UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE

https://archive-ouverte.unige.ch

Thèse 2022

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Three Essays on Corporate Finance

Zhang, Ye

How to cite

ZHANG, Ye. Three Essays on Corporate Finance. Doctoral Thesis, 2022. doi: 10.13097/archive-ouverte/unige:165135

This publication URL: https://archive-ouverte.unige.ch/unige:165135

Publication DOI: 10.13097/archive-ouverte/unige:165135

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.

Three Essays on Corporate Finance

Trois Essais Sur La Finance D'entreprise

THESIS
submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,

by

Ye ZHANG

Under the direction of
Prof. Harald HAU, supervisor

in fulfillment of the requirements for the degree of

Docteur ès économie et management
mention *finance*

Jury members:
Prof. Fabio TROJANI, Chair, University of Geneva
Prof. Harald HAU, Supervisor, University of Geneva
Prof. Boris NIKOLOV, HEC Lausanne, University of Lausanne

Thesis No. 115
October 2022

La Faculté d'économie et de management, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 5 octobre 2022

Dean
Markus MENZ

Impression d'après le manuscrit de l'auteur

Acknowledgements

I would like to start by thanking Prof. Harald Hau for his invaluable guidance, continuous support and giving me full freedom in research for the past six years. I am very grateful to other committee members: Prof. Fabio Trojani and Prof. Boris Nikolov. Their attitudes towards research and teaching deeply affect me. Their constructive suggestions significantly improve the thesis.

Respect also goes to all the teachers who have taught me in the past twenty three years. When three men meet together, one of them who is anxious to learn can always learn something of the other two. So moreover, I am thanking everyone in my life.

The sincerest gratitude goes to my coauthors: Prof. Mathias Hoffmann, Prof. Yi Huang and Rustam Abuzov. Working with them is always a pleasure. Their critical thinking and suggestions not only improve the research but also make me a better person.

I am grateful to all my colleagues and friends. Their support and caring make my life in Switzerland colorful, interesting and meaningful.

Finally, special thanks to my family for their infinite patience and unconditional support.

Abstract

This thesis focuses on empirical corporate finance by studying several entities' decision-making. In the first chapter, we investigate how local governments in China borrow from banks and at the end drive up housing prices. The mechanism is that local government financing vehicles could purchase land parcels at cheaper prices from the local governments and inflate the collateral valuation to borrow from banks. In the second chapter, we study whether negative sentiments shown in the IPO prospectus are related to the long-run returns of IPO firms. We find that there is a positive link and the impact of negative sentiments on retail investor attention could explain this relationship. In the third chapter, I examine the role that female solidarity plays on internal resource allocation in the context of Chinese listed firms. The results show that the same-gender linkages between female corporate executives and female political leaders positively influence the location of production decisions for Chinese exports. Various heterogeneity analyses support the explanation that female connection mitigates information friction.

Résumé

Cette thèse porte sur la finance d'entreprise empirique en étudiant plusieurs entités dans leur prise de décision. Dans le premier chapitre, nous examinons comment les gouvernements locaux en Chine empruntent aux banques et, à la fin, font monter les prix des logements. Notre histoire est que les véhicules de financement des gouvernements locaux pourraient acheter des parcelles de terrain à des prix moins chers auprès des gouvernements locaux et gonfler la valeur de la garantie pour emprunter auprès des banques. Dans le deuxième chapitre, nous étudions si les sentiments négatifs affichés dans le prospectus d'introduction en bourse sont liés aux rendements à long terme des entreprises introduites en bourse. Nous constatons qu'il existe un lien positif et que la relation passe par l'impact des sentiments négatifs sur l'attention des investisseurs particuliers. Dans le troisième chapitre, j'examine le rôle de la solidarité féminine sur l'allocation des ressources internes dans le contexte des entreprises cotées chinoises. Les résultats montrent que les liens de même sexe entre les femmes dirigeantes d'entreprise et les femmes dirigeantes politiques influencent positivement la localisation des décisions de production pour les exportations chinoises. Diverses analyses d'hétérogénéité soutiennent l'explication selon laquelle la connexion féminine atténue la friction de l'information.

Contents

Introduction

This thesis empirically analyzes the decision making of various entities in corporate finance. The first chapter focuses on the local government in China. The second chapter discusses the prospectus and long-term performance of IPOs in the United States. And the third chapter examines the internal resource allocation of China's export firms.

The Chinese government started a 4-trillion RMB fiscal stimulus package in November 2008 as a response to the global financial crisis. Specifically, the central government in China instructed the big four government-owned banks to expand lending. In the first chapter, we exploit a granular land transaction-level dataset to show how this lending boom interacted with monopolistic land supply by local governments to fuel China's post-2009 housing boom. We find that local governments with political connections benefited disproportionately from the lending boom by borrowing via local government financing vehicles (LGFVs). LGFVs purchased land from local governments with lower prices and posted land as collateral at inflated valuations. However, given the role of local governments as monopolistic suppliers of land usufruct rights, this policy also induced a scarcity of residential land supply to the private sector, and at the end, drove up land and house prices.

Literature shows that sentiments displayed in various corporate reports and news predict short-term returns and earnings. Although news are quite short-term, annual basis 10-K filings reflect firms comprehension of their future prospect as well as risks they may face, and S-1, a filing for SEC and investors to have more knowledge of the IPO firm, should be more long-term viewed. The second chapter examines the effect of sentiments in S-1 filings on IPOs' long-term performance for U.S. stocks listed during 1997-2016. We find that IPO stocks with higher (abnormal) negative sentiments displayed in the prospectus significantly outperform those with lower levels over the long run, using both pooled data and calendar time portfolio approaches. A variety of sentiment and performance measures reinforces the above finding. With retail investor attention as the mediator variable, we find that negative sentiment in the prospectus is associated with better long-term performance because IPOs demonstrating more negative sentiment obtain less retail investor attention and thus less buying pressure during IPO, and in the long run, when the demand from retail investors dissipates, there is a reversal in stock returns.

How do corporations make decisions about internal resource allocation? Does political connection play a role, especially in emerging economies with incomplete financial markets? In the third chapter, I explore the impact of linkages between corporate executives and local political leaders through gender commonality on decisions of production locations for export in the context of Chinese listed firms. I present evidence that same-gender linkages between female corporate executives and female political leaders positively influence the location of production decisions for Chinese exports. The effect only exists in financially constrained firms. By examining the ex-post performance, I find that firms with female connections perform better, even if the connection does not exist afterwards. This effect is more significant if the firm is younger, smaller, more concentrated, and faces fiercer local competitions. These results suggest that the most reasonable explanation is that female connection mitigates information friction.

In sum, the three chapters in the dissertation explain 1) how the local government borrows from bank when it has monopolistic control on land supply market in China; 2) how IPO firms draft their prospectuses during IPO process to attract retail investor attention and achieve long-term returns; 3) how female solidarity affects Chinese export firms' decision making on the location of productions.

Chapter 1

Super Debt Cities and the Geography of Housing Boom: the Role of Local Government Financing Vehicles[1]

Abstract

During the global financial crisis, the Chinese government instructed the big four government-owned banks to expand lending. We show that cities with strong political connections benefited more from the lending boom, and borrowed more through local government financing vehicles (LGFVs). By exploiting a granular land transaction-level dataset, we show the fundamental mechanism of local government finance: the monopoly power of land supply. Therefore, LGFVs could purchase the land with cheaper price from the local governments and inflate collateral valuation from the local banks. As a result, this lending boom contributed significantly to the house-price boom across China since 2009.

Key words: Fiscal Shocks, Credit Allocation, Real Estate, LGFV.
JEL codes: E50, E02, G28, H81, N25, O23, R30.

1.1 Introduction

What is the effect of an exogenous credit supply shock on the economy and on land prices in particular? A large and still growing literature has shown that exogenous credit supply shocks tend to have considerable effects on the real economy. In particular, there is considerable evidence from developed economies that positive credit supply shocks tend to increase prices for land and thus house prices. The transmission channel emphasized in

[1]This paper is coauthored with Prof. Mathias Hoffmann and Prof. Yi Huang. Mathias Hoffmann, University of Zurich. Yi Huang, Fudan University. We thank Zhiguo He, Yongheng Deng, Laura Liu, Wenlan Qian, Yu Qin, Wei Huang, Hong Ru, Choi Hyun Soo, Michael Weber and participants at the 7th Asian Bureau of Finance and Economic Research Annual Meeting and National University of Singapore, seminars at University of Geneva and Graduate Institute-Geneva.

this literature is that an increased supply of credit lowers lending rates, thus stimulating demand for mortgage loans and driving up house prices.

Much less is known about the impact of exogenous credit supply shocks on housing and land prices in emerging economies, though. Banking markets in emerging economies are often subject to considerable friction. In many countries, small private firms are de facto excluded from credit from the formal banking sector or may only have access to such credit if they can post housing or land as collateral. At the same time, the sale of land is an important source of government revenue in many emerging economies. This makes the process of land supply potentially subject to considerable political interference.

A key requirement for increases in housing demand to trigger relevant increases in house prices at all is that housing supply is sufficiently inelastic. Arguably, physical and geographical restrictions limit the supply of developable land in an exogenous way. For regulatory restrictions this is not obviously the case. In developed economies with their high levels of legal and democratic scrutiny, regulatory restrictions may be very hard to change and housing supply may therefore indeed be inelastic, at least in the short-run. In developing economies, however, such scrutiny is less direct and may sometimes not exist at all. In particular if land sales are an important source of government revenue, there is a strong incentive for local governments to expand the *de facto* supply of land by changing regulations, such as the designation of land use, in response to changes in demand. All of these considerations let us expect that the link between credit supply and housing prices is much different in an emerging economy.

