effect. When a borrower has larger sales, a greater net worth, or receives a larger loan size, the impact of uncertainty is weakened. In addition, this paper studies uncertainty from the lender side, which is ignored by the previous literature. I find evidence that lender uncertainty will play a role, but the magnitude is not as large as the borrower side uncertainty. Finally, this paper studies the cyclical impact of uncertainty and backs the credit ration hypothesis: When the economy is in a downturn, uncertainty's impact is more substantial.

A future extension to this paper includes a more advanced econometric method. The GARCH model can be used to get a better estimate of the uncertainty and a GMM model can be used to check the causality. In the meantime, most empirical findings in this paper have not been well modeled by current theory, thus I may conduct further theoretical works in the future on the lender side uncertainty and cyclical property.

References

Burkart M., D. Gromb, and F. Panunzi. 1997. Large Shareholders, Monitoring, and the Value of the Firm. The Quarterly Journal of Economics, 693-728.

Chava S., and M. R. Roberts. 2008. How Does Financing Impact Investment? The Role of Debt Covenants. The Journal of Finance, 63 (5), 2085-2121.

Chava S., and A. Purnanandam. 2011. The Effect of Banking Crisis on Bank-dependent Borrowers. Journal of Financial Economics, 99 (1), 116-135.

Chodorow-Reich G. 2014. The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008–9 Financial Crisis. The Quarterly Journal of Economics, 129 (1), 1-59.

Dennis S. A., and D. J. Mullineaux. 2000. Syndicated Loans. Journal of Financial Intermediation, 9, 404-426.

Diamond D. 1984. Financial Intermediation and Delegated Monitoring. Review of Economic Studies, 51 (3), 393-414.

Güner A.B. 2006. Loan Sales and the Cost of Corporate Borrowing. Review of Financial Studies, 19 (2), 687-716.

Gilchrist S., J. W. Sim, and E. Zakrajsek. 2014. Uncertainty, Financial Frictions, and Investment Dynamics. NBER Working Papers 20038.

Holmstrom B., and J. Tirole. 1997. Financial Intermediation, Loanable Funds, and the Real Sector. The Quarterly Journal of Economics, 663-691.

Hubbard R. G., K. N. Kuttner, and D. N. Palia. 2002. Are There Bank Effects in Borrowers' Costs of Funds? Evidence From a Matched Sample of Borrowers and Banks. The Journal of Business, 75 (4), 559-581.

Ivashina V. 2009. Asymmetric Information Effects on Loan Spreads. Journal of Financial Economics, 92, 300-319.

Ivashina V., and D. Scharfstein. 2010. Bank Lending during the Financial Crisis of 2008. Journal of Financial Economics, 97, 319-338

Jones J., W. Lang, and P. Nigro. 2005. Agent Bank Behavior in Bank Loan Syndication. Journal of Financial Research, 28, 385–402.

Lee S. W., and D. J. Mullineaux. 2004. Monitoring, Financial Distress, and the Structure of Commercial Lending Syndicates. Financial Management, 33, 107-130.

Leland H., and D. Pyle. 1977. Informational Asymmetries, Financial Structure, and Financial Intermediation. Journal of Finance, 32 (2), 371-387.

Mora N. 2014. Lender Exposure and Effort in the Syndicated Loan Market. Journal of Risk and Insurance, 82 (1), 205-252.

Murfin J. 2012. The Supply-side Determinants of Loan Contract Strictness. The Journal of Finance, 67 (5), 1565-1601.

Simons K. 1993. Why Do Banks Syndicate Loans? New England Economic Review of the Federal Reserve Bank of Boston, (Jan), 45-52.

Sufi A. 2007. Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans. The Journal of Finance, 62 (2), 629-668.

Standard and Poor. 2006. A Guide to the Loan Market. McGraw-Hill Companies, Inc.

Tables and Figures

Figure 2.1: Characteristic Functions

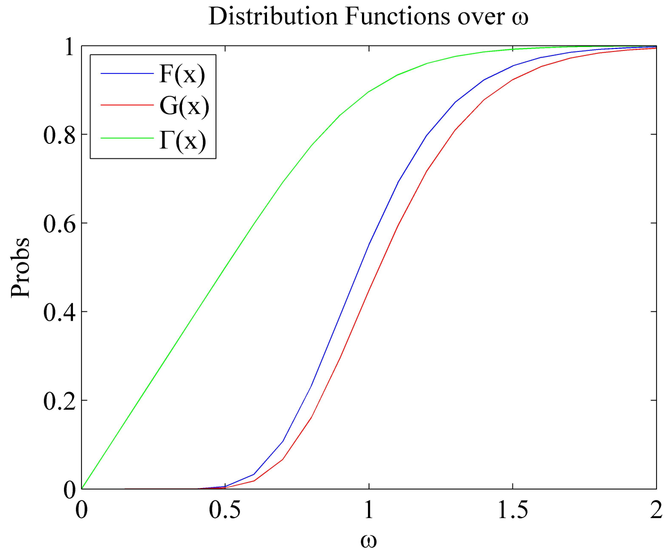

Figure 2.2: Impact of Uncertainty: Panel A

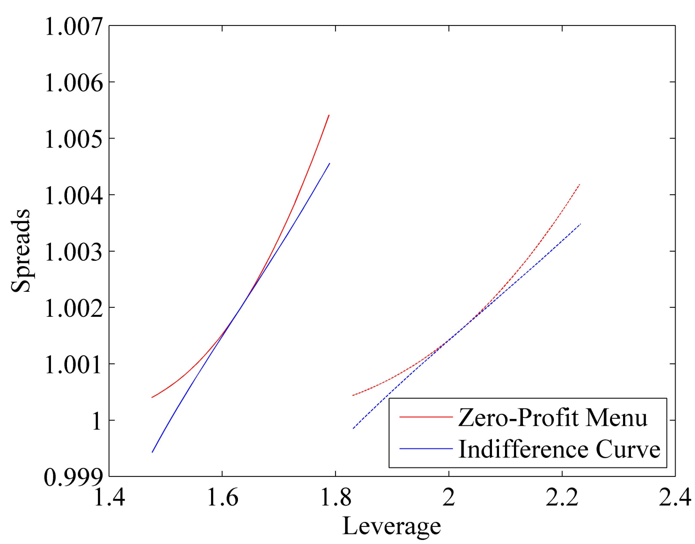

Figure 2.3: Impact of Uncertainty: Panel B

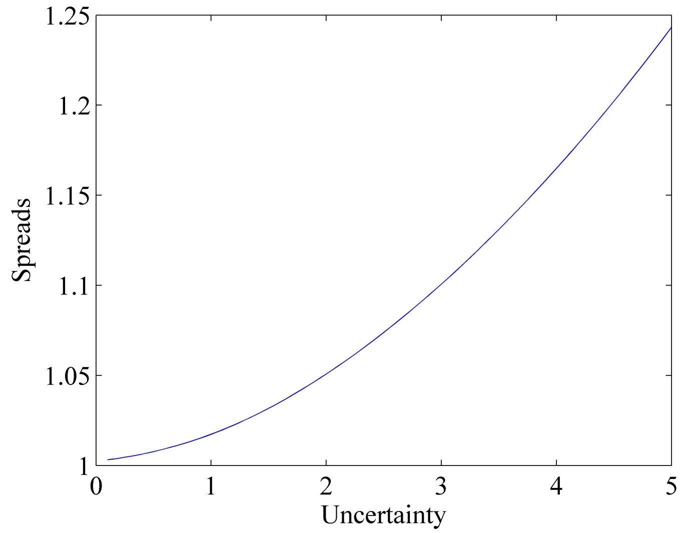

Figure 2.4: Uncertainty and Credit Spread

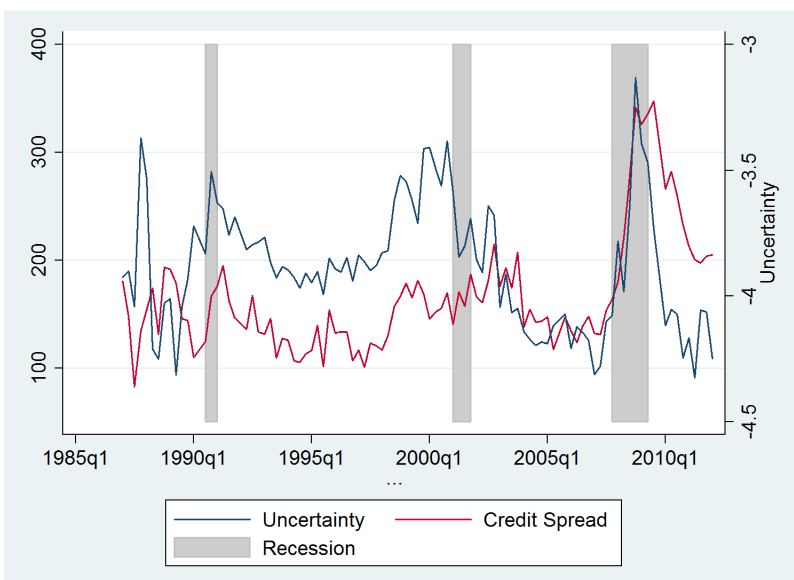

Figure 2.5: Loan Uncertainty and Credit Spread

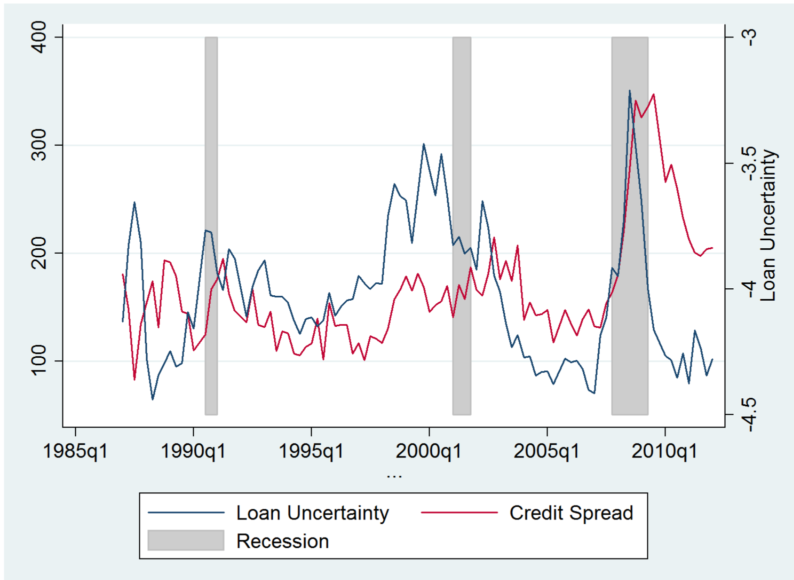

Table 2.1: Descriptive Statistics

Variable	Mean	Std. Dec.	10%	50%	90%
Credit Spread	170.47	121.11	35.00	150.00	325.00
Uncertainty	-3.89	0.51	-4.52	-3.91	-3.24
Loan Size	396M	830M	30M	180M	938M
Maturity	49.42	23.36	12.00	60.00	74.00
Sales (log)	5.68	1.58	3.77	5.60	7.82
Capital Return	0.05	0.25	-0.22	0.03	0.32
Debt Equity Ratio	1.21	8.72	0.09	0.69	2.46
Current Ratio	1.87	1.02	0.86	1.64	3.12
Net Worth (log)	6.14	1.53	4.26	6.05	8.21
Leverage	3.11	2.88	1.48	2.37	4.89
Lender Asset (log)	12.91	1.24	11.22	13.09	14.52
Capitalization Rate	0.07	0.02	0.04	0.08	0.10

Table 2.2: Sub-sample Evidence by Uncertainty Level

	Uncertainty Level		
Variable	Low	Medium	High
Credit Spread	129.37	165.79	220.53
Loan Size	575M	328M	202M
Maturity	50.96	50.14	47.65
Sales (log)	6.35	5.59	4.95
Capital Return	0.04	0.05	0.05
Debt Equity Ratio	1.01	1.18	1.68
Current Ratio	1.76	1.88	2.02
Net Worth (log)	6.87	6.05	5.35
Leverage	2.88	2.98	3.32
Lender Asset (log)	13.25	12.86	12.60
Capitalization Rate	0.08	0.07	0.07

Table 2.3: Lender's Uncertainty

	Uncertainty Level		
Variables	Low	Medium	High
Credit Spread	152.34	164.58	186.97
Loan Size	525M	349M	314M
Maturity	53.72	49.16	45.13
Sales (log)	6.08	5.56	5.38
Capital Return	0.07	0.06	0.02
Debt Equity Ratio	0.93	1.50	1.21
Current Ratio	1.83	1.89	1.88
Net Worth (log)	6.56	5.98	5.84
Leverage	3.15	3.17	3.01
Lender Asset (log)	13.40	12.54	12.55
Capitalization Rate	0.08	0.07	0.07

Table 2.4: Uncertainty and Credit Spread

Variables	Model I	Model II	Model III
Uncertainty	53.74***	55.07***	44.05***
	(26.73)	(27.34)	(21.16)
Loan Size (log)	-17.74***	-16.40***	-15.28***
	(-16.84)	(-16.93)	(-15.79)
Maturity (log)	14.00***	15.37***	12.36***
	(9.764)	(11.34)	(9.282)
Sales (log)	-17.61***	-14.25***	-13.67***
	(-14.81)	(-13.13)	(-8.765)
Current Ratio	-4.161***	-6.695***	-7.510***
	(-4.541)	(-7.988)	(-7.667)
Leverage	6.676***	5.395***	3.989***
	(17.00)	(15.01)	(10.09)
Net Worth (log)	5.817***	0.664	-5.455***
	(4.008)	(0.499)	(-3.281)
Stock Return	0.706	-8.548**	-7.934**
	(0.196)	(-2.564)	(-2.450)
Lender Asset (log)	23.42***	-4.889***	-4.397***
	(29.56)	(-4.213)	(-3.754)
Capitalization Rate	286.3***	-495.7***	-440.6***
	(6.665)	(-11.18)	(-9.862)
Financial Purpose	1.631***	1.591***	1.262***
Credit Rating	Yes	Yes	Yes
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.309	0.430	0.504

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 2.5: Uncertainty and Size Effect

Variables	Model IV	Model V	Model VI
Uncertainty	43.08***	43.14***	43.22***
	(20.64)	(20.66)	(20.77)
Uncertainty*Net Worth	-5.065***		
	(-5.071)		
Uncertainty*Sales		-4.725***	
		(-4.668)	
Uncertainty*Loan Size			-7.919***
			(-6.713)
Loan Size (log)	-15.14***	-15.14***	-15.30***
	(-15.65)	(-15.65)	(-15.84)
Maturity (log)	12.57***	12.45***	12.76***
	(9.442)	(9.352)	(9.590)
Sales (log)	-13.79***	-13.75***	-13.74***
	(-8.850)	(-8.828)	(-8.828)
Current Ratio	-7.364***	-7.381***	-7.277***
	(-7.521)	(-7.538)	(-7.437)
Leverage	3.941***	4.040***	4.038***
	(9.977)	(10.23)	(10.23)
Net Worth (log)	-5.430***	-5.433***	-5.311***
	(-3.270)	(-3.271)	(-3.200)
Stock Return	-9.229***	-8.779***	-8.997***
	(-2.844)	(-2.709)	(-2.780)
Lender Asset (log)	-4.294***	-4.307***	-4.243***
	(-3.670)	(-3.681)	(-3.629)
Capitalization Rate	-440.9***	-440.2***	-446.4***
	(-9.877)	(-9.861)	(-10.01)
Financial Purpose	1.631***	1.591***	1.262***
Credit Rating	Yes	Yes	Yes
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.309	0.430	0.504