In this paper, we exploit the quasi-natural experiment of China's 2009 fiscal stimulus to explore the mechanisms through which a rapid expansion in credit affects land and housing prices in an emerging economy. While officially designated as a "fiscal" program, this stimulus largely consisted of a gigantic expansion of credit by the the big four government-owned banks.

We argue that this credit supply shock laid the foundation of China's housing boom of the following years. It did so not primarily by stimulating household demand for credit but by changing the incentives for local officials to supply residential land to the primary market.

We identify the causal impact of the aggregate lending boom on local economies, by exploiting variation across cities in the degree of political connections to the central government. We first show that cities with strong political connections saw the largest increases in bank lending. They did so mainly because of an increase in the debt of local government financing vehicles (LGFVs), while credit to other state-owned enterprises (SOE) or domestic private firms did not increase to the same extent.

To explain our findings, we recognize that promotion of local officials in China depends on local economic performance. The lending boom provided incentives to local officials to quickly expand their local economies. While local governments are not allowed to directly borrow in financial markets, they are the monopoly suppliers of land usufruct rights. However, in many lower-tier cities, selling land to private enterprise would not have raised sufficient funds for a quick expansion of government expenditure. Therefore, local governments used local government financing vehicles which borrowed massively by posting residential land as collateral at inflated valuations. This quickly raised funds for infrastructure and industrial development, but led to a scarcity of residential land, driving up prices.

We use a unique transaction-level data set to identify this mechanism. Specifically, we compile information of land parcels for land sales and collateralization from the Ministry of Land and Resources. Based on this dataset, we show that the valuation of land as collateral vis-à-vis the lending bank was considerably higher in transactions by LGFVs in politically connected cities. These results are robust to very granular controls for differences in the quality of the land and for a host of city-level characteristics that could also have affected the transmission of the lending supply shock. Our findings suggest that officials in politically connected cities could indeed raise larger amounts from the stimulus than officials in other cities because, under the shadow of political protection, they could post land as collateral at inflated valuations.

In China's system of local government finance, funds in off-balance sheet LGFVs are much more flexible in their use than funds allocated through the official budgeting process that is under relatively close control of the central government. Hence, funds in LGFVs can more easily be used for the pursuit of the private career objectives of local officials. This gives these officials not only a strong incentive to raise funds for their LGFVs from the stimulus, but also to transfer land use rights from the local government to the LGFV at possibly low prices. Based on our transaction-level data set, we find that the transfer of land usufruct rights from local governments to their LGFVs did indeed happen at substantial discounts vis-à-vis the prices attained for comparable plots of lands sold to the private sector or in cities with weaker political connections.

Our paper relates to the literature along multiple dimensions. Mainly, we connect two large and important strands of the literature. Both questions have been attracted heating debates, yet there is no consensus on the answer: one that exploits regional variation of credit supply to explain house price movements; the other that studies the relationship between government borrowing and real effects from financial market or firm responses.

To what extent did the credit supply drive the housing boom and bust? This question is crucial to understanding the housing price cycles that impacts the macro fluctuations. Recent papers, such as Adelino, Schoar, and Severino (2012), Favara and Imbs (2015), Di Maggio and Kermani (2017) and Favilukis, Ludvigson, and Van Nieuwerburgh (2017), find that changes in credit conditions can explain the majority of the movements in house prices. By contrast, other papers, such as Kaplan, Mitman, and Violante (2017), argue that credit supply explains none of the boom and bust in house prices, which are instead driven by housing demand and expectations. Our paper studies the novel land allocation channel of the credit condition due to local government borrowing in China, which could explain the housing price movement puzzles.

Many studies have also discussed the real effects of government debt. Cohen, Coval, and Malloy (2011) and Graham, Leary, and Roberts (2014) show that government debt will crowd out corporate debt and investment by affecting investors' choices and the relative price of assets in the US. Greenwood, Hanson, and Stein (2010), Greenwood, Hanson, and Stein (2015), Krishnamurthy and Vissing-Jorgensen (2011), Krishnamurthy and Vissing-Jorgensen (2015) show the substitution effect between corporate debt and government debt. Becker and Ivashina (2018) document that rising domestic government bond holdings reduced corporate lending during the European sovereign debt crisis. Based on the cross-country data, Demirci, Huang, and Sialm (2019) shows that corporate leverage is lower in countries with higher government debt. Our paper extends all this research by linking local government borrowing through the land supply allocations and the effects on the asset pricing, particularly, the housing price.

Our paper also makes contributions to the burgeoning literature on China's local government debt. Huang, Pagano, and Panizza (2020) and Cong et al. (2019) provide similar evidence that local government debt crowds out private investment and product of capital in China. Ambrose, Deng, and Wu (2015) and Ang, Bai, and Zhou (2016) investigate two risk characteristics - real estate and political risk - in municipal corporate bond pricing. Bai, Hsieh, and Song (2016) and Chen, He, and Liu (2020) show the link between shadow banking activities and municipal corporate bond issuance. Liu, Lyu, and Yu (2017) investigates how local government fiscal conditions influence municipal corporate bond spreads. Our paper complements all this research by linking local government debt and the unintended consequence of Chinese housing booms.

The paper is organized as follows: Section 1.2 makes an introduction about the stimulus plan in 2009 and China's local government; Section 1.3 constructs a simple model to explain our story; Section 1.4 describes various data sources and variable constructions; in Section 1.5, we show the empirical results and then Section 1.6 provides some robustness results; the paper concludes in Section 1.7.

1.2 Background

1.2.1 Stimulus Plan in 2009

Chinese government introduced a stimulus plan to mitigate the slowdown of economic growth following the global financial crisis. This plan is composed of two parts: fiscal stimulus, also known as 4 trillion RMB plan, which amounts to about 15% of Chinese GDP in 2007; and credit stimulus, whose main objective is to inject liquidity to the banking system. The fiscal plan, officially announced by Premier Wen Jiabao on November 9, 2008, requires central government and local governments to spend 4 trillion RMB before the end of 2010 on infrastructure and social welfare projects. 1.18 trillion RMB was supposed to be funded by central government, and local governments have to finance the remaining 2.82 trillion RMB.

To support the 4 trillion investment plan, Chinese government implemented several policies to increase liquidity on banking system and as a result on the real economy. The People's Bank of China reduced commercial banks' required reserve ratio four times from 17.5% to 13.5% for small and medium sized banks, and from 17.5% to 15.5% for large banks from September 2008 until January 2010. The People's Bank of China eliminated the hard constraints on the credit quotas of commercial banks. The People's Bank of China also reduced the base one-year deposit and lending rate by 1.08 basis point in November 2008. Figure 1.1 and 1.2 show the lending balance and new credit of government-owned big four banks for the period 2005 to 2016[2]. From 2008 to 2009, due to the stimulus plan, the total loan balance increased from 14.32 trillion RMB to 19.08 trillion RMB with an increase rate of 33.24%. The year 2010 also experienced a soaring growth rate of 17.76% with a rise of total loan volume by 3.39 trillion RMB.

1.2.2 Local Government in China

Since the 1994 Budget Law, local governments in China are not allowed to borrow from commercial banks or to issue municipal bonds. However, the stimulus plan in 2009 required local government to finance up to 2.82 trillion RMB. To achieve this goal, central government issued bonds for local governments. The total amount is 200 billion

[2] The big four banks include: Bank of China, Industrial and Commercial Bank of China, China Construction Bank, and Agricultural Bank of China. New credit is defined as the difference of loan volume at the end of the period.

Figure 1.1: Lending Boom of Big 4 Banks: Loan Balance

The figure shows the loan balance and loan growth rate of the Big 4 banks which include: Bank of China, Industrial and Commercial Bank of China, China Construction Bank, and Agricultural Bank of China, for the period 2005 to 2016.

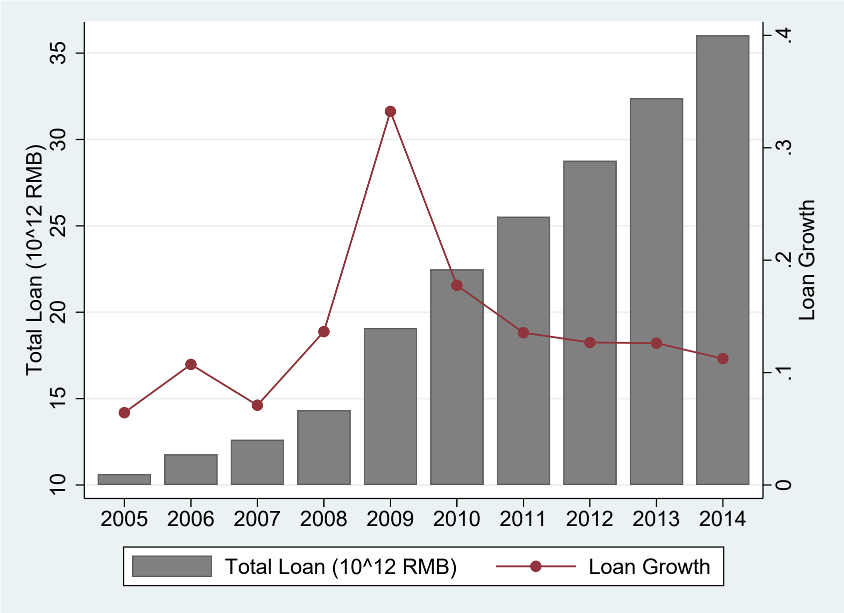
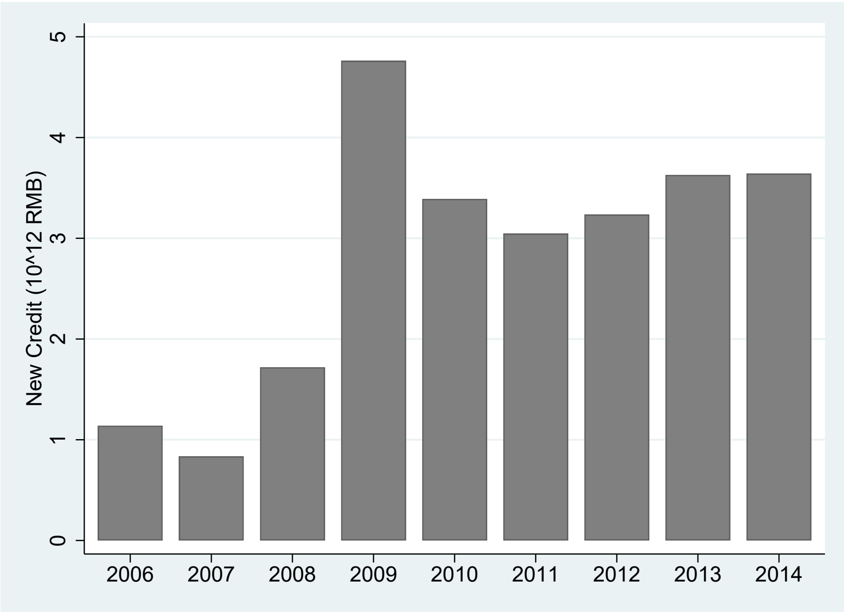


Figure 1.2: Lending Boom of Big 4 Banks: New Credit

he figure shows the new credit of the Big 4 banks for the period 2006 to 2014, where the new credit is defined as the difference of loan balance at the end of the period. The Big 4 banks include: Bank of China, Industrial and Commercial Bank of China, China Construction Bank, and Agricultural Bank of China.



RMB. More importantly, local governments established local government financing vehicles (LGFVs) to finance for investment projects. LGFV is an off-balance-sheet investment and financial platform to raise money for local government to invest in infrastructure projects. LGFV is a local state-owned-enterprise (SOE) with financial allocations, land, equity, or treasury bonds as its main asset.

In the March of 2009, the People's Bank of China and the China Banking Regulatory

Commission jointly proposed: "Support qualified local governments to establish financing platforms, issue corporate bonds, medium-term notes and other financing tools, and expand the financing channels for central government to fund investment projects". LGFVs borrowed from banks with local governments' explicit guarantees of debt repayments, and future revenues from local governments' land sales were used as collateral for bank loans (Liu and Xiong (2018)). Gao, Ru, and Tang (2020) estimated that bank loans account for more than 90% of new debts issued by LGFVs during the post-2008 stimulus period. According to Huang, Pagano, and Panizza (2016), local government debt almost quadrupled from 5.8% to 22% of GDP between 2006 and 2013. Due to the soaring of local government debt, on June 10, 2010, the State Council issued the "Notice on Strengthening the Management of Local Government Financing Platform Companies": LGFVs that are only responsible for public welfare projects and mainly rely on fiscal funds to repay debts will not be able to undertake financing tasks in the future; LGFVs with no stable cash flow can not borrow from commercial banks; local governments are strictly prohibited from providing illegal implicit guarantees to LGFVs[3].

With the fact that local government is the monopolistic supplier of land usufruct rights, during the stimulus period from 2009, a main arrangement for local government is that local government transferred land usufruct rights to LGFVs, and then LGFVs pledged these land parcels as collateral to borrow from banks. This arrangement is even more prevalent after June 2010 notice to meet the capital requirement[4].

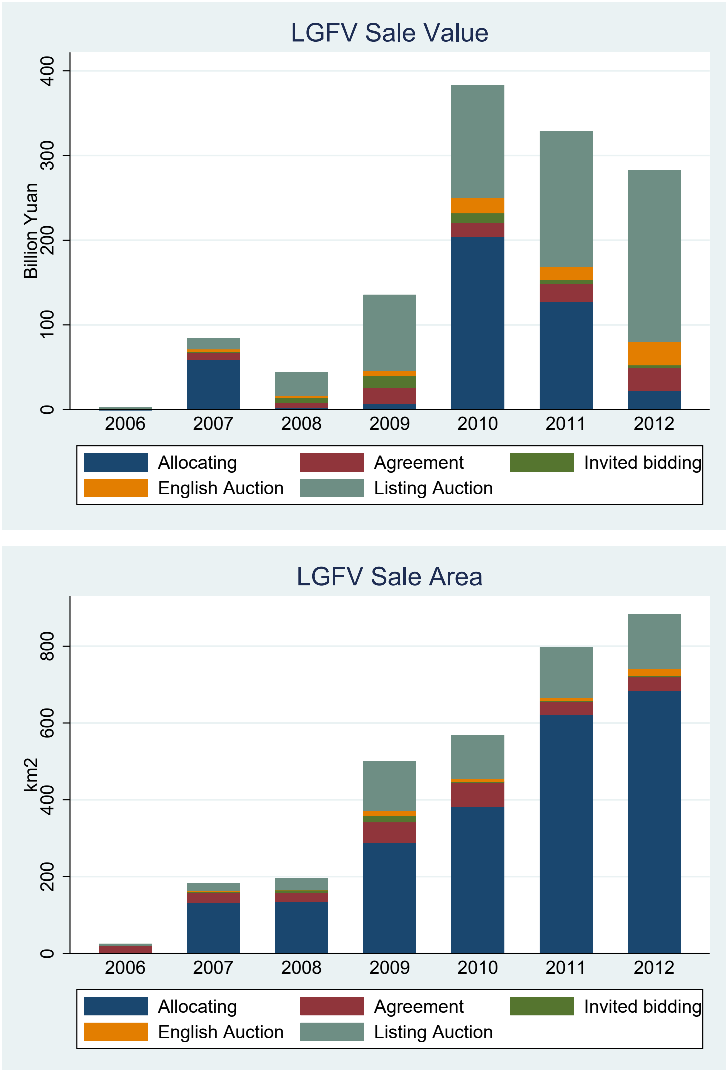
Figure 1.3 shows the value and area of land parcels purchased by LGFVs. There is a jump of area sold to LGFVs in 2009. Most of the land parcels were transferred to LGFVs through allocating and bilateral agreement and in most cases the price is much lower than the market price in these two supplying methods. Listing auction accounts for the major part for remaining land parcels provided to LGFVs and most of the revenue comes from listing auctions. This makes sense as listing auction is the most frequently used method of local government to sell land usufruct rights.

Figure 1.4 shows the LGFVs' collateral value and area with land parcels as collateral. The LGFVs' collateral value of land parcels has a sharp increase in 2008 and 2009, from 15.57 billion RMB in 2007 to 113.02 billion RMB in 2008, to 1150 billion RMB in 2009. This rise of value is a joint effect of collateral area and unit value rises. To be specific, the total collateral area of land parcels increases from 3.62 square kilometers in 2007 to 13.97 square kilometers in 2008, and then reaches the peak value of 119.75 square kilometers in 2009. At the same time, the unit value per square meter increases from 4,300 RMB in 2007 to more than 8,000 RMB in 2008 and 2009. Most of the proceeds from land collateralization are from residential land, commercial land and other types of land[5]. Figure 1.5 shows the average collateralization price for LGFVs and non-LGFVs of different land usages. The average collateralization price of industrial land is roughly similar for LGFVs and non-LGFVs. However, LGFVs benefit from higher collateralization unit price for residential land, commercial land and other types of land.

[3]See http://www.gov.cn/zwgk/2010-06/13/content_1627195.htm.
[4]See Bai, Hsieh, and Song (2016).
[5]Other types of land includes public land parcels, idle land parcels and land parcels with undefined usages.

Figure 1.3: Land Purchase by LGFVs

This figure shows the aggregate revenue and land area from the land sale transactions by LGFVs from the year 2006 to 2012, with different colors representing different supply methods which include allocating, bilateral agreement, invited biding, English auction and listing auction.