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 2.6: Uncertainty from Lender Side

Variables	Model VII	Model VIII
Uncertainty(borrower)	44.05***	42.94***
	(21.16)	(20.61)
Uncertainty(loan)		24.88***
		(6.721)
Loan Size (log)	-15.28***	-15.06***
	(-15.79)	(-15.59)
Maturity (log)	12.36***	12.33***
	(9.282)	(9.272)
Sales (log)	-13.67***	-13.56***
	(-8.765)	(-8.712)
Current Ratio	-7.510***	-7.565***
	(-7.667)	(-7.737)
Leverage	3.989***	4.058***
	(10.09)	(10.28)
Net Worth (log)	-5.455***	-5.181***
	(-3.281)	(-3.121)
Stock Return	-7.934**	-6.323*
	(-2.450)	(-1.951)
Lender Asset (log)	-4.397***	-2.946**
	(-3.754)	(-2.478)
Capitalization Rate	-440.6***	-418.2***
	(-9.862)	(-9.351)
Financial Purpose	1.631***	1.591***
Credit Rating	Yes	Yes
Time FE	No	Yes
Industry FE	No	No
R-squared	0.504	0.506

Robust t-statistics in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.7: Uncertainty and Credit Spread Cycles

Variables	Model IX	Model X	Model XI
Uncertainty(borrower)	42.94***	42.36***	42.65***
	(20.61)	(20.27)	(20.46)
Uncertainty (loan)	24.88***	24.03***	22.97***
	(6.721)	(6.481)	(6.151)
Uncertainty(borrower)*dy		-9.188***	
		(-3.496)	
Uncertainty(loan)*dy			-15.18***
			(-3.802)
Loan Size (log)	-15.06***	-15.08***	-15.11***
	(-15.59)	(-15.61)	(-15.64)
Maturity (log)	12.33***	12.36***	12.46***
	(9.272)	(9.302)	(9.373)
Sales (log)	-13.56***	-13.61***	-13.49***
	(-8.712)	(-8.748)	(-8.671)
Current Ratio	-7.565***	-7.530***	-7.591***
	(-7.737)	(-7.704)	(-7.767)
Leverage	4.058***	4.084***	4.050***
	(10.28)	(10.35)	(10.27)
Net Worth (log)	-5.181***	-5.205***	-5.298***
	(-3.121)	(-3.137)	(-3.193)
Stock Return	-6.323*	-5.476*	-5.540*
	(-1.951)	(-1.686)	(-1.707)
Lender Asset (log)	-2.946**	-2.997**	-2.948**
	(-2.478)	(-2.522)	(-2.482)
Capitalization Rate	-418.2***	-420.7***	-418.4***
	(-9.351)	(-9.410)	(-9.360)
Financial Purpose	1.631***	1.591***	1.262***
Credit Rating	Yes	Yes	Yes
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.309	0.430	0.504

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix

Table 2.8: Uncertainty with 1-year Window and Credit Spread

Variables	Model I	Model II	Model III
Uncertainty	85.02***	85.90***	74.50***
	(37.32)	(35.95)	(28.84)
Loan Size (log)	-15.56***	-14.96***	-14.30***
	(-15.09)	(-15.73)	(-14.98)
Maturity (log)	16.38***	14.91***	12.34***
	(11.70)	(11.23)	(9.412)
Sales (log)	-16.56***	-13.60***	-12.17***
	(-14.27)	(-12.78)	(-7.916)
Current Ratio	-4.654***	-7.270***	-7.564***
	(-5.208)	(-8.848)	(-7.840)
Leverage	6.131***	5.168***	3.743***
	(15.99)	(14.67)	(9.611)
Net Worth (log)	7.567***	2.879**	-3.067*
	(5.346)	(2.202)	(-1.870)
Stock Return	-4.461	-11.55***	-10.43***
	(-1.273)	(-3.534)	(-3.270)
Lender Asset (log)	23.85***	-4.484***	-4.009***
	(30.87)	(-3.944)	(-3.475)
Capitalization Rate	315.2***	-479.9***	-421.6***
	(7.522)	(-11.04)	(-9.579)
Financial Purpose	1.631***	1.591***	1.262***
Credit Rating	Yes	Yes	Yes
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.343	0.453	0.518

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Simple Volatility and Credit Spread

Variables	Model I	Model II	Model III
Uncertainty	55.78***	48.11***	38.60***
	(28.08)	(23.95)	(18.52)
Loan Size (log)	-17.75***	-16.57***	-15.35***
	(-16.91)	(-17.00)	(-15.80)
Maturity (log)	13.90***	15.16***	12.15***
	(9.729)	(11.11)	(9.082)
Sales (log)	-17.66***	-14.52***	-14.11***
	(-14.90)	(-13.30)	(-9.012)
Current Ratio	-4.893***	-6.909***	-7.597***
	(-5.347)	(-8.177)	(-7.721)
Leverage	6.644***	5.429***	4.027***
	(16.97)	(15.00)	(10.14)
Net Worth (log)	4.600***	-1.019	-6.782***
	(3.190)	(-0.764)	(-4.075)
Stock Return	4.297	-5.731*	-5.841*
	(1.197)	(-1.707)	(-1.795)
Lender Asset (log)	21.82***	-4.885***	-4.394***
	(27.67)	(-4.182)	(-3.736)
Capitalization Rate	229.2***	-500.1***	-445.5***
	(5.359)	(-11.20)	(-9.930)
Financial Purpose	1.631***	1.591***	1.262***
Credit Rating	Yes	Yes	Yes
Time FE	No	Yes	Yes
Industry FE	No	No	Yes
R-squared	0.313	0.422	0.499

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Uncertainty with 1-year Window from Lender Side and Credit Spread

Variables	Model VII	Model VIII
Uncertainty (borrower)	74.50***	73.25***
	(28.84)	(27.72)
Uncertainty (loan)		32.58***
		(7.557)
Loan Size (log)	-14.30***	-14.23***
	(-14.98)	(-14.66)
Maturity (log)	12.34***	12.58***
	(9.412)	(9.555)
Sales (log)	-12.17***	-12.07***
	(-7.916)	(-7.725)
Current Ratio	-7.564***	-7.466***
	(-7.840)	(-7.606)
Leverage	3.743***	3.729***
	(9.611)	(9.347)
Net Worth (log)	-3.067*	-3.016*
	(-1.870)	(-1.808)
Stock Return	-10.43***	-9.346***
	(-3.270)	(-2.882)
Lender Asset (log)	-4.009***	-1.731
	(-3.475)	(-1.436)
Capitalization Rate	-421.6***	-388.1***
	(-9.579)	(-8.637)
Financial Purpose	1.631***	1.591***
Credit Rating	Yes	Yes
Time FE	No	Yes
Industry FE	No	No
R-squared	0.518	0.518

Robust t-statistics in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.11: Simple Volatility from Lender Side and Credit Spread

Variables	Model VII	Model VIII
Uncertainty(borrower)	38.60***	37.01***
	(18.52)	(17.38)
Uncertainty(loan)		20.51***
		(5.762)
Loan Size (log)	-15.35***	-15.42***
	(-15.80)	(-15.57)
Maturity (log)	12.15***	12.45***
	(9.082)	(9.254)
Sales (log)	-14.11***	-14.09***
	(-9.012)	(-8.834)
Current Ratio	-7.597***	-7.495***
	(-7.721)	(-7.470)
Leverage	4.027***	3.976***
	(10.14)	(9.758)
Net Worth (log)	-6.782***	-6.860***
	(-4.075)	(-4.044)
Stock Return	-5.841*	-4.004
	(-1.795)	(-1.203)
Lender Asset (log)	-4.394***	-3.234***
	(-3.736)	(-2.672)
Capitalization Rate	-445.5***	-432.0***
	(-9.930)	(-9.450)
Financial Purpose	1.631***	1.591***
Credit Rating	Yes	Yes
Time FE	No	Yes
Industry FE	No	No
R-squared	0.504	0.506

Robust t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

3 Forecasting China's Economy: A Bayesian Approach

Abstract

While it is well known that many models used for Western economies do not perform well in explaining and forecasting China's economic data, I challenge this convention by building rigorous econometric forecasting models for China's economy. Different Bayesian Vector-autoregression (BVAR) models are built, revised, and evaluated. I also take advantage of the state-of-the-art methodology to further improve the model's performance. I find that utilizing the richer data set of additional macroeconomic data and sectoral data helps forecast GDP, CPI, and the interest rate. The large-scale BVAR model with 124 variables is the champion of my forecasting models and it is more effective in extracting information from a large database rather than factor models. Besides, it is found that the more theoretically grounded hierarchical BVAR model, which is supposed to optimize the values of prior parameters, is not very effective in improving forecasting performance.

3.1 Introduction

The importance of China's economy for the world can never be overstated and it has drawn a great deal of attention from both academia and industry. While researchers have made great efforts to study China's economy, many theories and models widely used in Western economies fail to work on China. Admittedly, this might be due to the poor quality of data, which ruins the foundation of successful economic analyses. Economists have been suspicious of the accuracy of Chinese official statistics for decades. However, it would be too pessimistic to conclude that China's economy can neither be well-forecasted nor explained by a rigorous economic model. First, as argued by Holz (2014), no compelling evidence has been discovered that the data is manipulated by the Chinese government and any accusation of such is by and large anecdotal. Second, even if the data released by the Chinese government have been manipulated, we can still build an effective model as long as the manipulation is not systemically biased. Moreover, if richer data are utilized, it is even possible to substantially alleviate the impact of the noise brought by such a manipulation. Third, the alternative cause of previous frustration makes sense as well: we simply haven't found the right model and the models used in Western economies have to be revised before they can be applied to China. There is no doubt that China's economy differs substantially from Western economies; thus it would be surprising if the model built for a Western economy can work for China in the absence of any change.

In light of the reasons above, the primary goal of this paper is to find a good econometric forecasting model for China. A group of econometric models, including the most state-of-the-art ones, are built and evaluated. I will also discuss the ways of improving these models specifically for China, which delivers promising forecasting performance.

The realm of the models I explore are Bayesian Auto-regression (BVAR) models, which are widely used in empirical macroeconomic research. The idea of BVAR modeling is to integrate the information from the Bayesian prior beliefs with the information from the data in order to shrink the dimension and alleviate the over-fitting problem suffered by simple VAR models. Although technically speaking, the priors can take a variety of forms, I start this analysis on one specific type of priors: the Minnesota priors advocated by Litterman (1980, 1986) and Doan et al. (1984), which deliver good forecast results and have been improved upon continuously by researchers in the past 30 years. When applied to the U.S. economy and other developed economies, it is found that BVAR models based on the Minnesota priors substantially outperform the VAR models thus reflecting the advantage of dimensionality shrinkage. However, the direct application of these models to China is frustrating: the standard BVAR model cannot even outperform the simple AR model and Random Walk model when forecasting some variables.

With an aim to improve the forecasting performance, I explore some of the most state-of-the-art methodologies. The motivation of these methodologies is to resolve the defect of the standard BVAR model: A typical BVAR model only includes 8 to 10 variables, which may not capture the comprehensive picture of the rich dynamics. One compelling example is the "prize puzzle" documented by Sims (1992) and Christiano et al. (1999).

They find that the price level's response to the interest rate violates the economic principle in a standard BVAR model and that adding more variables can alleviate these abnormalities. Following this logic, adding more variables or more information to the framework could improve the forecasting as well. However, this extension is not trivial because the standard model specification might need to be adjusted to accommodate the larger database. At the moment, there are three popular methodologies to incorporate a rich database to the BVAR models: 1) Large BVAR model; 2) Bayesian Factor-Augmented VAR (BFAVAR) model; and 3) Hierarchical BVAR model. I will document the details of the models in the following sections. All of these models do a decent job in forecasting the U.S. economy whereas the application to China is still an open question needing to be addressed.

As a preview, I find that a simple revision of parameter values in a small Sims&Zha model returns good results, which is very promising. More importantly, the large database contains important information that helps improve the model's performance. A Large BVAR model built from a BGR algorithm turns out to be the most effective. I also find ways of further improving the model's performance by proposing a version of the Large BVAR model for China, which delivers very nice forecasts for GDP, CPI, and interest rate and is the champion of all models. Concerning the BFAVAR model, I find that the model's performance is unbalanced: it does a better job in forecasting inflation and a worse job in forecasting GDP and the interest rate than the Large BVAR model. With respect to the Hierarchical model, it is found that the automatic prior optimization procedure works well for a smaller number of variables but does not function well when applied to the large database, which is the key reason why this approach cannot outperform the Large BVAR models. In addition, adding flexibility of the prior hyper parameters helps improve performance.

This paper contributes to the literature in the following aspects. First, this is the first paper applying and evaluating the most state-of-the-art BVAR forecasting methodologies to China's economy. It also rigorously documents the ways to further improve the model's performance of China's economy. Second, I find solid evidence that a rich database of macroeconomic, financial, and sectoral data helps improve the forecasting performance of key macro variables. Last but not least, I find some useful models for China, which delivers good forecasting results for GDP, inflation, and interest rate and outperform or at least stay competitive with the business forecasts.

The rest of the paper is organized as follows. Section 3.2 reviews the related literature, Section 3.3 documents the theories of the models used for forecasting, Section 3.4 describes the data, Section 3.5 reports the empirical results and provides comments and analyses, and Section 3.6 concludes the paper.

3.2 Literature Review

The BVAR literature can be traced back to the pioneering work of Sims (1980) on the VAR model, which is a flexible methodology that takes into account the interrelationships of macro variables. However, the VAR framework usually suffers from over-fitting problems, which results in an unreliable estimation and imprecise forecasts. As such, early researchers like Doan et al. (1984) and Litterman (1986) introduce the informative Bayesian priors to the VAR system. Their intuition is as follows: the VAR system encounters on over-fitting problem because the model asks too much information from the data. In the BVAR system, the data is only asked to provide part of the information; the rest of the information is provided by the informative Bayesian priors, thus alleviating the over-fitting problem. Their priors specification gained popularity in academia and was dubbed the Minnesota priors since most of the advocates come from the Minnesota Fed and University of Minnesota. This group of researchers continue to revise and sharpen the prior specification and many variants of the Minnesota priors have been proposed. Kadiyala and Karlsson (1997) make the prior's variance shrinkage with respect to lags faster and document the numerical methods for the Normal-Inverse-Wishart Distribution Priors, which is the foundation of the methodologies I utilize in this paper. The dummy observation priors that favor unit-root and cointegration are introduced by Doan et al. (1984), Sims (1992, 1993), and Sims and Zha (1998). These priors improve the forecast performance by taking into account the unit-root and cointegration characteristics of the macro variables. The rigorous methods of incorporating Bayesian priors to a structural VAR are proposed by Sims and Zha (1998), which further enhances the usage of Bayesian methods in macroeconomics. Koop and Korobilis (2010) write a comprehensive survey on the BVAR literature that covers both theory and the empirical uses of VAR, the time-varying-parameter VAR, BVAR, FAVAR, and BVAR with multivariate stochastic volatility.

As an important variant of the BVAR model, the large-scale Bayesian VAR model can be traced back to De Mol et al. (2008)'s work. They find that the model with Gaussian and double-exponential Bayesian priors deliver the similar forecasts as the principle component model. Moreover, they present an asymptotic property showing that Bayesian regression models tend to capture the common factors of data as the cross-section dimension increases for large panel data with a strong collinearity. Their theoretical result actually lays the foundation for a heuristic idea further developed by Banbura et al. (2010): the overall tightness of priors, τ, should be picked in relation with the number of variables in order to improve the forecasting performance of the model. Banbura et al. (2010) propose an empirical large BVAR model for U.S. monthly economic data by incorporating the macroeconomic variables and sectoral information. The overall tightness of priors is adjusted to keep the in-sample fitness of the model constant in the training sample. Banura et al. set up the Bayesian model with different scales: a small model with seven variable, a medium model with 20 variables, and a large model with 131 variables. It is found that the large model outperforms the smaller models and the improvement is biggest from small to medium size. Besides, the large model outperforms the factor model as well. The authors also apply the framework to the structural analysis of analyzing the monetary policy shock's impact. The impulse response functions are found to be robust for both the medium-size model and large model.

The FAVAR, or Factor-augmented VAR model is also a variant of the standard VAR model to utilize a larger database. Given that a typical VAR model only includes 8 to 10 variables, Bernanke et al. (2005) propose using factors extracted from a larger dataset instead of using the raw series. The authors estimate the model through both a two-step principal components approach and a Bayesian likelihood approach and the main results are robust across these approaches. They find that the FAVAR model produces a more plausible impulse response function of monetary policy shocks compared to a standard VAR model and thus helps solve the "price puzzle." The FAVAR framework is also applied to China's economy to analyze the monetary policy's impact. He et al. (2014) build a FAVAR model to study the effectiveness of monetary policy instruments. They find that market-based instruments like the repo rate and benchmark lending rate have little impact on the real economy, whereas non-market-based instruments like the total loan and money supply affects the real economy and inflation significantly. Fernald et al. (2014) build a dynamic FAVAR model for China's monthly data and present opposite results to He et al.'s work. The primary reason for this contradiction is that the authors calculate the real economic growth rate and inflation as factors from a set of variables instead of using the raw variables. They find that the interest rate and bank reserve requirement have substantial impacts on the economic activity and inflation in China. Interestingly, they find that the financial variables that are commonly believed to play an important role, like M2 and lending levels, do not have a significant impact once the policy variables are controlled. Their results imply that China's monetary policy's transmission channel is closer to the Western economies in recent years than what was previously thought.

Another way of improving the model's performance is to optimize the informativeness of priors. Giannone et al. (2015) propose a hierarchical model to add another layer of hyperpriors. In this model, the standard Minnesota prior's hyper parameters are characterized by the hyperpriors' distributions. The authors derive the closed form expression of the marginal likelihood (ML), which represents the in-sample fit and the penalty for the model's complexity at the same time. The empirical results in their paper shows that an estimation by maximizing the ML can deliver a better out-of-sample forecasting performance than both benchmark BVAR models and factor models. The authors explain that the advantage is due to the fact that a hierarchical model is able to pick the optimal values of the prior's parameters automatically for each round of estimation. This approach is superior to Banbura et al. (2010)'s method in two aspects. First, it is less ad hoc. The hierarchical model is able to pick many hyperparameters by maximizing the likelihood functions whereas the model created by Banbura et al. needs to arbitrarily pin down the parameters, with the exception of the overall tightness parameter, and must also pick the training sample. Therefore, it is likely that Giannone et al.'s approach can be applied to other economies whereas Banbura et al.'s approach is specific to the current U.S. economy. Second, the hierarchical model is more systematically consistent. Even though Banbura et al.'s model is found to be effective, it is still a heuristic approach at the end of the day. To the contrary, the likelihood maximization process is well-founded in theory.

While BVAR models have been widely used in forecasting Western economies, their application to China is very rudimentary. Summers and Zhang (1998) give a short introduction of the BVAR model and build a sample model for the Chinese data; however, no evaluation is conducted. Wang (2011) builds a BVAR model for the regional economy in Qinghai Province and finds that BVAR models outperform ARIMA models and VAR models in terms of forecasting. The BVAR model includes 4 variables and is built on annual data using the original Minnesota Prior. Kuang and Zhou (2015) build a BVAR model using Sims&Zha prior specification and find that the BVAR model's forecast outperforms that of VAR models; they also find that the BVAR's impulse response function of the monetary policy shock is more reliable. Higgins et al. (2016) build a monthly BVAR model for both unconditional and conditional forecasts. The results show that the BVAR model outperforms the AR model and Random Walk model generally in forecasting GDP, inflation, and interest rate. This is a remarkable achievement because the Random Walk model is believed to be the gold standard in forecasting China's economy. With respect to the conditional forecast, the authors find that if the investment growth rate slows down to 10%, the GDP growth rate would reduce to 6.2%, which implies that the impact of the recent policy of curbing investment is limited and under control.

3.3 Methodology

3.3.1 Benchmark BVAR Model

Let $y_t = [y_{1,t}, y_{1,t}, ...y_{n,t}]'$ refer to the vector whose elements are macro variables at period t. I can write down the VAR(p) model as follows:

$$y_t = C + B^{(1)}y_{t-1} + ...B^{(p)}y_{t-p} + \varepsilon_t,\ t = 1, 2, ..., T, \tag{3.1}$$

where the n-dimension error vector follows a multinomial distribution as $\varepsilon_t \sim N(0, \Sigma)$. The VAR models can simply be estimated by maximizing the likelihood function and the forecast of y of any period in the future can be calculated by plugging its corresponding lags and coefficient estimates in the model representation. While widely used, this simple model usually suffers from the curse of dimensionality, which causes the over-fitting of parameters and poor out-of-sample forecasts. In order to overcome these shortcomings, researchers propose the Bayesian VAR (BVAR) model. The idea is to introduce a reasonable prior belief in order to provide information as the complement to the information provided by data. Technically speaking, one can use any type of priors in this framework. However, in order to simplify the inference and expedite the calculation, there is some specific prior specification that is more often used. In this paper, I will focus on the conventional

Minnesota Priors setup proposed by Doan et al. (1984), Litterman (1986), Kadiyala and Karlsson (1997), and Sims and Zha (1998).

To better represent the Bayesian distributions, I rewrite Eq (3.1) as follows:

$$Y = XB + E, \tag{3.2}$$

where $B \equiv [C, B^{(1)}, ...B^{(p)}]'$, $\beta = vec(B)$, $E \equiv [\varepsilon_1, \varepsilon_2, ...\varepsilon_T]'$, $Y \equiv [y_1, y_2, ...y_T]'$, $E \equiv [\varepsilon_1, \varepsilon_2, ...\varepsilon_T]'$, $x_t = [1, y'_{t-1}, y'_{t-2}, ..., y'_{t-p}]$, and $X = [x'_1, x'_2, , ..., x'_T]$. Y and X are called the data matrix since they store the data of all periods. β is the coefficient vector with a dimension equal to $(np+1)n \times 1$ and each element in β corresponds to one element in the coefficient matrices. In addition, $\theta \equiv [C, B^{(1)}, B^{(2)}, ...B^{(p)}, \Sigma]$, $p(\theta)$ is the prior distribution, $p(y|\theta)$ is the conditional distribution of the data, and $p(\theta|y)$ is the posterior distribution. We have the following Bayesian Rule:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{3.3}$$

It can be proved that if priors follow distributions as below,

$$\beta \mid \Sigma \sim N(\underline{\beta}, \Sigma \otimes \underline{\Omega}) \text{ , and } \Sigma^{-1} \sim W(\underline{S}^{-1}, d) \text{ , } d = n+2, \tag{3.4}$$

then following Eq (3.3), the posteriors follow exactly the same classes of distributions as the priors. They are presented as follows:

$$\beta \mid \Sigma \sim N(\bar{\beta}, \Sigma \otimes \bar{\Omega}) \text{ , and } \Sigma^{-1} \sim W(\bar{S}^{-1}, d+T+2), \tag{3.5}$$

where $\bar{\beta} = vec(\bar{B})$, $\bar{B} = [\underline{\Omega}^{-1} + X'X]^{-1}[X'Y + \underline{\Omega}^{-1}\underline{B}]$, $\bar{\Omega} = [\underline{\Omega}^{-1} + X'X]^{-1}$, $\bar{S} = \underline{S} + \hat{E}'\hat{E} + (\bar{B} - \underline{B})'\underline{\Omega}^{-1}(\bar{B} - \underline{B})$, and $\hat{E} = Y - X * [(X'X)^{-1}(X'Y)]$. The parameters in the posterior distributions ($\bar{\beta}$, $\bar{\Omega}$, $\bar{S}$) can be written as the functions of the parameters in the prior distributions ($\underline{\beta}$, $\underline{\Omega}$, $\underline{S}$) and data matrices (X, Y), which implies that the posterior distribution integrates the information from the prior beliefs and data. This prior setup is very handy in terms of analytical inference, which makes it the benchmark model for an advanced prior setup.

Next let's pick the values of the parameters in prior distributions ($\underline{\beta}$, $\underline{\Omega}$, $\underline{S}$), which reflects reasonable prior beliefs. More precisely, we want to the achieve following goals:

$$E(B_{ij}^{(m)}) = \begin{cases} \delta_i, & j = i, m = 1 \\ 0, & otherwise \end{cases} \text{, and } cov(B_{ij}^{(m)}, B_{ks}^{(l)}) = \begin{cases} \frac{\tau^2 \underline{S}_{ik}}{m^\alpha \underline{S}_{jj}} & , \ j = s, \ m = l \\ 0, & otherwise \end{cases} . \quad (3.6)$$

When considering mixed stationarity, let δ_i be equal to 1 if the ith variable is non-stationary and equal to 0 if the variable is stationary. This setting reflects the prior belief that the non-stationary variable follows a random walk and the stationary variable has no persistence. In line with this, $\underline{\beta} = vec([0, diag(\delta_1, \delta_2, ..., \delta_n), 0, ..., 0]')$. On the other hand, many papers assume that all variables are non-stationary and δ_i is always equal to 1 in this case.

The second moment of coefficients are the tricky part of the prior setup. First, the factor α controls the rate at which the variance decreases with the lag order. Given that the means of the coefficients are equal to 0 except for before the first-order lag of dependent variables, it implies that the coefficients before the higher order of lags are more likely to be equal to zero. The intuition is straightforward: when explaining the current value of the dependent variable, what happens earlier is less relevant than what happens more recently. Second, the ration of $\frac{\underline{S}_{ik}}{\underline{S}_{jj}}$ adjusts the covariance of coefficients according to the scales and variability of explanatory variables. $\underline{S}_{jj}$ is the proxy of the variability or scale of the jth explanatory variable. It is measured by the sample variance of the residuals of the AR(p) model for the jth variable. The intuition is that the variances of coefficients are lower for those variables with higher variability, which can be traced back to the inference of OLS regressions. In an OLS model, the variance of coefficients is equal to $\sigma^2(X'X)^{-1}$ and $X'X$ is the proxy of the variability. When an explanatory variable represents a higher variability, its corresponding coefficient estimates have smaller variance. With respect to $\underline{S}_{ik}$, many papers let it be equal to 0 if $i \neq k$ while some papers measure it by the sample covariance between the residuals of ith and kth variables. In some versions of prior setup, people also assume that the variance of a coefficient before the dependent variable's own lags are bigger than what it is before other explanatory variables. I do not use that specification due to the concern of simplification. With that version of priors, the posterior distribution and prior distribution cannot be written in a concise form (the Kronecker Product of matrices), which would complicate the inference.

The most important parameter in the second moment specification is τ, which dictates the general tightness of prior beliefs. When τ is equal to a small value, it implies that we are more confident with our prior beliefs and more information would therefore come from the prior beliefs. In the extreme case, the posteriors would coincide with the prior beliefs when $\tau = 0$ and no information from the data is used. This would bring about a bad forecasting performance because as long as the model in our prior beliefs is not exactly the true model, the estimates would be biased and the forecast would be imprecise. On the other hand, a big value of τ is not a good choice either. When τ is large, the variance of all of the coefficients in the priors would be high, which implies that we are less confident with the prior beliefs or else the prior beliefs are not very informative. However, the motivation behind introducing the Bayesian framework is to provide some information from prior beliefs; thus when the prior beliefs fail to provide enough information, the BVAR model would be closer to the simple VAR model and we would encounter the over-fitting problem once again. Some evidence can be found from the function of $\bar{B} = [\underline{\Omega}^{-1} + X'X]^{-1}[X'Y + \underline{\Omega}^{-1}\underline{B}]$. $\underline{\Omega}$ is the parameter in the prior beliefs that dictates the level of the variance of the coefficients in the priors. When $\underline{\Omega}$ goes to infinity, which implies a less informative prior, the posterior mean $\bar{B} \to [X'X]^{-1}[X'Y]$, which is exactly the estimate of the simple OLS.

Since neither a small number or a big number is a good choice for τ, we must pick its value very carefully. There is a rich literature on the priors specification since Litterman (1986), Kadiyala and Karlsson (1997), and Sims and Zha (1998) and it is recommended that τ takes a value around 0.2 for a typical BVAR model. In this paper, I follow the literature in choosing τ and then detail more sophisticated methods in the following sections.

In order to achieve the goals documented earlier, the matrix $\underline{\Omega}$ need to be specified as below:

$$\underline{\Omega}_{(np+1)\times(np+1)} = \begin{bmatrix} V_c & 0_{1\times np} \\ 0_{np\times 1} & diag(\frac{\tau^2}{1^\alpha}, \frac{\tau^2}{2^\alpha}, \ldots, \frac{\tau^2}{p^\alpha}) \otimes diag(\underline{S}_{11}, \underline{S}_{22}, \ldots, \underline{S}_{nn}) \end{bmatrix}, \quad (3.7)$$

where V_c is the variance of the intercept. Following the literature, I use the uninformative prior for the intercept in the benchmark model and let $V_c = 1 \times 10^{10}$.

3.3.2 Dummy Observations

The prior setup in previous sections has been discovered to contain a temporal heterogeneity problem, which hurts the model's forecast performance. To be more specific, let $E_p(y_t|y_1, y_2, \ldots, y_p, \hat{\theta})$ refer to the deterministic component of the forecast, and it turns out to explain too much variation of the variables. Unfortunately, this component also varies substantially at different periods. In order to restrain the deterministic component's impact and improve the forecast performance, Doan et al. (1984), Sims (1992), and Sims and Zha (1998) recommend the inclusion of the dummy observations stacked at the beginning of the raw data.

The first type of dummy observations is called the sum-of-coefficients or unit-root dummy observation. It is constructed as follows:

$$Y_{unit-root} = diag(\delta_1 \bar{y}_1, \delta_2 \bar{y}_2, \ldots, \delta_n \bar{y}_n)/\mu \text{ , and } X_{unit-root} = [0_{n\times 1}, (1_{1\times p}) \otimes Y_{unit-root}], \tag{3.8}$$

where $\bar{y}_i$ is the mean of the first p observations for the ith variable and δ_i is the indicator for stationarity as defined earlier. According to these dummy observations, the dependent variables follow unit-root models. The corresponding prior is that the coefficients are centered at 1 before the dependent variable's first-order lags and 0 for other coefficients. The variance of the coefficients are determined by the value of μ. When $\mu \to 0$, we are more confident with the unit-root models; when $\mu \to \infty$, the prior beliefs are less informative and acts as if we does not introduce any unit-root prior beliefs.

The second type of dummy observation corresponds to cointegration priors. More precisely, the dummy observations has the following structure:

$$Y_{cointegration} = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n]/\eta \text{ , and } X_{cointegration} = [1/\eta, (1_{1\times p}) \otimes Y_{cointegration}]. \tag{3.9}$$

As $\eta \to 0$, there are strong beliefs on the cointegration among the variables; when $\eta \to \infty$, this prior is less informative. As implied by the name, this prior setup is motivated by the cointegration relationship of the variables. We will talk about the value of the parameters in the following sections.