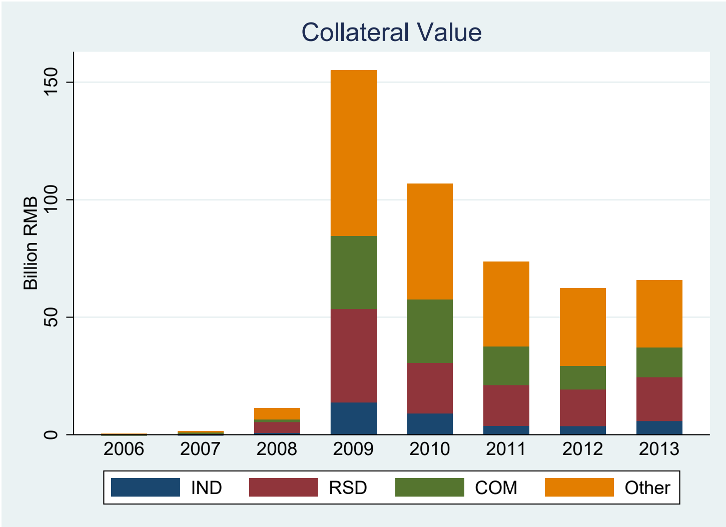


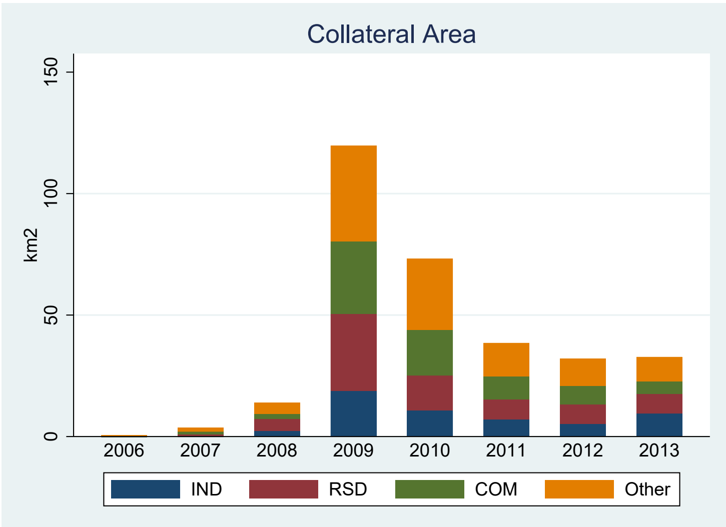
1.3 Theoretical Model

We consider a local government that aims to maximize the funds it can raise from land. It can do so in two ways. First, by selling land usufruct rights to the private market. Since the local government is the monopoly supplier of land use rights, it faces a downward sloping demand curve from the private sector such that the price is given by $P(Q)$ with $dP/dQ < 0$ where Q is the quantity provided to the market. Alternatively, the government can raise funds by borrowing, collateralizing land at an exogenous valuation V. Our model is static and we do not consider the implications that raising funds by collateralization may have for future net government revenue. This assumption reflects the short-term nature of the incentive structure of local officials in China who have to meet short-run growth targets in order to achieve promotions. We therefore believe that it is reasonable to assume that officials' objective is to maximize the funds available for the pursuit of these short-run objectives.

Figure 1.4: Land Collateralization by LGFVs

This figure shows the aggregate land value and land area from the land collateral transactions by LGFVs from the year 2006 to 2012, with different colors representing different types of land, including industrial land(IND), residential land(RSD), commercial land(COM), and all other types of land(Other).





In the baseline setting, providing land to the market is assumed to have zero marginal cost and each period, the government can sell one unit of land only. Then the fiscal revenue of the government becomes:

$$R(Q) = P(Q) \times Q + (1 - Q)V$$

The FOC is

$$\frac{dP}{dQ}Q + P(Q) - V = 0$$

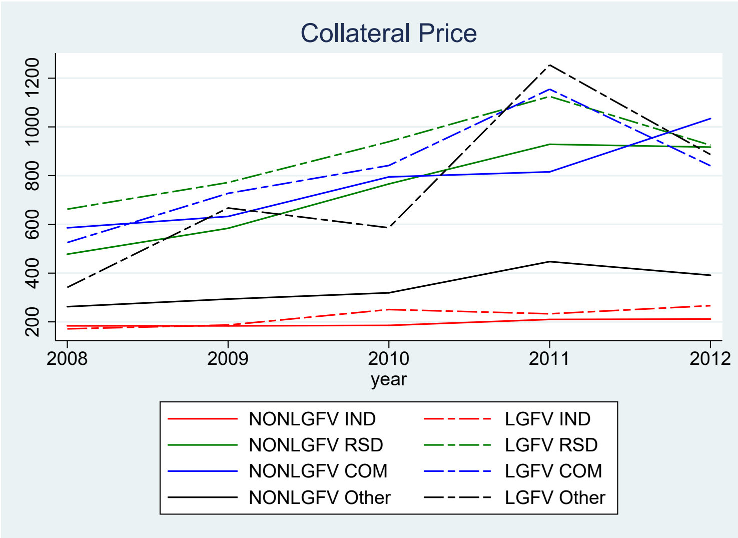
or, defining the elasticity $\epsilon = -\frac{dP}{dQ} \times \frac{Q}{P} > 0$:

$$1 - \epsilon = \frac{V}{P}$$

It is instructive to first consider the case when $V = 0$. Then we have $\epsilon = 1$, the standard result that a monopolist with zero marginal cost will supply land until the

Figure 1.5: Unit Price of Land Collateralization Transactions

This figure shows the unit price of land collateralization transactions by firms that are not LGFV(NONLGFV, the solid lines) and LGFVs(LGFV, the dashed lines) from the year 2008 to 2012, with different colors representing different types of land, including industrial land(IND), residential land(RSD), commercial land(COM), and all other types of land(Other).



marginal revenue created by additional supply is just equal to the loss in marginal revenue created by the reduction in prices. If $V > 0$, we will have $V/P = 1 - \epsilon > 0$ and thus $\epsilon < 1$. Hence,the local government will sell a lower quantity of land at a higher price than it would without the outside option of collateralization.

Note that it is always optimal for the government to use collateralization, even though $V < P$ whenever $\epsilon < 1$. To see why, assume that we start with $V = 0$. Then $\epsilon = 1$ and let $P_{\epsilon=1}$ and $Q_{\epsilon=1}$ denote the optimum price and quantity of land supplied in this case, respectively.

Now consider an increase in the collateral value by $\Delta V > 0$. The additional revenue the government can generate by collateralizing a quantity ΔQ instead of selling it in the market is $\Delta V \times \Delta Q + \Delta P \times (Q_{\epsilon=1} - \Delta Q)$, whereas the reduction in revenue from reducing sales to the market is $\Delta Q \times P_{\epsilon=1}$. Since $\epsilon = 1$, it must be the case that $\Delta Q \times P_{\epsilon=1} = \Delta P \times Q_{\epsilon=1}$, so that the marginal revenue is $\Delta V \Delta Q - \Delta P \times \Delta Q$, which will be positive, as long as $\Delta P < \Delta V$.

In particular, we also get that

$$\frac{dP}{dV} = \frac{1}{1-\epsilon} > 1$$

and the elasticity of prices with respect to exogenous changes of valuation is unity.

$$\frac{dP}{dV} \times \frac{V}{P} = 1$$

In this simple model, exogenous movements in the value of collateral will act like supply shocks to the land market, shifting land prices in direct proportion to fluctuations in the value of collateral. In our empirical analysis, we will use the China's 4 trillion fiscal stimulus as the source of exogenous variation in land supply. The country-wide expansion in credit gave politically connected local governments an opportunity to raise more revenue by restricting supply to the local market, driving up land prices in the process.

Elastic land supply Our conclusions are unaffected by the assumption of a fixed land supply. To see this, suppose that the government can adjust the aggregate land supply $\overline{Q}$ at expense $E(\overline{Q})$ with $E'\left(\overline{Q}\right) > 0$..

$\overline{Q} = S + C$ is the sum of land sold and land collateralized. Then the government raises revenue

$$R(\overline{Q}) = P(S) \times S + CV - E(\overline{Q})$$

where S is the quantity of land sold and C the area that gets collateralized, so that $\overline{Q} = S + C$. The first order condition for land sales is

$$\frac{dP}{dS} \times S + P(S) - \frac{dE}{dS} = 0$$

and for collateralized land

$$V - \frac{dE}{dC} = 0$$

The marginal cost of providing land is the same irrespective of whether the land is sold or collateralized, i.e.

$$\frac{dE}{dC} = \frac{dE}{dS}$$

and we therefore have the same first-order condition as before

$$\frac{dP}{dS} S + P(S) - V = 0$$

and therefore again

$$1 - \epsilon = \frac{V}{P}$$

Hence, while the absolute amount of land supplied will clearly depend on the cost of providing this land, the relative supply of land to the private market and for collateralization is determined by the same considerations as before. This, in turn implies that an exogenous increase in the value of collateral will continue to lead to a proportional increase in land prices.

Endogenous valuation So far we have assumed that the collateral value is independent of the quantity collateralized. Now we relax this assumption, so that $V = V(1 - Q)$ and revenue becomes

$$R(Q) = P(Q) \times Q + (1 - Q)V(1 - Q)$$

Then the FOC is

$$\frac{dP}{dQ} Q + P(Q) - V + \frac{dV}{dQ}(1 - Q) = 0$$

or

$$\left(\frac{dP}{dQ} \times \frac{Q}{P} + 1\right) P(Q) = V\left(1 - \frac{dV}{dQ}\frac{(1 - Q)}{V}\right)$$

Again defining

$$\epsilon = -\frac{dP}{dQ} \times \frac{Q}{P} \text{ and using } C = 1 - Q\text{,so that } \frac{dV}{dQ}\frac{(1 - Q)}{V} = -\frac{dV}{dC}\frac{C}{V} = \eta$$

we get

$$\frac{V}{P} = \frac{1 - \epsilon}{1 - \eta}$$

Towards an instrument Let us define government revenue per unit of land sold (or collateralized)

$$\frac{R(\overline{Q})}{\overline{Q}} = P(Q) \times \frac{Q}{\overline{Q}} + \frac{(\overline{Q} - Q)}{\overline{Q}} V = P(Q) \times (1 - \omega) + \omega V$$

where ω is is the share of collateralized land. Then let us define *excess* government revenue, $ER(\overline{Q})$ as the ratio of actual revenue divided by the revenue that would be attained if the government was to sell $\overline{Q}$ in the private market. Then we have

$$ER(\overline{Q}) = \frac{R(\overline{Q})}{P\overline{Q}} = 1 + \omega \frac{V - P}{P}$$

1.4 Data Description

In this section, we first introduce our city-level data from various sources and the definition of our key variable, political connection. Then we explain how to identify LGFV companies. Finally, we describe our land transaction dataset for land sale and land collateralization, and the spatial matching approach of our sample. We end this section with a summary of the above mentioned data.