The dummy observations setup can be derived through a Theil mixed estimation. Following this method, the prior specification in Section 3.3.1 can also be represented through dummy observations. More precisely, we can add the following dummy observations at the beginning of the real data:

$$Y_{coef} = \begin{bmatrix} diag(\delta_1 \underline{S}_{11}^{0.5}, \delta_2 \underline{S}_{22}^{0.5}, \ldots, \delta_n \underline{S}_{nn}^{0.5})/\tau \\ 0_{[n(p-1)+1]\times n} \end{bmatrix},$$

$$X_{coef} = \begin{bmatrix} 0_{np\times 1} & diag(\frac{1}{1^{\alpha/2}\tau}, \frac{1}{2^{\alpha/2}\tau}, \ldots, \frac{1}{p^{\alpha/2}\tau}) \otimes diag(\underline{S}_{11}^{0.5}, \underline{S}_{22}^{0.5}, \ldots, \underline{S}_{nn}^{0.5}) \\ \frac{1}{Vc^{0.5}} & 0_{1\times np} \end{bmatrix}, \quad (3.10)$$

$$Y_{cov} = diag(\underline{S}_{11}^{0.5}, \underline{S}_{22}^{0.5}, \ldots, \underline{S}_{nn}^{0.5}) \text{ , and } X_{cov} = 0_{n\times(np+1)}. \quad (3.11)$$

Matrix (3.10) corresponds to the coefficient's prior distribution and Matrix (3.11) corresponds to the error term covariance matrix's prior distributions. If we use the improper prior distribution $p_0(\theta) = |\underline{S}|^{-\frac{n+3}{2}}$ and stack Matrix (3.10) and Matrix (3.11) at the beginning of the raw data matrix, the posterior distribution would coincide with Distribution (3.5).

3.3.3 Large BVAR Model

In the following three sections, I discuss the more sophisticated ways of improving the forecast performance of the BVAR model. As discussed in Section 3.1, the larger database is found to be helpful and I explore three approaches in utilizing the larger datasets consecutively. The first approach is the Large BVAR model, which has proven to be a nice tool for the U.S. economy. I detail the methodology in this section and would discuss its revision when applied to China's data in Section 3.5.2. According to Banbura et al. (2010), the value of τ should be chosen in relation to the number of variables. More precisely, Banbura et al. (2010) pick a training sample before the evaluation sample and focus on 3 key variables – Employment, CPI, and Federal Funds Rates. First, a 3-variable VAR model is estimated on the training sample for the key variables to calculate the in-sample MSFE (mean square forecast error). Second, a BVAR model is set up on the training sample and τ is chosen to equate the in-sample MSFE of BVAR model and that of 3-variable VAR model. To be specific, the MSFE and τ are calculated as follows:

$$msfe_i^{(model)} = \frac{1}{T_0 - p} \sum_{t=p}^{T_0-1} (y_{i,t+1|t}^{(model)} - y_{i,t+1})^2 \text{ and } model = c, \tau, \tag{3.12}$$

$$fit_i^{model} = 1 - \frac{msfe_i^{(model)}}{msfe_i^{(rw)}} \text{ and, } FIT^{model} = \frac{1}{I} \sum_{i=1}^{I} fit_i^{model}, \tag{3.13}$$

$$\tau = arg\min_{\tau} |Fit^c - Fit^\tau|, \tag{3.14}$$

where $msfe_i^{(\tau)}$, $msfe_i^{(c)}$, and $msfe_i^{(rw)}$ are the MSFE for the ith variable within the training sample for the BVAR model with a certain value of τ, the target model (3-variable VAR model), and the Random Walk model, respectively. T_0 refers to the sample size of the training sample and I is the set of target variables. In Banbura et al. (2010)'s paper, I includes Employment, CPI, and Federal Funds Rates. As implied by Equation (3.14), we pick the value of τ with the smallest difference of the in-sample fitness between the target model and BVAR model. In order to account for the scales and variability for different variables, the MSFE is standardized by the MSFE from the Random Walk model for each variable. After we pick the value of τ, we can apply its value to estimate the parameters of the BVAR model and then calculate the forecast.

Given the detailed algorithm, let us now focus on its intuition. First of all, it is reasonable to pick a higher overall tightness of the Bayesian priors when the number of the variables in the model increases because more variables result in a more severe over-fitting problem. A tighter prior brings about more aggressive shrinkage, which can counteract the impact of a larger database. Given this principle, we still need a relatively objective algorithm to pick τ and Banbura et al. look into the source of the problem: since the Bayesian priors is used to fight the in-sample over-fitting, a proper τ should be the one that keeps the in-sample fitness constant when the number of variables changes. In the next step, the authors pick the

3-variable VAR model as the target model because this small model is thought to represent modest over-fitting.

One may argue that when there are so many parameters involved in the BVAR model, why should only the value of τ be picked instead of adjusting other parameters at the same time? My answer is as follows: After careful investigation, τ is the key parameter or the most sensitive parameter in the system and its value has a huge impact on the estimation and forecast. Besides, it is hard to find a rule that picks all of the parameter values at the same time and this rule of picking only τ's value is one step ahead to set up the priors in a more objective and data-dependent way. I would later discuss the methods of picking the values of a group of hyperparameters in the hierarchical model, which is not as successful as the more heuristic Large BVAR model. In the empirical analysis section, I will return to this point and discuss this issue in a more quantitative way.

3.3.4 FAVAR Model

Another way of utilizing more data is through a FAVAR, or Factor-augmented VAR model, proposed by Bernanke et al. (2005). Instead of adjusting the Bayesian priors to accommodate the data, this method transforms the data to fit the model. More precisely, this methodology extracts factors from the data and integrates them with the standard VAR model.

Let f_t be the m dimension vector and each element of it corresponds to one unobservable factor; x_t is the k dimension vector for the big dataset. The FAVAR model assumes that the y_t and f_t follow the following specification:

$$\begin{bmatrix} f_t \\ y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} f_{t-1} \\ y_{t-1} \end{bmatrix} + v_t \text{ , and } x_t = \Lambda^f f_t + e_t. \tag{3.15}$$

There are two available estimation approaches: a Bayesian likelihood and two-step principal components. Although the Bayesian likelihood approach is more consistent theoretically, I use the two-step principal components approach because it is easier to implement and produces better empirical results according to Bernanke et al. (2005). In the first step, I extract the principle component factor f_t from x_t. In the second step, f_t and the key macroeconomic variables y_t are pooled together in the standard VAR model. I don't distinguish between the fast-moving and slow-moving variables as do Bernanke et al. because I focus mainly on forecasting instead of a structural analysis.

After careful investigation, it is found that the FAVAR model's over-fitting problem is still significant because its MSFE is even higher than the Random Walk model and AR model. In order to improve the model's performance, I incorporate the Bayesian priors to the FAVAR model following Banbura et al. (2010). The values of the parameters are picked in exactly the same way as Large BVAR model, and the factors are treated as stationary variables.

3.3.5 Hierarchical Model

This method is the most objective and theoretically consistent in that it can pick the optimal value of a group of hyperparameters automatically. Actually, the Large BVAR model can also be regarded as a way of parameter optimization by picking the value of τ in relation to the number of variables. Although empirically powerful, the Large BVAR method is still to some extent arbitrary and is not strongly theoretically-grounded. Motivated by this issue, Giannone et al. (2015) propose a hierarchical model to add hyperpriors to priors. To be specific, $p(\theta)$ is replaced with $p(\theta|\tau)$ where τ is a hyperparameter and θ is the prior's parameter; thus, the prior's density is $p(\theta) = \int p(\theta|\tau)p(\gamma)d\gamma$. The corresponding posterior density $p(\gamma|y) \propto p(y|\gamma)p(\gamma)$, where $p(\gamma)$ is the prior density of the hyperparameters γ and $p(y|\gamma)$ is the marginal likelihood (ML) and follows the expression below:

$$p(y|\gamma) = \int p(y|\theta,\gamma)p(\theta|\gamma)d\theta. \tag{3.16}$$

Following the prior specification in Section (3.3.1), it can be proved that

$$p(y|\gamma) \propto \underbrace{|(V_\varepsilon^{posterior})^{-1}V_\varepsilon^{prior}|^{\frac{T-p+d}{2}}}_{\text{Fit}} \cdot \underbrace{\prod_{t=p+1}^{T} |V_{t|t-1}|^{-\frac{1}{2}}}_{\text{Penalty for model compexity}}, \tag{3.17}$$

where $V_\varepsilon^{posterior}$ and V_ε^{prior} are the means of the error term covariance matrix for the posterior and prior distributions, $V_{t|t-1} \equiv E_\Sigma[var(y_t|y_{t-1,}y_{t-2},\ldots,\Sigma)]$, which is the variance of a one-step-ahead out-of-sample forecast of y average across Σ. Equation (3.17) represents the advantage of ML in capturing the trade-off between the in-sample fitness and out-of-sample forecast. The first term of RHS represents the in-sample fitness effect. With a smaller $V_\varepsilon^{posterior}$, ML is higher and the fitness is better. The second term of RHS captures the penalty for model complexity. When the model is too complex and suffers from the over-fitting problem, the variance of the out-of-sample forecast is larger, and thus lowers the value of ML. When the second term dominates the function, ML could be a good measure of the out-of-sample forecast performance.

Following Giannone et al.'s work, I set up the hyperpriors in this way: τ, θ, η follow a Gamma distribution and the diagonal vector of $\underline{S}$ follows an inverse-Gamma distribution. The details of the values of the parameters are discussed in Section 3.5. Giannone et al. explain the reason why this hierarchical model can deliver good forecast results: the model can automatically pick the optimal value of parameters in relation to the data. One experiment is to shut down the hyperpriors except for τ. The result is that τ gets smaller when the number of variables increases. The mechanism seen here is similar to the mechanism in Banbura et al. (2010)'s model.

In general, the hierarchical model is superior to Banbura et al.'s approach in the following aspects. First, it is more strongly theoretically grounded and consistent. Second, it is able to pick the optimal value of all parameters together rather than only one parameter of τ. Third, the optimization procedure is more well-rounded and does not only adjust with the size of the model. However, the hierarchical model also has its disadvantages. First, the hyperprior distribution is still arbitrary and empirical analysis shows that the value of the hyperpriors can dramatically affect the results. Second, the performance of the model is not very robust. While working well for the study of the U.S. economy, the performance is not strong in the context of applying the model to the China's economy. I will discuss more details of the hierarchical model in Section 3.5.

3.4 Data

The primary database of the analysis covered in this paper comes from the Center for Quantitative Economic Research (CQER) of the Federal Reserve Bank of Atlanta. The data is compiled by Higgins and Zha (2015) and includes 49 quarterly macroeconomic variables. One of the features of this database is that the quarterly expenditure-side GDP breakdown is interpolated through the econometric methods. Because of this, I can add consumption and investment, which are otherwise unavailable to our BVAR model. I take 34 series from CQER and all of the key macro variables in the benchmark model comes from this database. The variables from CQER are listed in Table 3.1. The notation of $\times$ in the "Log" column means that the variable is log transformed. An $\times$ in the column of "Non-stat" indicates that the variable is non-stationary. In factor analysis, I use the first difference with respect to the non-stationary variable. It is important to note that it is not necessary to take the first difference for the non-stationary variable in the BVAR model. An $\times$ in the column of "SA" means that the variable is seasonally adjusted through an X12 approach.

In order to build the model a utilizing large dataset, I collect additional 90 macro variables and sectoral variables from the CEIC database, which are listed in Table 3.2. Although CEIC is the database most widely used for China's economic research, its data structure is cumbersome and I clean the data in the following aspects. First, I fill in the missing values of the data by a linear interpolation. Sometimes the values of some variables in January are missing and the values in February store the sum of the two months. The new values are calculated for the two months by assuming the sum of the two values is the original value stored in February and the month-over-month growth rate for the two months are the same. Second, some variables are available only through growth rate and I transform the growth rate into the level value by picking some base year level. Third, for variables only available in a year-to-date value, I calculate the value of each period by taking the first difference within each year. Fourth, no seasonally adjusted series is used from CEIC and I conduct the seasonal adjustment on my own through the X12 approach using the program provided by the U.S. Census Bureau. Fifth, the monthly frequency variables are converted into quarterly frequency variables by taking the average. Table 3.2 contains the description of the data from CEIC and the meaning of the items are exactly the same as in Table 3.1.

The cleaned data are pooled together and there are 124 series in total that reflect a wide perspective of the economy, such as interest rate, equity market, output, export, government activities, prices, etc. Following the practice of Higgins et al. (2016), the benchmark BVAR model includes variables of Real GDP, CPI, 1-day Repo Rate, Net Export over GDP, Real Consumption, Real Investment, and 1-year Deposit Rate. The sample ranges from 1996Q1 to 2014Q4 in which all the variables have no missing values.

Following the BCEI forecast methodology and Higgins et al. (2016), the log level variables such as GDP and CPI are converted into annualized growth rates. To be specific, I encode $t = \{yr, qtr\}$, where yr represents the calendar year and qtr represents the calendar quarter. The annual growth rate is calculated as

$$\tilde{y}^i_{yr,qtr} = 100\left[\frac{\frac{\sum_{qtr=1}^{4} exp(y^i_{yr,qtr})}{4}}{\frac{\sum_{qtr=1}^{4} exp(y^i_{yr-1,qtr})}{4}} - 1\right]. \tag{3.18}$$

The first forecast date is labeled as $T+1 = \{yr*, qtr*\}$. If $yr < yr*$ or if $yr = yr*$ but $qtr* < qtr$, then $y^i_{yr,qtr}$ is the actual value; otherwise, $y^i_{yr,qtr}$ is the value predicted by the model. The intuition is that the predicted annual growth rate is updated for each quarter before the end of the year. Generally speaking, the closer it is to the end of the year, the better the forecast becomes as more actual series are used. Note that this approach only applies to the growth variables and I simply take the average within a year for variables like the interest rate, which is already expressed as a percent.

RMSE is the key evaluation measure of the forecast used in this paper and it is calculated following the literature as

$$rmse^{(model)}_{i,h} = \sqrt{\frac{1}{T_1 - T_0 + 1}\sum_{t=T_0}^{T_1}(\tilde{y}^{(model)}_{i,t+h|t} - \tilde{y}_{i,t+h})^2}, \tag{3.19}$$

where $\tilde{y}^{(model)}_{i,t+h|t}$ refers to the forecast h periods ahead into the future calculated by the model for variable i at period t; $\tilde{y}_{i,t+h|t}$ is the true value corresponding to the forecast; T_0 and T_1 are the starting and ending point of the evaluation sample, respectively.

The key macro variables are GDP, CPI, and the 1-day Repo Rate, which are plotted in Figure 3.1 to Figure 3.3. The figures clearly show that China has experienced a remarkable economic boom from the beginning of the 21st century until the Great Financial Crisis. Both GDP and Inflation plummeted after the crisis and the interest rate drops substantially as a result of the implementation of an expansive monetary policy. Thanks to the unprecedented government intervention, China's economy sees a robust recovery around 2010, but the economy seems to run into a long-run downturn cycle thereafter. Another striking pattern is that the interest rate was once extremely high at the beginning of the sample but has decreased and since 2000 remained at a relatively low level. In the meantime, China is undergoing the economic reform, which may affect the dynamics of the variables significantly. For example, the interest rate has not been used as a policy tool until recent years, thus there may be changes in the mechanism of the monetary policy. Due to the facts outlined above, it is extremely difficult to model the economy of China and one should not be surprised that many models used in other economies do not work well for this economy.

3.5 Empirical Evidence

3.5.1 Benchmark Model

In order to evaluate the model's performance, I not only evaluate the BVAR models but also Random Walk model, AR model, and simple VAR model. The Random Walk model is of particular interest because it is usually regarded as the gold standard of forecasts. Because of this, I also report the RMSE of the different models divided by the RMSE of the Random Walk model, which is called Relative RMSE in the following analysis. The specification of the Random Walk model follows the practice of Atkeson and Ohanian (2001) and Higgins et al. (2016). More precisely, the forecast for the ith variable multi-period in the future is calculated recursively as

$$y^i_{t+1} = y^i_t + (y^i_t - y^i_{t-4})/4. \tag{3.20}$$

The empirical analysis starts with two benchmark models. The first one is the benchmark model discussed in Section 3.3.1, which is basically the hybrid of the Minnesota Priors and Sims&Zha priors, which I call Minnesota Revised model thereafter. In this benchmark model, $\tau = 0.2$, $\alpha = 2$, $\eta = 1$, and $\mu = 1$ following Giannone et al. (2015). The second benchmark model is the Sims&Zha Model for the U.S. economy in which $\mu_1 = 1$, $\mu_2 = 1$, $\mu_3 = 1$, $\mu_4 = 1$, $\mu_5 = 1$, and $\mu_6 = 1$ following the practice of Sims and Zha (1998).

Table 3.3 and Table 3.4 report the RMSE of GDP, CPI, and the 1-day Repo Rate for different models. The first column refers to the variables and the second column refers to the forecast horizon covering the next 4 years. 1-day Repo Rate is denoted by "R" in the table. The rest of the columns report the RMSE for different models: RW stands for the Random Walk model, AR is the Auto-regression model, VAR refers to the Vector Auto-regression model, and Minnesota Revised and Sims&Zha are the two benchmark models. All autoregressive models in this paper include 5 lags. One obvious pattern is that the RMSE becomes higher for the longer forecast horizon. Table 3.5 reports the Relative RMSE of GDP, CPI, and 1-day Repo Rate for different models and the labels of the columns are the same as in Table 3.3. If the Relative RMSE is greater than one, it simply means that the rigorous econometric model cannot even outperform the Random Walk model. It can be seen that the Minnesota Revised model does a poor job particularly for the GDP forecast and that the Sims&Zha model is the best model I have at the moment.

3.5.2 Evaluation of Large BVAR Model

First, let's study the Large BVAR model strictly following Banbura et al. (2010). More precisely, let $\alpha = 2$ and $\mu = 10\tau$, the cointegration priors are turned off , and the Training Sample is the first 10 year's data. The large model includes all 124 variables and 5 lags. The RMSE and Relative RMSE are reported in Table 3.6 and Table 3.7, respectively. The performance of the large model is labeled by BGR, which is compared with the two small benchmark models. The optimal value $\tau = 0.07$ is calculated by the algorithm. Obviously, the BGR model outperforms the Random Walk model and Minnesota Revised model for all horizons and it also beats the Sims&Zha model with the exception of the short-horizon forecast of interest rate. The magnitude of the improvement is also impressive, particularly for the long-run forecast of GDP: the RMSE of GDP for the BGR model is almost half of the competing benchmark models.

The next issue is the feasibility of further improving the Large Model's performance. As discussed in Section 3.3.3, this methodology picks the value of parameters in a heuristic way and lacks a theoretical background. First, the value of α , μ, and η are arbitrarily chosen. Second, although the methodology follows the intuition that τ should be smaller as the number of variables increases, the rule of picking τ is not necessarily optimal. With these limitations in mind, I explore the performance of the model by changing the values of the parameters.

Table 3.8 reports the Relative RMSE when arbitrarily assigning different values to τ. Note that the values of other parameters are the same as the original BGR model. We can see that the Relative RMSE of GDP and CPI become smaller as τ decreases from 0.07 to 0.03, whereas the forecast of the interest rate is worse. Table 3.9 presents the Relative RMSE with a different α. The improvement is significant when α increases from 2 to 4. From these practices, it is reasonable to argue that the more aggressive shrinkage for the Bayesian priors may help improve the forecast. As of now, another problem I need to address is that τ is automatically picked by the BGR algorithm. If I want the target τ to be smaller, the targeting rule must be changed. Since τ is determined by the in-sample fitness in the BGR algorithm, I report the in-sample fitness for different models in Table 3.10. The upper panel of the table reports the in-sample fitness of the 3 key macro variables for Large BVAR models with different degree of shrinkage. It can be seen that the model with higher degrees of shrinkage, which are Test 3 and Test 6, tend to represent lower in-sample fitness. The model whose parameter values are the same as the original BGR algorithm delivers the highest in-sample fitness. The lower panel of Table 3.10 reports the in-sample fitness of potential target models. VAR3 refers to the VAR model with 3 key macro variables and

VAR2 refers to the VAR model with 2 macro variables (GDP and CPI). VAR3 delivers an in-sample fitness closest to the original BGR model labeled by Test 1 because the value of τ here is exactly chosen by targeting the fitness of VAR3. Among different potential target models, the AR model delivers the smallest in-sample fitness. Provided that lower in-sample fitness comes with more aggressive shrinkage, I build the Revised BGR algorithm by targeting the in-sample fitness of GDP and CPI calculated by the AR model. Besides, the value of α is raised to 4 in the new algorithm.

The performance of the revised Large BVAR models are reported in Table 3.11 and Table 3.12. The BGR Revised column stands for the large model using the revised BGR algorithm ; $\tau = 0.03$, which is picked by the algorithm. The improvement is greater for the short-run forecast of GDP and the long-run forecast of CPI and the interest rate.

Next let's discuss the impact of the Sims&Zha prior's parameters. It is important to note that for all of the models above do not have cointegration priors; however, macro variables tend to present cointegration relationships. Inspired by this fact, I add a cointegration prior with different tightness to the revised BGR large model and report the performance in Table 3.13 and Table 3.14. It is clear that the forecast of GDP is better with a moderate tightness of the cointegration prior, which is $\eta = 1$. Although some of the forecasts of CPI and the interest rate are marginally worse, the gains of the GDP forecast dominates considering the magnitude of the changes. As such, the moderate level tightness of the cointegration prior generally helps improve the forecast performance of the model. But if I go to the extreme by setting a very tight prior, such as $\eta = 0.3$, the forecasts would become much worse as shown by the very right column in Table 3.14. A potential explanation is that the over-tight Bayesian prior tends to put too much weight on the prior beliefs and mutes the information from the real data.

The other component of the Sims&Zha prior is the unit root priors. Different values of μ are tested and the performance is reported in Table 3.15 and Table 3.16. Note that the models are large BVAR models built on the revised BGR algorithm with $\eta = 1$. In the 3rd column, $\mu = 10\tau$ serves as the benchmark. While some values of μ bring about better forecasts for some horizons compared to the benchmark, either the improvement is marginal or the forecasts of the other horizons become much worse. As a result, it is reasonable to follow the original BGR algorithm by setting $\mu = 10\tau$.

The last factor that might affect the forecast performance is stationarity. Most of the BVAR papers assume that all variables are non-stationary whereas BGR uses mixed stationarity. In Table 3.17 and Table 3.18, I report the forecast performance for Revised BGR models with $\mu = 10\tau$ and $\eta = 1$. The "Mixed" column is for the model with mixed stationarity and the "All-stat" column is for the model assuming non-stationarity for all variables. It can be seen that the impact of the stationarity is mixed. The forecast of GDP becomes worse whereas the forecast of the interest rate improves. Considering the magnitude of change in forecasting, I assume non-stationarity for all variables in the priors since it delivers a more balanced forecast performance across variables.

3.5.3 Evaluation of Factor Model

In Section 3.3.4, I discuss that the Bayesian Factor-augmented VAR (BFAVAR) model can be built by incorporating the Bayesian priors to the FAVAR model through a BGR algorithm. Since I also develop a revised BGR algorithm that has more aggressive shrinkage and contains a cointegration prior, the first question to ask is how the two BGR algorithms of adding Bayesian priors affect the performance. The results are presented in Table 3.19 and Table 3.20. Column BFAVAR3 and Column BFAVAR5 refer to the BFAVAR model with 3 factors and 5 factors, respectively. The second row of the tables record the BGR algorithm. The designation "Original" means that the model applies the original BGR algorithm, whereas "Revised" means that the model applies the Revised BGR algorithm. For clarification, the Revised BGR algorithm targets the AR model to pick the value of τ;

$\alpha = 4$, $\mu = 10\tau$, and $\eta = 1$. A simple comparison across different columns shows that the BFAVAR model using a revised algorithm outperforms the original algorithm. The biggest improvement comes from the forecast of GDP since the original BGR algorithm is particularly poor.

Another important specification of the BFAVAR model is the number of factors. While some of the literature discusses the optimal number of factors, most papers report the performance of a group of models with different factors. In this analysis, I report the performance when there are one, three, and five factors in the models using a revised BGR algorithm for the Bayesian priors setup. The results are reported in Table 3.21 and Table 3.22. The first finding is that the 5-factor model is generally worse compared to the other two models; however, it is is harder to rank the performance of the 1-factor model and 3-factor model. The 3-factor model prevails when forecasting GDP whereas the 1-factor model does a better job with the long-run forecast of CPI and interest rate.

It is also interesting to look into the impact of stationarity. Since the 1-factor and 3-factor models are tied in performance, I report the stationarity's impact for both of them as shown in Table 3.23 and Table 3.24. Generally the setup of non-stationarity for all variables helps improve the performance, particularly for the forecast of GDP and the long-run forecast of CPI. Among the models, it is found that the 3-factor model with non-stationarity for all variables delivers better results.

3.5.4 Evaluation of Hierarchical Model

In this section, I investigate the impact of prior optimization through a hierarchical model following Giannone et al. (2015). As discussed in Section 3.3.5, this methodology is superior to a Large BVAR model because it is more theoretically consistent and well-rounded. The value of the parameters for the Bayesian priors is not arbitrarily chosen, which saves effort in an empirical sense. First, I build the model following Giannone et al. (2015). The modes of the parameters are $\eta_{mode} = 1$, $\mu_{mode} = 1$, and $\tau_{mode} = 0.2$; for the standard deviation , $\eta_{sd} = 1$, $\mu_{sd} = 1$, and $\tau_{sd} = 0.4$. α has no hyperprior distribution and $\alpha = 2$. The elements of the principle diagonal of matrix $\underline{S}$ follow the inverse-Gamma distribution and its value is equal to the value of corresponding elements in $\underline{S}$ as calculated in Section 3.3.1. The scale and shape of the distribution is equal to $(0.02)^2$. The hierarchical model's performance is reported in column GLP Original in Table 3.25 and Table 3.26. The second row refers to the stationarity and its meaning is the same as in previous sections. It can be seen that non-stationarity for all variables gives us better performance.

From the practice in Section 3.5.2, I learned that the degree of shrinkage plays a substantial role for model performance; however, in the original GLP's hierarchical model, α is set to a constant. One natural variant is allowing α to have a hyper distribution as well and exploring the performance. For the revised GLP model, let α follow a Gamma distribution, such as $\alpha_{mode} = 1$ and $\alpha_{sd} = 1$. The performance is reported in the GLP Revised columns. As expected, the revised GLP model achieves substantial improvement for the GDP and CPI forecasts. Considering the importance of the variable and the magnitude of change, the revised GLP model with all non-stationary variables generally stands out among the hierarchical model group.

3.5.5 Champions of Models

This section compares the performance of the best models (Champions) from the different model groups. In addition to the 3 types of models using a large database, I also report the performance of the champion of the small models, which include 8 key macro variables. It is found that the Sims&Zha model is better than the revised Minnesota model and I pick the good parameter combination through a trial-and-error process. More precisely, $\mu_1 = 0.3$, $\mu_2 = 1$, $\mu_3 = 1$, $\mu_4 = 2$, $\mu_5 = 10$, and $\mu_6 = 1$. Compared to the Sims&Zha prior parameters in the benchmark model, the new parameter combination has more aggressive shrinkage for the lag decaying factor, overall tightness, and the unit-root prior.

With respect to the models built from a large database, here is their detailed specification: The Large BVAR model is built through a revised BGR algorithm meaning that $\alpha = 4$, $\mu = 10\tau$, and $\eta = 1$. I target the AR model's in-sample fitness for GDP and CPI. In addition, all variables are assumed to be non-stationarity in the priors. The BFAVAR model incorporates Bayesian priors through a revised BGR algorithm, which has been discussed previously, to a 3-factor FAVAR model and all variables are non-stationary in the priors. Hierarchical model is built following a revised GLP algorithm in which α also has a hyper prior distribution and all variables are non-stationary in the priors.

The BCEI forecast is also included in the analysis and it reflects the forecast of the corporate world. The BCEI forecast is updated monthly for GDP and CPI forecast 2 year into the future. Since my models are built on quarterly data, I convert the BCEI forecast series to a quarterly frequency by taking a simple average. The missing value for forecast is recorded as "NA" in the performance tables.

The performance of models is reported in Table 3.27 and Table 3.28. Let us start the discussion with the small model. It can be seen that the small model is already very successful is able to outperform the BCEI forecast for the one-year horizon. Actually, its one-year forecast of CPI is the best among all models. It is very impressive that a small model of only 8 variables can deliver such good results thanks to the Sims&Zha priors. On the other hand, the large model group shows that a large database is helpful in improving the performance. The most successful large model is the Large BVAR model built from a BGR algorithm and it particularly reduces the RMSE for the short-run forecast of GDP. The BFAVAR model also gives a decent forecast performance with its strength in forecasting CPI and it is the only econometric model that beats BCEI's CPI forecast for the two-year horizon. Unfortunately, its forecast of GDP and interest rate is worse, which makes its overall performance not well-balanced. The hierarchical model built through the GLP algorithm is the worst large model primarily because the algorithm requires a numerical method to search for the mode of the parameters and it does not work well on large-dimension data.

As such, I argue that the large database helps improve the forecasting performance and that the revised BGR algorithm turns out to be the most successful methodology. The gains of the large database mainly comes from the improvement for the short-run GDP forecast, long-run CPI forecast, and long-run interest rate forecast compared to the best small model. To visualize the performance, the forecasts of the models in this section are plotted for Figure 3.4 to Figure 3.9.

3.6 Conclusion

This paper discusses the method of building an econometric forecasting model in the Bayesian realm. While many people doubt the feasibility of a true econometric forecasting model for China, I find some promising models that perform very well. The standard small 8-variable BVAR model specified for the U.S. economy cannot deliver good performance; however, a simple revision of the parameters improves the performance significantly. The most important finding in this paper is that a rich dataset of financial data, sectoral data, and economic data turns out to contain important information for forecasting GDP, CPI, and interest rate. A Large BVAR model with 124 variables is the best model available and generally outperforms other models particularly in terms of forecasting GDP. The BFAVAR model also delivers a decent forecasting performance but is more unbalanced than the Large BVAR model. The hierarchical model, which is a more theoretically-grounded model in terms of picking the parameter's value, is not very effective in improving the model performance for China's economy. Another point worth noting is that the champion of the models in this paper can outperform the BCEI forecast for GDP and CPI in the short-run, which provides more confidence in applying the findings and methodologies in this paper to guiding real-world decisions.

References

Atkeson A., and L. E. Ohanian. 2001. Are Phillips Curves Useful for Forecasting Inflation? Federal Reserve Bank of Minneapolis Quarterly Review, 25 (1), 2-11.