1.4.1 Data Source and Variable Construction

Our city-level data is from three sources. Urban Statistical Yearbook of China provides macro variables for each city such as GDP, aggregate income, population, house prices and so forth. Aggregate loan growth to various sectors for each city is obtained from China Banking Regulatory Commission (CBRC), an agency to regulate and monitor China's banking sector. Political connection represents the relatedness of each city with the central government in Beijing and it should not be correlated with the city's economic performance. So we define political connection (PC) in one city as the total number of national, ministry, or other leaders[6] in central government that were born or native from that city. For that purpose, we hand collected all personal information for officials working in the central government from the official website of the State Council [7] and then identify their native cities, born cities, and finally aggregate the number of officials into the PC measure[8].

China Banking Regulatory Commission(CBRC) provides lists for companies qualified as LGFVs and the date when companies exit the list. We use the lists from year 2011 to year 2014. And we get another name list from Wind Information for LGFVs that have issued "Chengtou Bond" before year 2014. We combine the name lists from the two sources and identify the LGFV names without considering the exact bond issue date and date of the list, since LGFVs typically exist for extended period of time without changing its business nature(Gao, Ru, and Tang (2020)). In this way, we identify 2885 (778) local government financing vehicles from year 2008 to 2012, for which names, locations, can be matched with our land sale (collateral) transaction data set[9].

[6] Such as the chairman of the Chinese People's Political Consultative Conference (CPPCC).

[7] http://www.gov.cn.

[8] If official O_1 is native from city A, born in city B, O_1 is counted twice, once for city A, once for city B; if official O_2 is native from and born in city A, O_2 is only counted once, for city A.

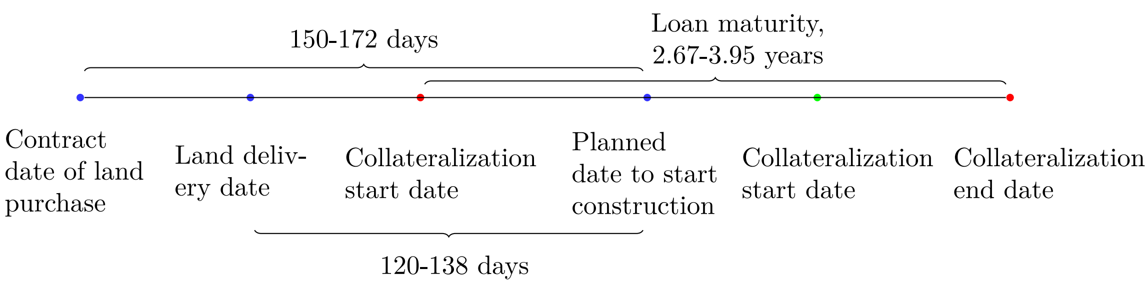
[9] 3856 deals for land collateralization; 14794 deals for land sale.

The Ministry of Land and Resources makes all the land sale, transfer, collateralization transactions publicly available on their website [10]. We manually collected all the transactions related to sale and collateralization. The land sale dataset contains information about names of seller, names of buyer, buyer's industry, contract date, address, land area, total payment, land quality, stipulated land usage, selling method and others[11]. The land collateralization dataset contains information about borrower, lender, start time, end time, address, land area, stipulated land usage and others.

The key challenge is to identify the spread between the purchase price and collateral valuation of the same land parcel, $Spread = log(1+sale\ price) - log(1+valuation)$. Since the dataset doesn't provide an identification number for each land parcel, similar to the literature, we apply a spatial matching approach. To be specific, for each land parcel sold, we generate a sample of collateralized land parcels pledged by the same entity, within a 500-meter radius, 1,000-meter radius and 1,500-meter radius from the center of the land parcel[12]. In addition, to ensure the land quality does not change, we also require the collateralization start date to be less than half a year after the delivery date of the land sale transaction. The choice of half of a year is due to the fact that, on average, it takes 150-172 days for the company to start construction after the delivery date of the land purchase[13]. Since the land quality when the land parcel is pledged should be the same as that when the land parcel is sold, the collateralization start date should be before the construction starting point[14]. Moreover, most data is missing for the construction date variable. Thus, we simply choose half a year as a criterion. In the robustness test, we also use one year, one and half a year to test our model. For a graphic look of the matching process, refer to Figure 1.7. Altogether we identify 5,719 land parcels within the 1,500-meter radius, 5,143 land parcels within the 1,000-meter radius and 4,485 land parcels within the 500-meter radius. Figure 1.8 shows the geographic distribution of the sample. Among the sample, 328 land parcels within the 1,500-meter radius, 282 land parcels within the 1,000-meter radius and 252 land parcels within the 500-meter radius are purchased and pledged by LGFVs.

Figure 1.6: Dates of Land Transactions

This figure shows all the important dates and time intervals for the land sale and land collateralization transactions.



[10] https://www.landchina.com.

[11] Land quality is a natural number scaled from 1-20 evaluated by the official-in-charge, the lower the number, the higher the quality of the land parcel. Stipulated land usage is the usage regulated by the government, including industrial land, residential land, commercial land and other land. Selling method includes allocating, negotiating, auction and leasing.

[12] For each land parcel p being sold, if more than one land parcels being collateralized are matched, *valuation* refers to the average collateral valuation.

[13] See more details in Figure 1.6.

[14] The collateralization start date should be the red point, not the green point in Figure 1.6.

Figure 1.7: Matching Process

This figure shows our spatial matching method. For each land parcel sold, we generate a sample of collateralized land parcels pledged by the same entity, within a 500-meter radius, 1,000-meter radius and 1,500-meter radius from the center of the land parcel sold, and the collateralization start date should be less than half a year after the delivery date of the land sale transaction.

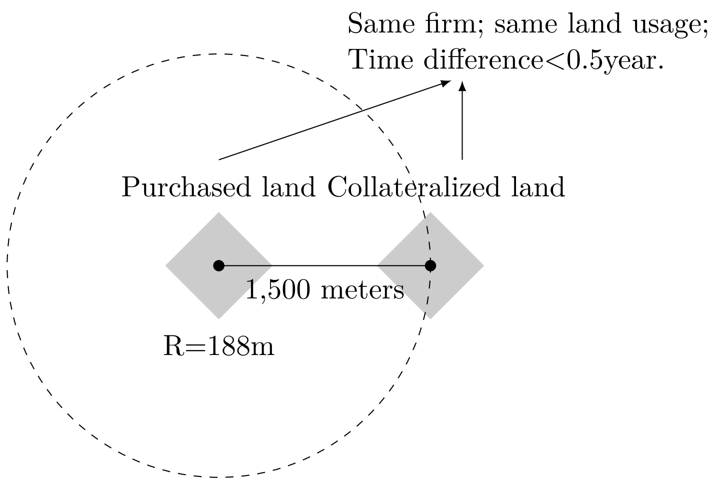
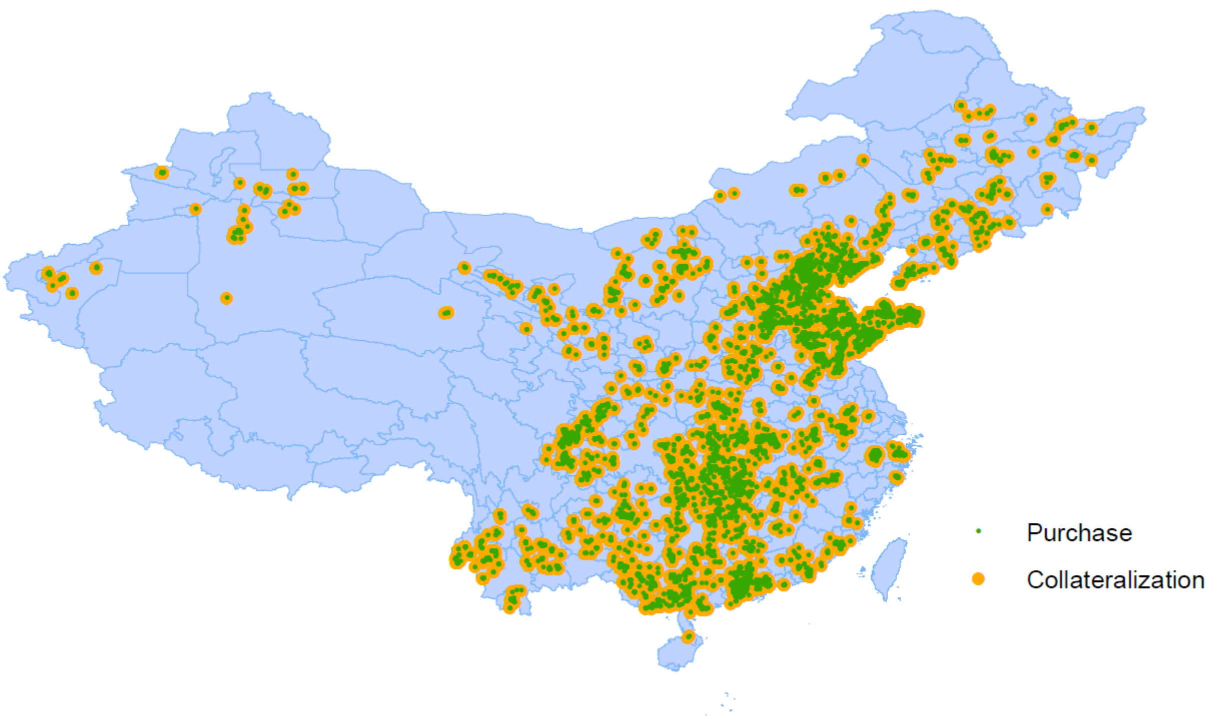


Figure 1.8: Geographic Distribution of the Matched Sample

This figure presents the geographic distribution of the matched sample. The green dots represent the land parcels sold by the local government. The yellow dots represent the land parcels collateralized to the bank.



1.4.2 Summary Statistics

Our sample period is from year 2008 to year 2012. We start with year 2008 is due to the fact that the stimulus plan went into effect at the end of year 2008, and from 2008, a new cohort of central government officials is in power until year 2012. Therefore, to make reliable implications from our difference in difference regression in Section 1.5, we need sample periods before 2009. To mitigate the endogeneity problem, we fix the political connection measure at the start of the sample [15]. Therefore, if we choose year 2007 as the start of our sample, political connection in 2007 can not reflect the political connection from 2008, since the officials in central government changed from 2008. The sample ends in year 2012 for two reasons: the stimulus plan finished; the central government officials finished their 5-year term at the end of year 2012 if they start in year 2008.

Table 1.1 summarizes the city-level macro variables. The mean (median) growth rate of all SOE loans is 15.7% (8.9%) during the sample period for 1,233 city-year observations, where all SOE firms include LGFVs and other SOEs. The growth rate of loans to LGFV firms, with mean(median) of 19.5% (9.4%), is higher than growth rate of loans to other SOEs, with mean(median) of 13.5% (8.56%). The growth rate of private loans is the largest, 21.2% on average. Although more than half of the cities in the sample has no political connection, the average political measure is 0.424, with the maximal political connection of 4, which means 4 officials from the central government are native or born from the same city. During the sample period, a representative city experience 15% growth of GDP per capita, and 9.88% growth of income per capita, and 0.78% increase in population.

Table 1.1: Summary Statistics of Macro Variables in City Level

This table shows the summary statistics of macro variables for each year during the sample period 2008-2012. The first 4 variables measure the log lending growth to all SOEs, LGFVs, other SOEs and private sectors. PC stands for political connection. GDP denotes GDP growth. Income denotes income growth. Population denotes population growth.

	(1) N	(2) mean	(3) sd	(4) min	(5) p25	(6) p50	(7) p75	(8) max
Growth All SOE Loans	1,233	0.157	0.290	-0.681	-0.0159	0.0890	0.275	1.758
Growth LGFV Loans	1,204	0.195	0.385	-0.736	-0.0446	0.0940	0.349	2.353
Growth SOE Loans	1,220	0.135	0.286	-0.739	-0.0305	0.0856	0.250	1.908
Growth Private Loans	1,242	0.212	0.213	-0.691	0.105	0.196	0.290	2.532
PC	1,428	0.424	0.824	0	0	0	1	4
GDP	1,411	0.150	0.0723	-0.191	0.108	0.151	0.195	0.516
Income	1,407	0.0988	0.0643	-0.392	0.0689	0.0976	0.129	0.577
Population	1,421	0.00784	0.0290	-0.265	0.00188	0.00609	0.0112	0.571

Table 1.2 summarizes the variables related to land transactions where the pledged land parcels are within 1,500-meter radius of the land sale parcels. The sample contains 5,719 matched land parcels. The average spread is 0.347, which represents a 41.5% appreciation from sale price to collateralization valuation. 5.75% of the entities in the sample are LGFV firms. The average political connection for the matched sample is 0.749. Since we require the land collateralization start date is less than half a year after the delivery date of land sales, the average time difference between the two dates is 0.242 year, which is equivalent to 88 days. The maturity of land collateralization is about 3 years. The average area of land parcels in the matched sample is 1.66 hectare, and the floor-area-ratio of a representative land parcel is 1.48[16].

[15] See more details in Section 1.5.

[16] 1.66 = exp(0.979)-1; 1.48 = exp(0.908) - 1.

Table 1.2: Summary Statistics for Land Transactions within 1,500-Meter Radius

This table shows the summary statistics of matched land transactions when the collaterialized land parcels is within 1,500 meters from the center of land parcels purchased by the same entity. Spread = log(1 + collateral valuation) - log(1 + sale price); LGFV is a dummy variable equal to 1 if the buyer for land sale transactions or borrower for land collateralization transactions is LGFV; PC denotes political connection; Timediff denotes the date difference between the contract date of land sale transactions and collateralization start date of land collateralization transactions; Maturity measures the maturity of the land collateralization transactions; Logarea is log(1 + area) where area is the collateralization area; logFAR = log(1 + floor-area-ratio).

	(1) N	(2) mean	(3) sd	(4) min	(5) p25	(6) p50	(7) p75	(8) max
Spread	5,719	0.347	1.187	-3.138	-0.344	0.137	0.792	5.354
LGFV	5,719	0.0574	0.233	0	0	0	0	1
PC	5,719	0.749	1.050	0	0	0	1	4
Timediff	5,719	0.242	0.140	0	0.127	0.235	0.359	0.500
Maturity	5,719	2.979	2.461	0.00270	0.986	1.976	2.962	10
Logarea	5,719	0.979	0.869	0.000300	0.251	0.821	1.486	6.562
LogFAR	5,719	0.908	0.413	0	0.693	0.788	1.224	1.792

1.5 Empirical Methodology and Results

In this section, we explain our empirical methods and hypothesis step by step. We begin with the hypothesis that following the fiscal stimulus, political connection with the central government does promote the lending growth, specifically lending growth of LGFV firms. Then we try to identify the mechanism behind, that is, political connections affect the loan growth through the land transaction sale price and collateral valuation of LGFV firms. At the end, we aggregate the spread between the land purchase price and collateral valuation to explain the house price boom. After each regression model, we report the corresponding empirical results.

1.5.1 Lending Growth

In China, promotion of local officials mainly depends on local economic performance. The lending boom since 2009 provided an opportunity for local officials to quickly expand their local economies. Local political connection to the central government, has been shown in many other research to be an unignorable channel to impact the local economies through credit supply[17]. While local governments are not allowed to directly borrow in financial markets, they establish local government financing vehicles (LGFV) to borrow on behalf of the local government. Thus we expect that cities with political connection to experience higher lending growth following the fiscal stimulus comparing to cities without political connection. And a higher proportion of this lending growth is going through the LGFVs.

To test the causality between political connection and loan growth, we employ the following regression model based on a double Diff-in-Diff (DDD) strategy:

$$Outcome_t^c = \alpha \times PC_{2008}^c \times LG_t + \beta' \mathbf{X_t^c} + city + province \times year + \epsilon_t^c \tag{1.1}$$

where the dependent variable $Outcome_t^c$ is the lending growth for city c in year t to different sectors: state-owned-enterprises (SOE), private firms. Besides the lending growth to all SOEs for city c, we also exploit the effect for a subsample of lending growth to LGFVs in city c, local SOEs in city c. The key explanatory variable is $PC_{2008}^c \times LG_t$, the product of political connection in the year 2008 for city c and country-wide loan growth in year t.

[17]See Ru (2018), Huang, Pagano, and Panizza (2016).

As year 2008 is the beginning of our sample, and to rule out any correlation between the political connection and city-level outcomes, we fix the political connection in the year 2008. Therefore, our assumption is that the national-wide fiscal stimulus affects municipal lending growth through the political connection of municipal government with the central government. In all city-level specifications, we include a set of city-time varying control variables, $\boldsymbol{X_t^c}$, including GDP growth, income growth, population growth. In addition, to get rid of any common factors among cities, or for the same province in the same year, we include the *city* fixed effect and *province* $\times$ *year* fixed effect. The standard errors are clustered at province and year level.

The results are reported in Table 1.3. Column (1), where the dependent variable is the lending growth to all SOEs, including LGFVs and local SOEs, shows that the coefficient for $PC_{2008}^c \times LG_t$ is 0.280, statistically significant at 5% level. As for the economic significance, one unit increase of political connection leads to 0.18 standard deviation increase of lending growth to all SOEs[18]. Column (2)-(4) shows that the lending growth concentrates on the LGFVs. Specifically, column (2), where the dependent variable is the lending growth to LGFVs, shows that the coefficient for $PC_{2008}^c \times LG_t$ is 0.270, statistically significant at 10% level. As for the economic significance, one unit increase of political connection leads to 0.13 standard deviation increase of lending growth to LGFVs during the fiscal stimulus[19]. In the contrast, statistically, $PC_{2008}^c \times LG_t$ does not affect lending to local SOEs in column (3), or firms in the private sectors in column (4).