Banbura M., D. Giannone, and L. Reichlin. 2010. Large Bayesian VARs. Journal of Applied Econometrics, 25 (1), 71–92.

Bernanke B., J. Boivin, and P. Eliasz. 2005. Measuring Monetary Policy: A Factor Augmented Autoregressive (FAVAR) Approach. Quarterly Journal of Economics, 120 (1), 387-422.

Chang C., K. Chen, D. Waggoner, and T. Zha. 2015. Trends and Cycles in China's Macroeconomy. NBER Macroeconomics Annual 2015, Volume 30.

Chow G. C. 1986. Chinese Statistics. The American Statistician, 40 (3), 191–196.

Christiano L. J., M. Eichenbaum, and C. L. Evans. 1999. Monetary Policy Shocks: What Have We Learned and To What End? Handbook of Macroeconomics, 1, 65-148.

De Mol C., D. Giannone, and L. Reichlin. 2008. Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components? Journal of Econometrics, 146, 318–328.

Del Negro, M., and F. Schorfheide. 2004. Priors from General Equilibrium Models for VARS. International Economic Review, 45 (2), 643–673.

Doan T., R. Litterman, and C. Sims. 1984. Forecasting and Conditional Projection Using Realistic Prior Distributions. Econometric Reviews, 3 (1), 1-100.

Fang Y., and J. Wu. 2009. China's Inflation: External Shocks or Monetary Easing? Evidence from BVAR Forecasting Models. International Finance Research, 4, 72-78.

Fernald J. G., M. M. Spiegel, and E. T. Swanson. 2014. Monetary Policy Effectiveness in China: Evidence from a FAVAR Model. Journal of International Money and Finance, 49, 83-103.

Giannone D., M., Lenza, and G. E. Primiceri. 2015. Prior Selection for Vector Autoregression. The Review of Economics and Statistics, 97 (2), 436-451.

He Q., P. Leung, and T. T. Chong. 2013. Factor-augmented VAR Analysis of the Monetary Policy in China. China Economic Review, 25, 88-104.

Higgins P., and T. Zha. 2015. China's Macroeconomic Time Series: Methods and Implications. Unpublished Manuscript, Federal Reserve Bank of Atlanta.

Higgins P., T. Zha, and K. Zhong. 2016. Forecasting China's Economic Growth and Inflation. NBER Working Paper 22402, July 2016.

Holz C. A. 2014. The Quality of China's GDP Statistics. China Economic Review, 30, 309–338.

Kadiyala K. R., and S. Karlsson. 1997. Numerical Methods for Estimation and Inference in Bayesian VAR Models. Journal of Applied Econometrics, 12 (2), 99–132.

Koop G., and D. Korobilis. 2010. Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. Foundations and Trends in Econometrics, 3(4), 267– 358.

Kuang M., Z. Zhou. 2015. China Macroeconomic System Dynamic Conduct, Conduct Reliability, and Dynamic Mechanisms of Monetary Policy. Economic Research, 50 (2), 31-46.

Litterman R. 1980. A Bayesian Procedure for Forecasting with Vector Autoregression. MIT, Department of Economics Working Paper.

_____. 1986. Forecasting with Bayesian Vector Autoregression - Five Years of Experience. Journal of Business and Economic Statistics, 4, 25-38.

Sims C. A.. 1980. Macroeconomics and Reality. Econometrica, 48, 1-48.

_____. 1992. Bayesian Inference for Multivariate Time Series with Trend. Mimeo, Princeton University.

_____. 1993. A Nine-variable Probabilistic Macroeconomic Forecasting Model. Business Cycles, Indicators and Forecasting, 179-212.

Sims C. A., and T. Zha. 1998. Bayesian Methods for Dynamic Multivariate Models. International Economic Review, 949-968.

Summers P. M., and S. Zhang. 1998. BVAR Quarterly Forecasting Model. Quantitative Economic Methodology and Research. 9, 29-33.

Wang F. 2011. Regional Economic Forecasting Model Based on the Bayesian Vector Autoregression: Evidence from Qinghai Province. Journal of Quantitative Economics, 28 (2), 95-100.