Table 1.3: Loan Growth

This table shows the regression results of $Outcome_t^c = \alpha \times PC_{2008}^c \times LG_t + \beta' \mathbf{X_t^c} + city + province \times year$. The dependent variable $Outcome_t^c$ is the lending growth for city c in year t to all SOEs, LGFVs, other SOEs and private sectors. $PC_{2008}^c \times LG_t$ is the product of political connection in the year 2008 for city c and country-wide loan growth of the Big 4 Banks in year t. $\boldsymbol{X_t^c}$, includes GDP growth, income growth, population growth. *city* fixed effect and *province* $\times$ *year* fixed effect are included. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Lending to ...	(1) All SOE	(2) LGFV	(3) Other SOE	(4) Private
$PC_{2008}^c \times LG_t$	0.280**	0.270*	0.00395	-0.0330
	(0.0838)	(0.113)	(0.0997)	(0.0504)
GDP per cap	0.132	0.255	0.0634	0.0821
	(0.137)	(0.151)	(0.0762)	(0.102)
Income	-0.0332	-0.0730	-0.107*	-0.114
	(0.0862)	(0.164)	(0.0402)	(0.0859)
Population	0.0523	0.193	0.0727	0.488*
	(0.121)	(0.184)	(0.105)	(0.179)
Observations	1,233	1,204	1,220	1,242
R-squared	0.554	0.654	0.499	0.386
City FE	YES	YES	YES	YES
Prov*Year FE	YES	YES	YES	YES

To rule out any other confounding factors that may bias the estimation of $PC_{2008}^c \times LG_t$ on lending growth, we did a set of robustness tests. First, to ensure that the political connection affects local lending growth only through the national-wide lending growth, not through other time-varying national trends, we add $PC_{2008}^c * \boldsymbol{\delta_t}$ as additional controls, where $\boldsymbol{\delta_t}$ is a vector of aggregate time varying variables, including current account/GDP,

[18]0.280*0.186/0.290, where the average LG_t is 0.186, the standard deviation of lending growth to all SOEs is 0.290.

[19]0.270*0.186/0.385, where the average LG_t is 0.186, the standard deviation of lending growth to all SOEs is 0.385.

real GDP growth, real interest rate. Second, other city-characteristics could modulate the impact of the national lending boom on local lending growth, so we control for $LG_t \times \boldsymbol{\gamma^c}$, where $\boldsymbol{\gamma^c}$ is a vector of time-invariant city-level characteristics, including city type and administrative area. In the third and fourth setting, we additionally control for $\mathbf{X^c_t} \times PC^c_{2008}$ and $\mathbf{X^c_t} \times LG_t$, where $\mathbf{X^c_t}$ has the same definition as the control variables in equation (1.1), taking into account that political connection may affect local lending through city-time varying variables, or that the mechanism from national lending to local lending could be from other city-time varying channels. At last, we put all the above additional controls in one regression and also include $\boldsymbol{\delta_t} \times \boldsymbol{\gamma^c}$ as controls.

The results are reported in Table 1.4. To save space and make comparison easier, we only report the coefficients of $PC^c_{2008} \times LG_t$. As demonstrated in column (1), the effects of $PC^c_{2008} \times LG_t$ on lending growth of all SOEs stay statistically significant, and its economic significance is at least as large as the base case. When we consider the lending growth of LGFV firms, all coefficients of $PC^c_{2008} \times LG_t$ are significant at least on the 10% level and the economic significance is as large as that in the base case. In column (3) and (4), we find consistent result that $PC^c_{2008} \times LG_t$ has no effect on lending growth of local SOEs or firms in the private sectors.

Table 1.4: Robustness Test on Table 1.3

This table shows the robustness regression results of $Outcome^c_t = \alpha \times PC^c_{2008} \times LG_t + \beta' \mathbf{X^c_t} + city + province \times year$. The dependent variable $Outcome^c_t$ is the lending growth for city c in year t to all SOEs, LGFVs, other SOEs and private sectors. $PC^c_{2008} \times LG_t$ is the product of political connection in the year 2008 for city c and country-wide loan growth of the Big 4 Banks in year t. In row **B**, $\boldsymbol{X^c_t}$, includes GDP growth, income growth, population growth; in row **1**, controls additionally include $PC^c_{2008} * \boldsymbol{\delta_t}$, where $\boldsymbol{\delta_t}$ is a vector of aggregate time varying variables, including current account/GDP, real GDP growth, real interest rate; in row **2**, controls additionally include $LG_t \times \boldsymbol{\gamma^c}$, where $\boldsymbol{\gamma^c}$ is a vector of time-invariant city-level characteristics, including city type and administrative area; in row **3**, controls additionally include $\mathbf{X^c_t} \times PC^c_{2008}$; in row **4**, controls additionally include $\mathbf{X^c_t} \times LG_t$, where $\mathbf{X^c_t}$ has the same definition as the control variables in equation (1.1); in row **5**, controls include all the above additional controls in one regression and also include $\boldsymbol{\delta_t} \times \boldsymbol{\gamma^c}$. *city* fixed effect and *province* $\times$ *year* fixed effect are included. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Lending to...	(1) All SOEs	(2) LGFV	(3) Other SOEs	(4) Private
B.$PC^c_{2008} \times LG_t$	0.280**	0.270*	0.00395	-0.0330
	(0.0838)	(0.113)	(0.0997)	(0.0504)
1. $PC^c_{2008} \times LG_t$	0.527***	0.515**	0.0146	0.00153
	(0.0318)	(0.151)	(0.153)	(0.0489)
2. $PC^c_{2008} \times LG_t$	0.285**	0.352**	0.0296	0.0462
	(0.0852)	(0.119)	(0.104)	(0.0529)
3. $PC^c_{2008} \times LG_t$	0.571***	0.554*	0.0720	0.0455
	(0.0720)	(0.224)	(0.147)	(0.0530)
4. $PC^c_{2008} \times LG_t$	0.282**	0.355*	0.0263	0.0460
	(0.0880)	(0.129)	(0.105)	(0.0739)
5. $PC^c_{2008} \times LG_t$	0.500**	0.652*	0.109	0.194
	(0.150)	(0.265)	(0.210)	(0.130)

In summary, in this section, we find consistent result to show that following the fiscal stimulus plan of China in 2009, political connections with the central government lead to higher growth of lending to SOEs and specifically, LGFVs, but not to local SOEs or private firms.

1.5.2 Transaction-Level Results

The above section provides evidence that fiscal stimulus affects the local lending growth through political connection to the central government, and the main effect is on LGFVs. In this part, we exploit the mechanism behind with our land sale and land collateralization transaction-level dataset.

In China's system of local government finance, funds in off-balance sheet LGFVs are much more flexible in their use than funds allocated through the official budgeting process that is under relatively close control of the central government. Hence, funds in LGFVs can more easily be used for the pursuit of the private career objectives of local officials. Land is an important source of capital to establish LGFVs since local government is the monopolistic supplier of land usufruct rights. Thus, officials in local government have a strong incentive not only to raise funds for their LGFVs from the stimulus, but also to transfer land use rights from the local government to the LGFVs at possibly low prices. As estimated by Chen, He, and Liu (2020), 90% of the increase in local government debt during the stimulus period comes from bank loans. Although local government provides implicit guarantee for LGFV borrowings, after the strengthening of the management on LGFVs on June 10, 2010, LGFVs have to pledge land usufruct rights or other assets as collateral to borrow from banks[20]. Therefore, we expect the price of land sale transactions for LGFVs is lower during the stimulus period compared to non-LGFVs, especially for cities with political connection, and the valuation of land collateral transactions is higher during the the stimulus period compared to non-LGFVs, especially for cities with political connection.

To estimate the higher spread, price discount and overvaluation obtained by the LGFVs, we employ the following regression models:

$$Outcome^d = \alpha \times LGFV^d \times PC_{2008}^{c(d)} \times LG_t + Ctrls + FEs + \epsilon^d \tag{1.2}$$

where the dependent variable $Outcome^d$ is the transaction level outcome, including $Spread$ between collateral valuation and purchase price, purchase price, collateral valuation for land transaction d[21]. The key independent variable is the triple interaction term $LGFV^d \times PC_{2008}^{c(d)} \times LG_t$. $LGFV^d$ is a dummy variable, which equals to 1 if the buyer(borrower) of land transaction d is LGFV, 0 otherwise. $PC_{2008}^{c(d)}$ is the city-level political connection in 2008 for land transaction d in city c. LG_t represents the country-wide loan growth in year t when the collateralization happens. The lower order interaction terms $LGFV^d \times PC_{2008}^{c(d)}$, $LGFV^d \times LG_t$, $LGFV^d$ are included, while $PC_{2008}^{c(d)} \times LG_t$, $PC_{2008}^{c(d)}$, LG_t are absorbed by the fixed effects. We also control for a number of transaction-level control variables, including the time difference between purchase date and collateralization start date, loan maturity, log of area, floor-area-ratio, land quality(level). Additionally, to ensure that we are actually comparing the land parcels located in the same city of the same year, with same stipulated usage and supplied by the government with the same method, we include a high-dimensional control of fixed effects, including city-year fixed effect, usage fixed effect, supply method fixed effect. The standard errors are clustered at province and year level.

The results are reported in Table 1.5, 1.6, 1.7 for spread, purchase price, collateral valuation respectively. In column (1) of Table 1.5, the coefficient of $LGFV^d \times PC_{2008}^{c(d)} \times LG_t$ is 2.859, significant at 1% level for the sample where the purchased land parcels and collateralized land parcels are within 500 meters. Regarding to the economic significance, one LGFV firm located in the city with 1 unit of political connection will benefit 21.44% increase of spread compared to non-LGFV firms in the same city in 2009[22]. The coefficients get bigger when we allow the distance of the purchased land and collateralized land to be farther. The coefficient is 3.526 if the distance of the purchased land parcels

[20] More details in Section 1.2.

[21] All the variables are logarized.