Tables and Figures

Figure 3.1: GDP, Annual Growth Rate

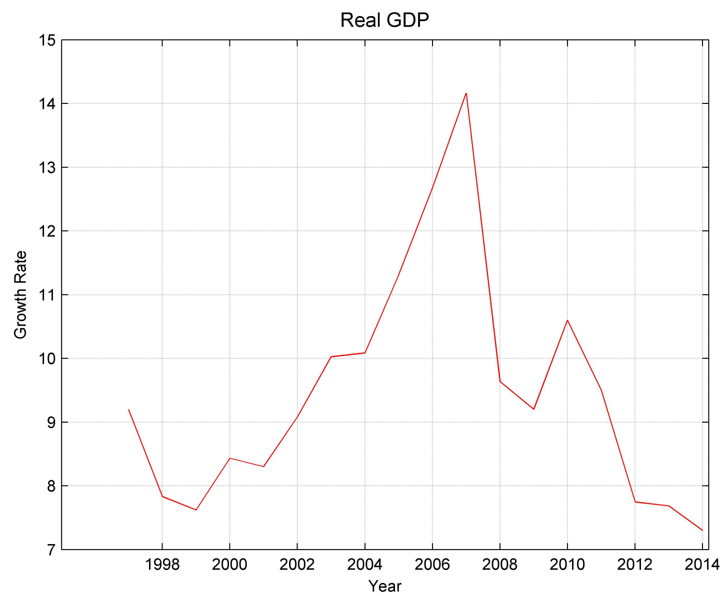

Figure 3.2: CPI, Annual Growth Rate

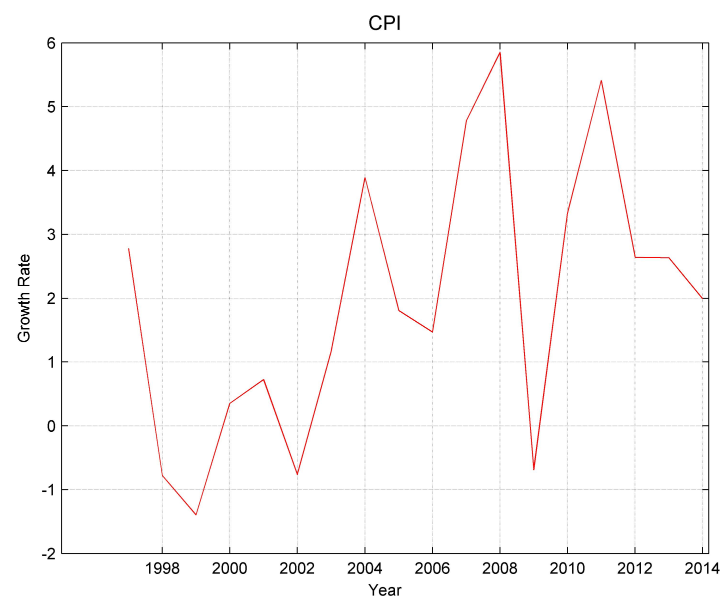

Figure 3.3: 1-day Repo Rate, Annual Average

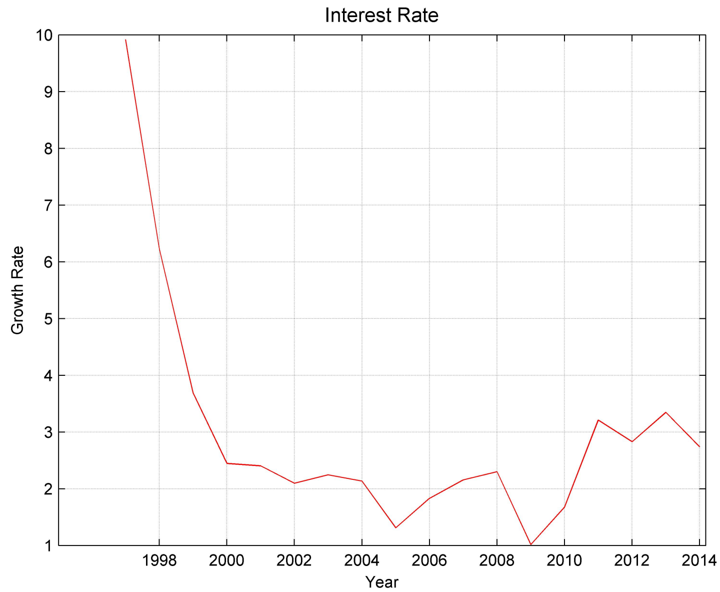

Figure 3.4: GDP Forecast, Panel A

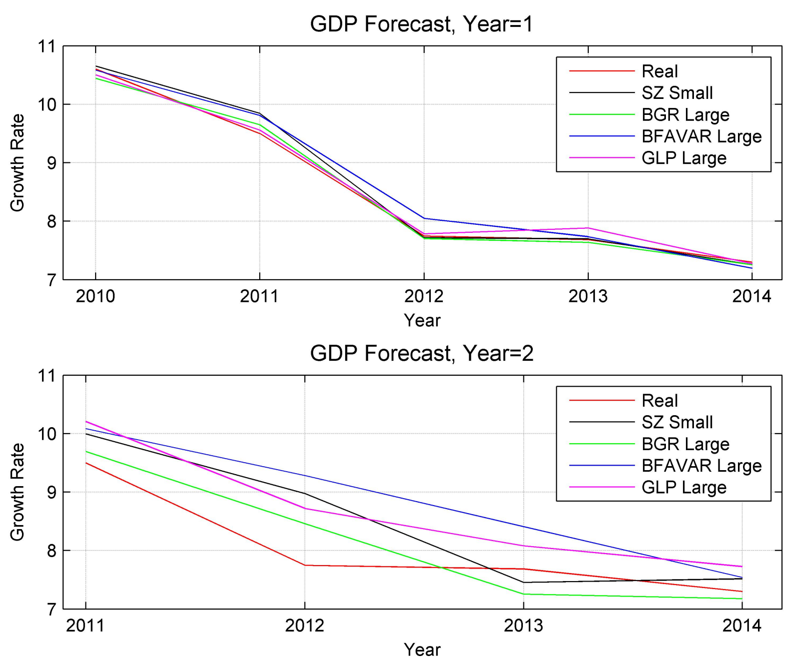

Figure 3.5: GDP Forecast, Panel B

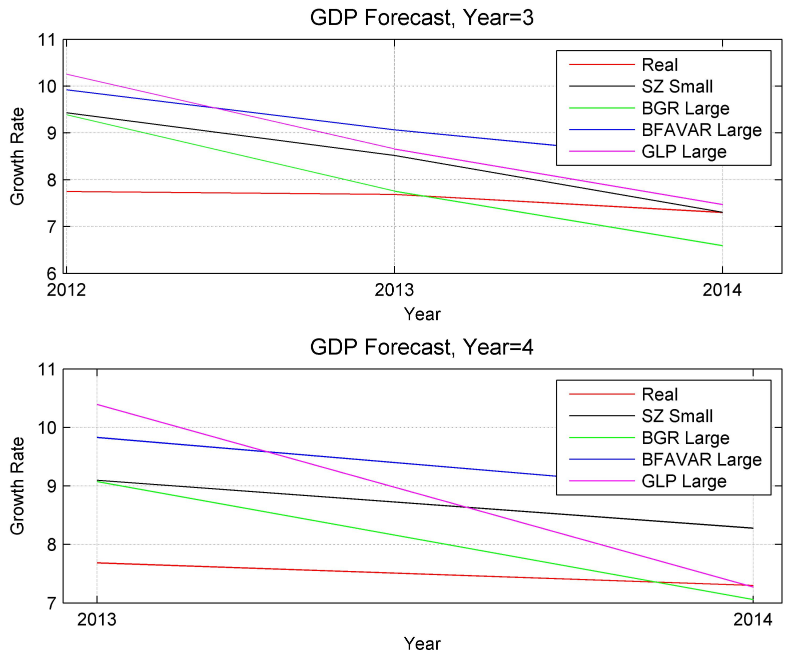

Figure 3.6: CPI Forecast, Panel A

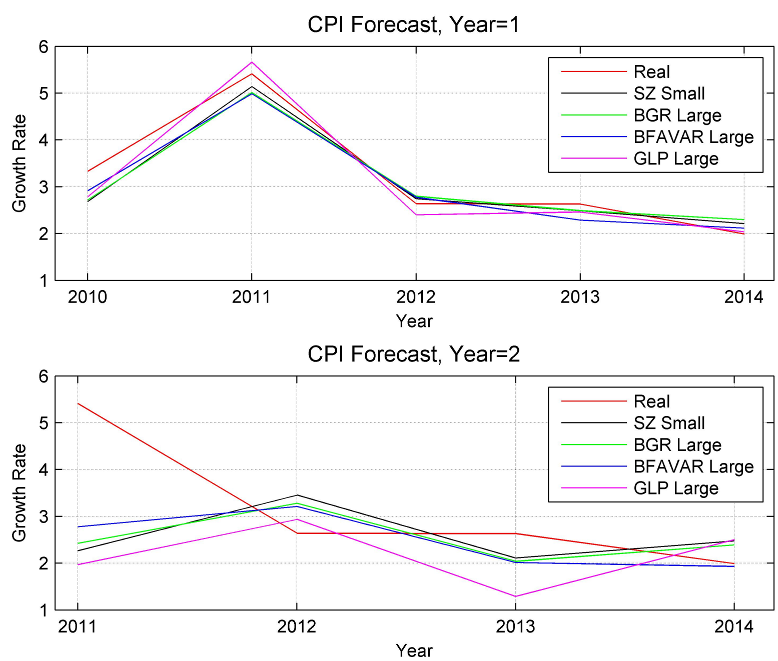

Figure 3.7: CPI Forecast, Panel B

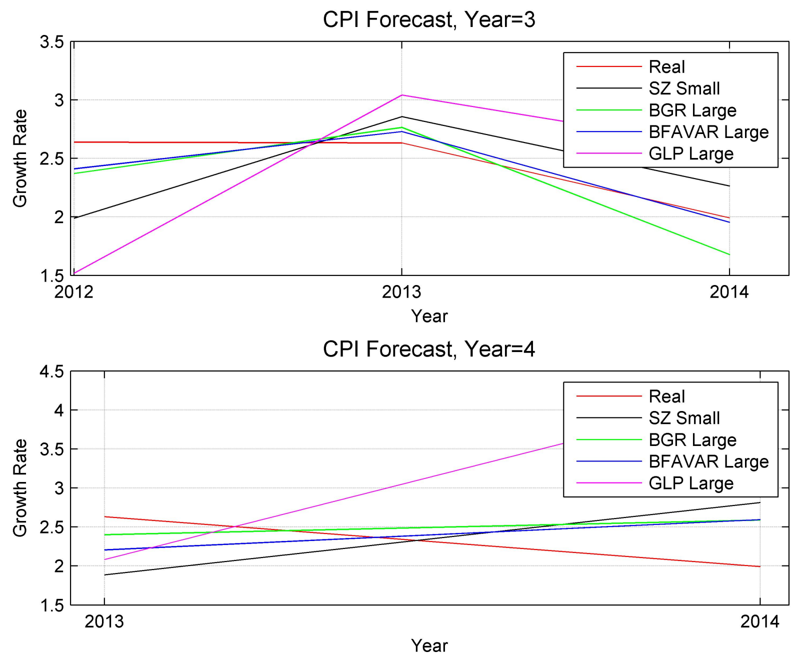

Figure 3.8: Interest Rate Forecast, Panel A

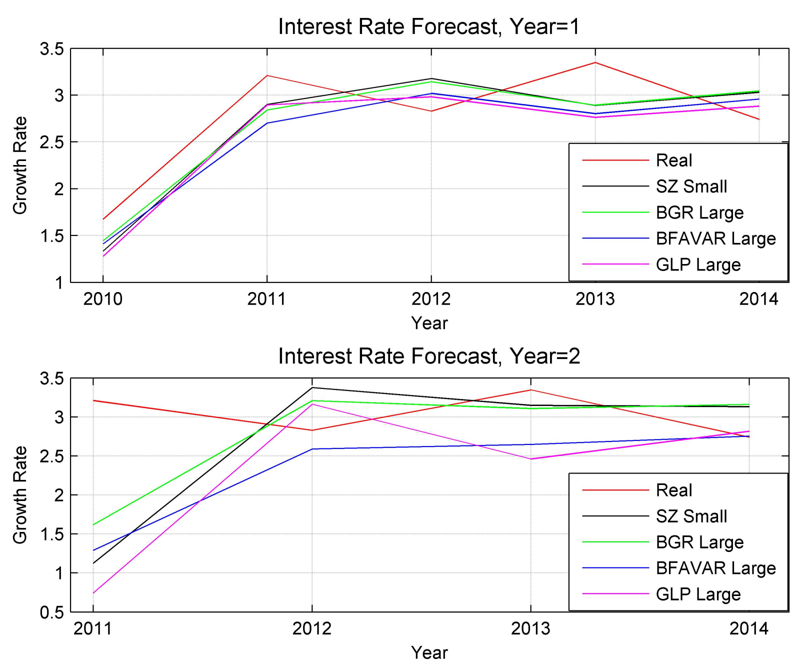

Figure 3.9: Interest Rate Forecast, Panel B

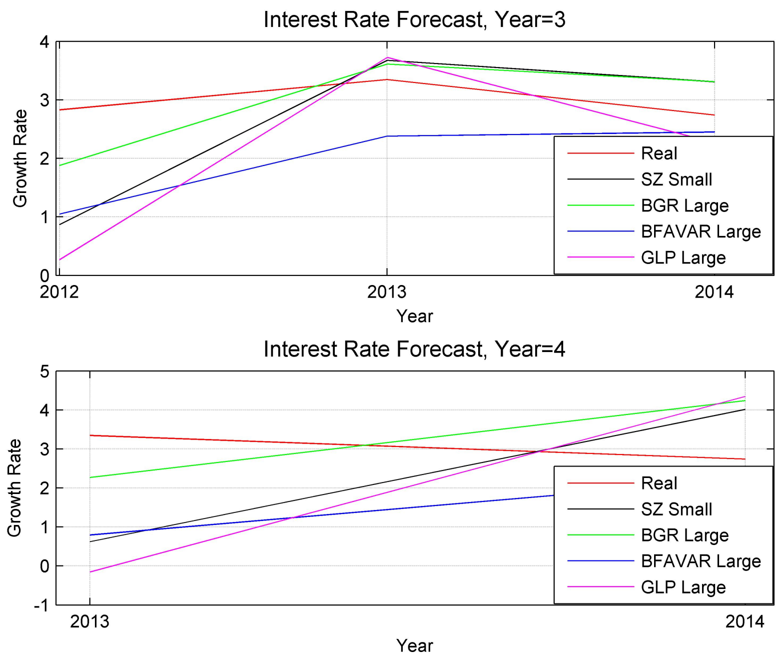

Table 3.1: Data from CQER

Variable Name	Log	Non-stat	SA
Real GDP by Value Added (RMB billion)	×	×	×
Consumer Price Index	×	×	×
1-day Repo Rate			
Real Net Export over GDP		×	×
Real Consumption	×	×	×
Real Investment	×	×	×
M2	×	×	
1-year Deposit Rate			
Retail Price Index	×	×	×
Fixed Asset Investment Price Index	×	×	×

Variable Name	Log	Non-stat	SA
Price Index for Gross Fixed Capital Formation	×	×	×
Implicit Price Deflator for GDP by Value Added	×	×	×
GDP by Expenditure (RMB billion)	×	×	×
Retail Sales of Consumer Goods (RMB billion)	×	×	×
Fixed Asset Investment (RMB billion)	×	×	×
Exports of Goods Reported by the Chinese Customs (RMB million)	×	×	×
Imports of Goods Reported by the Chinese Customs (RMB million)	×	×	×
Government Consumption by Expenditure (RMB billion)	×	×	×
Changes in Inventories (RMB billion)			×
Gross Fixed Capital Formation with No Inventories (RMB billion)	×	×	×
Gross Fixed Capital Formation: Government (RMB billion)	×	×	×
Gross Fixed Capital Formation: Households	×	×	×
Aggregate Average Nominal Wages	×	×	×
Reserve Money (RMB billion)	×	×	
M0 (RMB billion)	×	×	
Required Reserve Ratio		×	
Actual Reserve Ratio		×	
Excess Reserve Ratio		×	
Time Deposits Rate: 3 Months			

Variable Name	Log	Non-stat	SA
End-of-quarter Financial Institution Loans Outstanding: Total	×	×	
End-of-quarter Financial Institution Loans Outstanding: Short-term	×	×	
End-of-quarter Financial Institution Loans Outstanding: Medium and Long Terms	×	×	
New Bank Loans to Non-financial Enterprises (RMB billion): Short Term			
New Bank Loans to Non-financial Enterprises (RMB billion): Short Term and Bill Financing			
New Bank Loans to Non-financial Enterprises (RMB billion): Medium and Long Terms		×	

Table 3.2: Data from CEIC

Variable Name	Log	Non-stat	SA
Central Bank Benchmark Interest Rate: Loans to FI: Less Than 20 days			
Central Bank Benchmark Interest Rate: Loans to FI: 3 Month or Less			
Central Bank Benchmark Interest Rate: Loans to FI: 6 Month or Less			
Central Bank Benchmark Interest Rate: Loans to FI: 1 Year			

Variable Name	Log	Non-stat	SA
Nominal Lending Rate: Within 1 Year (Including 1 Year)			
Nominal Lending Rate: 1-5 Year (Including 5 Year)			
Nominal Lending Rate: 1-5 Year (Including 5 Year)			
Household Savings Deposits Rate: Demand			
Household Savings Deposits Rate: Time: 6 Month			
Household Savings Deposits Rate: Time: 2 Year			
Household Savings Deposits Rate: Time: 3 Year			
Interbank Offered Rate: Weighted Avg: 1 Month			
FX Rate: SAFE: Month Avg: RMB to USD	×		
FX Rate: SAFE: Month Avg: RMB to JPY	×	×	
Effective Exchange Rate Index: BIS: Real	×	×	
Effective Exchange Rate Index: BIS: Nominal	×	×	
Index: Shanghai Stock Exchange: Composite	×	×	
Index: Shanghai Stock Exchange: A Share	×	×	
Index: Shanghai Stock Exchange: Industrial	×	×	
Index: Shanghai Stock Exchange: Commerce	×	×	
Index: Shanghai Stock Exchange: Properties	×	×	
Turnover: Volume: Shanghai SE: Total	×	×	
Turnover: Volume: Shanghai SE: Stock	×	×	
Turnover: Volume: Shanghai SE: Stock: A Share	×	×	
Market Capitalization: Shanghai SE: Stock	×	×	

Variable Name	Log	Non-stat	SA
Market Capitalization: Shanghai SE: A Share	×	×	
Index: Shenzhen Stock Exchange: Composite	×	×	
Index: Shenzhen Stock Exchange: A Share	×	×	
Turnover: Volume: Shenzhen SE: Total	×	×	
Turnover: Volume: Shenzhen SE: Stock	×	×	
Turnover: Volume: Shenzhen SE: Stock: A Share	×	×	
Market Capitalization: Shenzhen SE: Stock	×	×	
Market Capitalization: Shenzhen SE: A Share	×	×	
PE Ratio: Shenzhen SE: All Share	×	×	
PE Ratio: Shenzhen SE: A Share	×	×	
Govt Revenue	×	×	×
Govt Revenue	×	×	×
Consumer Confidence Index	×	×	
Consumer Expectation Index	×	×	
Export sitc: Primary Product (PP)	×	×	×
Export sitc: Manufacture	×	×	×
Import sitc: Primary Product (PP)	×	×	×
Import sitc: Manufacture	×	×	×
Export: Asia	×	×	×
Export: Africa	×	×	×
Export: Europe	×	×	×
Export: Latin America	×	×	×
Export: North America	×	×	×
Export: Oceania	×	×	×
Import: Asia	×	×	×

Variable Name	Log	Non-stat	SA
Import: Africa	×	×	×
Import: Europe	×	×	×
Import: Latin America	×	×	×
Import: North America	×	×	×
Import: Oceania	×	×	×
Official Reserve Asset: Foreign Reserve(FR)	×	×	
Turnover: Interbank Loan: Total	×	×	
Turnover: Interbank Loan: 7 Day	×	×	
Turnover: Interbank Loan: 1 Month	×	×	
Turnover: Interbank Loan: 3 Month	×	×	
Property Price: YTD Avg: Overall	×	×	×
Import Price: Crude Oil	×	×	×
FAI: ytd: Primary Industry	×	×	×
FAI: ytd: Secondary Industry (SI)	×	×	×
FAI: ytd: Tertiary Industry	×	×	×
Real Estate Inv: ytd	×	×	×
Real Estate Inv: ytd: Residential Building	×	×	×
Real Estate Inv: ytd: Office Building	×	×	×
Real Estate Inv: ytd: Commercial Building	×	×	×
Real Estate Inv: ytd: Other Building	×	×	×
Building Sold: ytd	×	×	×
CPI: MoM: Food	×	×	×
CPI: MoM: Clothing	×	×	×
CPI: MoM: Household Facility, Article & Maintenance Service (HA)	×	×	×

Variable Name	Log	Non-stat	SA
CPI: MoM: Medicine, Medical Care & Personal Article (MP)	×	×	×
CPI: MoM: Transportation and Communication (TC)	×	×	×
CPI: MoM: Recreational, Educational, Cultural Article & Service (RE)	×	×	×
CPI: MoM: Residence	×	×	×
GDP: Primary Industry	×	×	×
GDP: Secondary Industry(SI)	×	×	×
GDP: SI: Industry	×	×	×
GDP: SI: Construction	×	×	×
GDP: Tertiary Industry(TI)	×	×	×
GDP: TI: Transport, Storage and Post	×	×	×
GDP: TI: Wholesale and Retail Trade	×	×	×
GDP: TI: Accommodation and Catering Trade	×	×	×
GDP: TI: Financial Intermediation	×	×	×
GDP: TI: Real Estate	×	×	×
GDP: TI: Other	×	×	×

Table 3.3: RMSE, Benchmark Model Panel A

Variable	Horizon	RW	AR	VAR
GDP	Year 1	0.30	0.47	0.73
	Year 2	1.22	2.03	2.57
	Year 3	2.11	3.50	3.87
	Year 4	2.76	4.63	7.41
CPI	Year 1	0.60	0.50	1.22
	Year 2	2.12	2.00	4.35
	Year 3	1.95	2.49	2.29
	Year 4	2.68	3.35	4.75
R	Year 1	0.73	0.52	0.86
	Year 2	1.76	0.95	2.45
	Year 3	2.26	0.88	3.88
	Year 4	3.91	1.19	6.80

Table 3.4: RMSE, Benchmark Model Panel B

Variable	Horizon	RW	Minnesota Revised	Sims&Zha
GDP	Year 1	0.30	0.90	0.22
	Year 2	1.22	2.42	0.73
	Year 3	2.11	2.97	1.93
	Year 4	2.76	3.55	2.51
CPI	Year 1	0.60	0.50	0.48
	Year 2	2.12	2.07	2.14
	Year 3	1.95	1.20	0.84
	Year 4	2.68	1.47	1.35
R	Year 1	0.73	0.65	0.58
	Year 2	1.76	1.53	1.36
	Year 3	2.26	1.89	1.70
	Year 4	3.91	2.73	2.68

Table 3.5: Relative RMSE, Benchmark Model

Variable	Horizon	AR	VAR	Minnesota Revised	Sims&Zha
GDP	Year 1	1.56	2.42	3.01	0.75
	Year 2	1.66	2.10	1.98	0.59
	Year 3	1.66	1.83	1.41	0.91
	Year 4	1.68	2.69	1.29	0.91
CPI	Year 1	0.84	2.05	0.83	0.81
	Year 2	0.94	2.05	0.98	1.01
	Year 3	1.28	1.18	0.61	0.43
	Year 4	1.25	1.78	0.55	0.50
R	Year 1	0.71	1.18	0.89	0.80
	Year 2	0.54	1.40	0.87	0.78
	Year 3	0.39	1.72	0.84	0.75
	Year 4	0.30	6.80	0.70	0.69

Table 3.6: RMSE, Large BVAR Model

			Small		Large
Variable	Horizon	RW	Minnesota Revised	Sims&Zha	BGR
GDP	Year 1	0.30	0.90	0.22	0.22
	Year 2	1.22	2.42	0.73	0.49
	Year 3	2.11	2.97	1.93	1.06
	Year 4	2.76	3.55	2.51	1.13
CPI	Year 1	0.60	0.50	0.48	0.39
	Year 2	2.12	2.07	2.14	1.91
	Year 3	1.95	1.20	0.84	0.63
	Year 4	2.68	1.47	1.35	0.82
R	Year 1	0.73	0.65	0.58	0.60
	Year 2	1.76	1.53	1.36	1.41
	Year 3	2.26	1.89	1.70	1.62
	Year 4	3.91	2.73	2.68	2.19

Table 3.7: Relative RMSE, Large BVAR Model

Variable	Horizon	Small		Large
		Minnesota Revised	Sims&Zha	BGR
GDP	Year 1	3.01	0.75	0.73
	Year 2	1.98	0.59	0.40
	Year 3	1.41	0.91	0.50
	Year 4	1.29	0.91	0.41
CPI	Year 1	0.83	0.81	0.66
	Year 2	0.98	1.01	0.90
	Year 3	0.61	0.43	0.32
	Year 4	0.55	0.50	0.31
R	Year 1	0.89	0.80	0.83
	Year 2	0.87	0.78	0.80
	Year 3	0.84	0.75	0.72
	Year 4	0.70	0.69	0.56

Table 3.8: Relative RMSE, Large BVAR Model with Different τ

Variable	Horizon	Test 1	Test 2	Test 3
		$\tau = 0.07$ $\alpha = 2$	$\tau = 0.05$ $\alpha = 2$	$\tau = 0.03$ $\alpha = 2$
GDP	Year 1	0.74	0.64	0.55
	Year 2	0.68	0.38	0.45
	Year 3	0.62	0.47	0.50
	Year 4	0.49	0.38	0.39
CPI	Year 1	0.87	0.65	0.68
	Year 2	0.77	0.86	0.80
	Year 3	0.33	0.27	0.22
	Year 4	0.35	0.28	0.28
R	Year 1	0.75	0.84	0.86
	Year 2	0.54	0.81	0.77
	Year 3	0.43	0.73	0.69
	Year 4	0.32	0.57	0.55

Table 3.9: Relative RMSE, Large BVAR Model with Different α

Variable	Horizon	Test 4	Test 5	Test 6
		$\tau = 0.07$ $\alpha = 2$	$\tau = 0.07$ $\alpha = 4$	$\tau = 0.07$ $\alpha = 5$
GDP	Year 1	0.74	0.43	0.44
	Year 2	0.42	0.36	0.39
	Year 3	0.50	0.46	0.48
	Year 4	0.40	0.37	0.37
CPI	Year 1	0.66	0.69	0.71
	Year 2	0.91	0.76	0.74
	Year 3	0.33	0.39	0.43
	Year 4	0.31	0.23	0.24
R	Year 1	0.83	0.84	0.85
	Year 2	0.80	0.74	0.73
	Year 3	0.71	0.69	0.68
	Year 4	0.56	0.54	0.53

Table 3.10: In-sample Fitness

Variable	Test 1	Test 3	Test 6
	$\tau = 0.07$ $\alpha = 2$	$\tau = 0.03$ $\alpha = 2$	$\tau = 0.07$ $\alpha = 5$
GDP	0.69	0.46	0.48
CPI	0.76	0.59	0.60
R	0.77	0.41	0.58
Variable	VAR3	VAR2	AR
GDP	0.64	0.53	0.48
CPI	0.77	0.70	0.44
R	0.78	NaN	0.41

Table 3.11: RMSE, Revised Large BVAR Model

Variable	Horizon	RW	BGR	BGR Revised
			$\tau = 0.07$ $\alpha = 2$	$\tau = 0.03$ $\alpha = 4$
GDP	Year 1	0.30	0.22	0.14
	Year 2	1.22	0.49	0.61
	Year 3	2.11	1.06	1.15
	Year 4	2.76	1.13	1.14
CPI	Year 1	0.60	0.39	0.48
	Year 2	2.12	1.91	1.60
	Year 3	1.95	0.63	0.30
	Year 4	2.68	0.82	0.58
R	Year 1	0.73	0.60	0.61
	Year 2	1.76	1.41	1.22
	Year 3	2.26	1.62	1.35
	Year 4	3.91	2.19	1.85

Table 3.12: Relative RMSE, Revised Large BVAR Model

Variable	Horizon	BGR	BGR Revised
		$\tau = 0.07$ $\alpha = 2$	$\tau = 0.03$ $\alpha = 4$
GDP	Year 1	0.73	0.46
	Year 2	0.40	0.50
	Year 3	0.50	0.54
	Year 4	0.41	0.41
CPI	Year 1	0.66	0.80
	Year 2	0.90	0.75
	Year 3	0.32	0.16
	Year 4	0.31	0.22
R	Year 1	0.83	0.84
	Year 2	0.80	0.69
	Year 3	0.72	0.60
	Year 4	0.56	0.47

Table 3.13: RMSE, Cointegration's Impact

Variable	Horizon	RW	BGR Revised $\eta = inf$	BGR Revised $\eta = 1$	BGR Revised $\eta = 0.3$
GDP	Year 1	0.30	0.14	0.12	0.53
	Year 2	1.22	0.61	0.51	1.48
	Year 3	2.11	1.15	1.05	1.92
	Year 4	2.76	1.14	1.01	1.81
CPI	Year 1	0.60	0.48	0.47	0.42
	Year 2	2.12	1.60	1.59	1.78
	Year 3	1.95	0.30	0.33	1.17
	Year 4	2.68	0.58	0.52	0.90
R	Year 1	0.73	0.61	0.62	0.59
	Year 2	1.76	1.22	1.24	1.21
	Year 3	2.26	1.35	1.40	1.39
	Year 4	3.91	1.85	1.92	1.79

Table 3.14: Relative RMSE, Cointegration's Impact

Variable	Horizon	BGR Revised $\eta = inf$	BGR Revised $\eta = 1$	BGR Revised $\eta = 0.3$
GDP	Year 1	0.46	0.41	1.77
	Year 2	0.50	0.42	1.21
	Year 3	0.54	0.50	0.91
	Year 4	0.41	0.36	0.66
CPI	Year 1	0.80	0.78	0.71
	Year 2	0.75	0.75	0.84
	Year 3	0.16	0.17	0.60
	Year 4	0.22	0.19	0.34
R	Year 1	0.84	0.85	0.82
	Year 2	0.69	0.71	0.69
	Year 3	0.60	0.62	0.61
	Year 4	0.47	0.49	0.46

Table 3.15: RMSE, Unit Root's Impact

Variable	Horizon	RW	BGR Revised $\mu = 10\tau$	BGR Revised $\mu = 1$	BGR Revised $\mu = 0.3$
GDP	Year 1	0.30	0.11	0.11	0.13
	Year 2	1.22	0.55	0.55	0.58
	Year 3	2.11	1.11	1.11	1.16
	Year 4	2.76	1.15	1.15	1.15
CPI	Year 1	0.60	0.47	0.47	0.49
	Year 2	2.12	1.56	1.56	1.58
	Year 3	1.95	0.66	0.66	0.29
	Year 4	2.68	0.57	0.57	0.51
R	Year 1	0.73	0.62	0.62	0.62
	Year 2	1.76	1.24	1.24	1.25
	Year 3	2.26	1.49	1.49	1.40
	Year 4	3.91	2.16	2.16	1.92

Table 3.16: Relative RMSE, Unit Root's Impact

Variable	Horizon	BGR Revised $\mu = 10\tau$	BGR Revised $\mu = 1$	BGR Revised $\mu = 0.3$
GDP	Year 1	0.41	0.36	0.42
	Year 2	0.42	0.45	0.48
	Year 3	0.50	0.53	0.55
	Year 4	0.36	0.42	0.42
CPI	Year 1	0.78	0.79	0.82
	Year 2	0.75	0.74	0.75
	Year 3	0.17	0.34	0.15
	Year 4	0.19	0.21	0.19
R	Year 1	0.85	0.85	0.85
	Year 2	0.71	0.71	0.71
	Year 3	0.62	0.66	0.62
	Year 4	0.49	0.55	0.49

Table 3.17: RMSE, Stationarity's Impact

Variable	Horizon	RW	BGR Revised Mixed	BGR Revised Non-Stat
GDP	Year 1	0.30	0.12	0.13
	Year 2	1.22	0.51	0.53
	Year 3	2.11	1.05	1.14
	Year 4	2.76	1.01	1.18
CPI	Year 1	0.60	0.47	0.46
	Year 2	2.12	1.59	1.59
	Year 3	1.95	0.33	0.33
	Year 4	2.68	0.52	0.58
R	Year 1	0.73	0.62	0.57
	Year 2	1.76	1.24	1.10
	Year 3	2.26	1.40	0.94
	Year 4	3.91	1.92	1.48

Table 3.18: Relative RMSE, Stationarity's Impact

Variable	Horizon	BGR Revised Mixed	BGR Revised Non-Stat
GDP	Year 1	0.41	0.43
	Year 2	0.42	0.43
	Year 3	0.50	0.54
	Year 4	0.36	0.43
CPI	Year 1	0.78	0.77
	Year 2	0.75	0.75
	Year 3	0.17	0.17
	Year 4	0.19	0.22
R	Year 1	0.85	0.78
	Year 2	0.71	0.63
	Year 3	0.62	0.41
	Year 4	0.49	0.38

Table 3.19: RMSE, BFAVAR Model with Different Algorithm

Variable	Horizon	BFAVAR3 Original	BFAVAR3 Revised	BFAVAR5 Original	BFAVAR5 Revised
GDP	Year 1	1.42	0.29	1.55	0.25
	Year 2	3.27	1.20	1.92	1.30
	Year 3	3.64	2.10	2.41	2.32
	Year 4	2.70	2.57	2.27	2.91
CPI	Year 1	0.52	0.41	0.48	0.42
	Year 2	2.21	1.38	2.34	1.41
	Year 3	1.27	0.93	1.46	0.87
	Year 4	0.86	1.38	0.73	1.45
R	Year 1	0.72	0.59	0.73	0.60
	Year 2	1.79	1.18	1.85	1.22
	Year 3	2.16	1.28	2.54	1.36
	Year 4	2.78	1.82	3.04	1.92

Table 3.20: Relative RMSE, BFAVAR Model with Different Algorithm

Variable	Horizon	FAVAR3 Original	FAVAR3 Revised	FAVAR5 Original	FAVAR5 Revised
GDP	Year 1	4.72	0.97	5.15	0.84
	Year 2	2.67	0.98	1.57	1.06
	Year 3	1.72	0.99	1.14	1.10
	Year 4	0.98	0.93	0.82	1.06
CPI	Year 1	0.88	0.68	0.81	0.71
	Year 2	1.04	0.65	1.10	0.66
	Year 3	0.65	0.48	0.75	0.45
	Year 4	0.32	0.52	0.27	0.54
R	Year 1	0.98	0.81	1.00	0.82
	Year 2	1.02	0.67	1.05	0.69
	Year 3	0.96	0.57	1.12	0.60
	Year 4	0.71	0.47	0.78	0.49

Table 3.21: RMSE, BFAVAR Model with Different Factors

Variable	Horizon	RW	BFAVAR1 Revised	BFAVAR3 Revised	BFAVAR5 Revised
GDP	Year 1	0.30	0.33	0.29	0.25
	Year 2	1.22	1.28	1.20	1.30
	Year 3	2.11	2.20	2.10	2.32
	Year 4	2.76	2.76	2.57	2.91
CPI	Year 1	0.60	0.41	0.41	0.42
	Year 2	2.12	1.31	1.38	1.41
	Year 3	1.95	0.80	0.93	0.87
	Year 4	2.68	1.25	1.38	1.45
R	Year 1	0.73	1.40	0.59	0.60
	Year 2	1.76	1.03	1.18	1.22
	Year 3	2.26	1.09	1.28	1.36
	Year 4	3.91	1.57	1.82	1.92

Table 3.22: Relative RMSE, BFAVAR Model with Different Factors

Variable	Horizon	BFAVAR1 Revised	BFAVAR3 Revised	BFAVAR5 Revised
GDP	Year 1	1.09	0.97	0.84
	Year 2	1.04	0.98	1.06
	Year 3	1.04	0.99	1.10
	Year 4	1.00	0.93	1.06
CPI	Year 1	0.70	0.68	0.71
	Year 2	0.62	0.65	0.66
	Year 3	0.41	0.48	0.45
	Year 4	0.47	0.52	0.54
R	Year 1	1.92	0.81	0.82
	Year 2	0.59	0.67	0.69
	Year 3	0.48	0.57	0.60
	Year 4	0.40	0.47	0.49

Table 3.23: RMSE, Stationarity on BFAVAR Model

		BFAVAR1	BFAVAR1	BFAVAR3	BFAVAR3
Variable	Horizon	Mixed	Non-stat	Mixed	Non-stat
GDP	Year 1	0.33	0.30	0.29	0.24
	Year 2	1.28	1.14	1.20	0.94
	Year 3	2.20	1.99	2.10	1.63
	Year 4	2.76	2.46	2.57	1.91
CPI	Year 1	0.41	0.46	0.41	0.45
	Year 2	1.31	1.36	1.38	1.43
	Year 3	0.80	0.26	0.93	0.24
	Year 4	1.25	0.36	1.38	0.53
R	Year 1	1.40	1.45	0.59	0.57
	Year 2	1.03	1.14	1.18	1.20
	Year 3	1.09	1.21	1.28	1.29
	Year 4	1.57	1.79	1.82	1.94

Table 3.24: Relative RMSE, Stationarity on BFAVAR Model

		BFAVAR1	BFAVAR1	BFAVAR3	BFAVAR3
Variable	Horizon	Mixed	Non-stat	Mixed	Non-stat
GDP	Year 1	1.09	1.00	0.97	0.81
	Year 2	1.04	0.93	0.98	0.77
	Year 3	1.04	0.94	0.99	0.77
	Year 4	1.00	0.89	0.93	0.69
CPI	Year 1	0.70	0.77	0.68	0.75
	Year 2	0.62	0.64	0.65	0.67
	Year 3	0.41	0.13	0.48	0.12
	Year 4	0.47	0.14	0.52	0.20
R	Year 1	1.92	2.00	0.81	0.79
	Year 2	0.59	0.65	0.67	0.68
	Year 3	0.48	0.53	0.57	0.57
	Year 4	0.40	0.46	0.47	0.50

Table 3.25: RMSE, Hierarchical Model

		GLP Original	GLP Original	GLP Revised	GLP Revised
Variable	Horizon	Mixed	Non-stat	Mixed	Non-stat
GDP	Year 1	0.40	0.27	0.34	0.21
	Year 2	1.26	0.90	0.97	0.86
	Year 3	2.05	2.50	2.04	1.87
	Year 4	3.65	3.55	3.67	2.37
CPI	Year 1	0.52	0.51	0.46	0.45
	Year 2	2.26	1.84	1.97	1.98
	Year 3	4.56	1.49	3.38	0.97
	Year 4	2.01	2.19	2.12	1.96
R	Year 1	0.49	0.68	0.66	0.65
	Year 2	1.48	1.51	1.73	1.71
	Year 3	2.93	1.57	2.54	2.01
	Year 4	4.80	2.41	3.09	3.40

Table 3.26: Relative RMSE, Hierarchical Model

		GLP Original	GLP Original	GLP Revised	GLP Revised
Variable	Horizon	Mixed	Non-stat	Mixed	Non-stat
GDP	Year 1	1.34	0.88	1.14	0.69
	Year 2	1.03	0.74	0.80	0.70
	Year 3	0.97	1.18	0.97	0.88
	Year 4	1.32	1.29	1.33	0.86
CPI	Year 1	0.86	0.85	0.77	0.76
	Year 2	1.06	0.87	0.93	0.93
	Year 3	2.34	0.77	1.74	0.50
	Year 4	0.75	0.82	0.79	0.73
R	Year 1	0.67	0.93	0.90	0.90
	Year 2	0.84	0.86	0.99	0.97
	Year 3	1.29	0.69	1.12	0.89
	Year 4	1.23	0.62	0.79	0.87

Table 3.27: RMSE, Champions of Models

			Small	Large		
Variable	Horizon	BCEI	Sims&Zha	BGR	BFAVAR	GLP
GDP	Year 1	0.39	0.19	0.13	0.24	0.21
	Year 2	0.72	0.74	0.53	0.94	0.86
	Year 3	NA	1.13	1.14	1.63	1.87
	Year 4	NA	1.26	1.18	1.91	2.37
CPI	Year 1	0.63	0.43	0.46	0.45	0.45
	Year 2	1.49	1.69	1.59	1.43	1.98
	Year 3	NA	0.46	0.33	0.24	0.97
	Year 4	NA	0.80	0.58	0.53	1.96
R	Year 1	NA	0.57	0.57	0.57	0.65
	Year 2	NA	1.30	1.10	1.20	1.71
	Year 3	NA	1.41	0.94	1.29	2.01
	Year 4	NA	2.28	1.48	1.94	3.40

Table 3.28: Relative RMSE, Champions of Models

			Small	Large		
Variable	Horizon	BCEI	Sims&Zha	BGR	BFAVAR	GLP
GDP	Year 1	1.30	0.63	0.43	0.81	0.69
	Year 2	0.59	0.61	0.43	0.77	0.70
	Year 3	NA	0.53	0.54	0.77	0.88
	Year 4	NA	0.46	0.43	0.69	0.86
CPI	Year 1	1.05	0.73	0.77	0.75	0.76
	Year 2	0.70	0.80	0.75	0.67	0.93
	Year 3	NA	0.24	0.17	0.12	0.50
	Year 4	NA	0.30	0.22	0.20	0.73
R	Year 1	NA	0.78	0.78	0.79	0.90
	Year 2	NA	0.74	0.63	0.68	0.97
	Year 3	NA	0.62	0.41	0.57	0.89
	Year 4	NA	0.58	0.38	0.50	0.87