[22] 21.44% = exp(2.859 * 0.369 - 0.326 - 1.449*0.369) - 1, the loan growth rate in 2009 is 0.369.

and collateralized land parcels is less than 1,000 meters, and 3.621 if the distance of the purchased land and collateralized land is less than 1,500 meters. The coefficients for $LGFV^d \times LG_t$ are negative, reflecting the fact that compared to LGFVs in politically connected cities, LGFVs located in cities with no political connection experience lower spread. Similarly, the negative coefficients for $LGFV^d \times PC^{c(d)}_{2008}$ show that in low national loan growth period, LGFVs located in cities with political connection experience lower spread. Consistent with findings in prior literature, land area has a significantly negative effects on the spread, reflecting a marginally decreasing phenomenon.

Table 1.5: Land Transactions: Spread

This table shows the regression results of $Outcome^d = \alpha \times LGFV^d \times PC^{c(d)}_{2008} \times LG_t + Ctrls + FEs + \epsilon^d$ where the dependent variable $Outcome^d_t$ is the spread, defined as log difference between collateral valuation and purchase price per square meter for transaction d. $LGFV^d$ is a dummy variable, which equals to 1 if the buyer(borrower) of transaction d is LGFV, 0 otherwise. PC^c_{2008} is the city-level political connection in 2008 for transaction d which is located in city c. LG_t represents the country-wide loan growth of the Big 4 Banks in year t when the collateralization happens. The lower order interaction terms $LGFV^d \times PC^c_{2008}$, $LGFV^d \times LG_t$, as well as each variable in the interaction term $LGFV^d$ are included, while $PC^c_{2008} \times LG_t$, PC^c_{2008}, LG_t are absorbed by the fixed effects. Control variables include the time difference between purchase and collateralization, loan maturity, log of area, floor-area-ratio and land quality(level). Fixed effects include city-year fixed effect, usage fixed effect, supply method fixed effect. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	Distance, *meters*		
Spread	500	1000	1500
$LGFV^d$	-0.0187	0.279	0.164
	(0.330)	(0.382)	(0.345)
$LGFV^d \times PC^c_{2008} \times LG_t$	2.859***	3.526***	3.621***
	(0.162)	(0.269)	(0.669)
$LGFV^d \times PC^c_{2008}$	-0.326***	-0.527**	-0.495**
	(0.0601)	(0.175)	(0.176)
$LGFV^d \times LG_t$	-1.449***	-2.322***	-2.096**
	(0.107)	(0.307)	(0.505)
Timediff	-0.0507	-0.0791	0.000546
	(0.0929)	(0.129)	(0.0751)
Maturity	0.0501***	0.0519***	0.0493***
	(0.00960)	(0.0102)	(0.00971)
Logarea	-0.142***	-0.153**	-0.153***
	(0.0226)	(0.0337)	(0.0258)
logFAR	-0.199*	-0.233*	-0.303**
	(0.0927)	(0.0885)	(0.0966)
Level	0.0121	0.0110	0.0127
	(0.0102)	(0.0133)	(0.0123)
Observations	4,485	5,143	5,719
R-squared	0.547	0.528	0.510
Controls	YES	YES	YES
CityYear FE	YES	YES	YES
Other FEs	YES	YES	YES

Table 1.6, 1.7 show that the higher spread for LGFV firms located in cities with political connection during high loan growth period is from the lower purchase price when they buy the land, and higher collateral valuation when they pledge the land to banks. Specifically, in the sample where the distance between the purchased land and collateralized land is less than 500 meters, the coefficient of $LGFV^d \times PC^{c(d)}_{2008} \times LG_t$ is 1.577 for collateral valuation, which is significant at 1% level, which can be translated to 48.25% of higher valuation economically in 2009 for LGFV firms located in cities with one unit of political connection compared to non-LGFV firms in the same city[23]. The coefficient of $LGFV^d \times PC^{c(d)}_{2008} \times LG_t$ is -1.233 for purchase price, which is marginally

[23] 48.25% = exp(0.423 + 1.577*0.369 - 0.189 - 1.144*0.369) - 1, the loan growth rate in 2009 is 0.369.

significant but when we allow the distance to be further, the coefficients become significant at 1% level.

Table 1.6: Land Transactions: Sale Price

This table shows the regression results of $Outcome^d = \alpha \times LGFV^d \times PC^{c(d)}_{2008} \times LG_t + Ctrls + FEs + \epsilon^d$ where the dependent variable $Outcome^d_t$ is the log purchase price per square meter for transaction d. $LGFV^d$ is a dummy variable, which equals to 1 if the buyer(borrower) of transaction d is LGFV, 0 otherwise. PC^c_{2008} is the city-level political connection in 2008 for transaction d which is located in city c. LG_t represents the country-wide loan growth of the Big 4 Banks in year t when the collateralization happens. The lower order interaction terms $LGFV^d \times PC^c_{2008}$, $LGFV^d \times LG_t$, as well as each variable in the interaction term $LGFV^d$ are included, while $PC^c_{2008} \times LG_t$, PC^c_{2008}, LG_t are absorbed by the fixed effects. Control variables include the log of area, floor-area-ratio, land quality(level). Fixed effects include city-year fixed effects, usage fixed effects, supply method fixed effects. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	Distance, *meters*		
Purchase Price	500	1000	1500
$LGFV^d$	0.377	0.0718	0.0231
	(0.264)	(0.470)	(0.311)
$LGFV^d \times PC^c_{2008} \times LG_t$	-1.233	-1.491***	-1.808***
	(0.661)	(0.297)	(0.228)
$LGFV^d \times PC^c_{2008}$	0.270	0.289	0.287
	(0.199)	(0.201)	(0.163)
$LGFV^d \times LG_t$	0.558	1.797	2.107**
	(0.915)	(1.235)	(0.669)
Logarea	-0.0832	-0.0836	-0.0667
	(0.0489)	(0.0744)	(0.0800)
LogFAR	0.656***	0.624***	0.602***
	(0.118)	(0.130)	(0.0870)
Level	-0.0282**	-0.0339**	-0.0327**
	(0.00929)	(0.00953)	(0.00725)
Observations	3,716	4,315	4,872
R-squared	0.709	0.683	0.690
Controls	YES	YES	YES
CityYear FE	YES	YES	YES
Other FEs	YES	YES	YES

1.5.3 House Prices

As shown in Section 1.3, when local government has options to collateralize land, it is always optimal to let LGFV to collateralize land parcels instead of selling the land parcels to private sector in the primary market. Moreover, local government has incentives to develop industrial land to boost the local economy. Thus local government posts more residential land as collateral at inflated valuations to borrow from banks. This led to a scarcity of residential land, leading to a boom of house prices. Therefore, our hypothesis is that cities with more political connections during the fiscal stimulus period post more residential land parcels as collateral, which decreases the supply of residential land to the market, and as a result, house price rises. This is a supply side story. However, several policies have been implemented with the expectation to increase the house demand by the Chinese government at the same time as the stimulus plan. For example, in October 2008, the minimum mortgage interest ratio dropped from 85% to 70% of the benchmark lending interest rate; the down payment ratio declined from 30% to 20% for the first home purchase; the loan interest rate of housing provident fund decreased by 0.27%[24].

[24]See http://www.gov.cn/gzdt/2008-10/22/content_1127938.htm, http://www.gov.cn/ztzl/2009-01/04/content_1195355.htm.

Table 1.7: Land Transactions: Collateral Valuation

This table shows the regression results of $Outcome^d = \alpha \times LGFV^d \times PC_{2008}^{c(d)} \times LG_t + Ctrls + FEs + \epsilon^d$ where the dependent variable $Outcome_t^d$ is the log of collateral valuation per square meter for transaction d. $LGFV^d$ is a dummy variable, which equals to 1 if the buyer(borrower) of transaction d is LGFV, 0 otherwise. PC_{2008}^c is the city-level political connection in 2008 for transaction d which is located in city c. LG_t represents the country-wide loan growth of the Big 4 Banks in year t when the collateralization happens. The lower order interaction terms $LGFV^d \times PC_{2008}^c$, $LGFV^d \times LG_t$, as well as each variable in the interaction term $LGFV^d$ are included, while $PC_{2008}^c \times LG_t$, PC_{2008}^c, LG_t are absorbed by the fixed effects. Control variables include the loan maturity, log of area, floor-area-ratio. Fixed effects include city-year fixed effects and usage fixed effects. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	Distance, *meters*		
Collateral Valuation	500	1000	1500
$LGFV^d$	0.423*	0.498*	0.303
	(0.163)	(0.231)	(0.233)
$LGFV^d \times PC_{2008}^c \times LG_t$	1.577***	1.619***	1.007**
	(0.0558)	(0.225)	(0.305)
$LGFV^d \times PC_{2008}^c$	-0.189**	-0.248**	-0.148
	(0.0640)	(0.0775)	(0.0819)
$LGFV^d \times LG_t$	-1.144*	-1.207	-0.220
	(0.525)	(0.656)	(0.838)
Maturity	0.0295	0.0298	0.0311*
	(0.0295)	(0.0167)	(0.0135)
Logarea	-0.217**	-0.210*	-0.214**
	(0.0677)	(0.0804)	(0.0590)
LogFAR	0.526**	0.475**	0.463**
	(0.142)	(0.140)	(0.121)
Level	-0.0332*	-0.0313**	-0.0303*
	(0.0146)	(0.0112)	(0.0133)
Observations	4,051	4,620	5,088
R-squared	0.684	0.669	0.654
Controls	YES	YES	YES
CityYear FE	YES	YES	YES
Other FEs	YES	YES	YES

With the demand side story as a confounding factor for the rise of house prices, the main challenge to test the above supply side story empirically lies in that we need an instrument that purely measures the supply side effect, that is, the effect of $LGFV^d * PC_{2008}^c * LG_t$ on $Spread$, ruling out any potential city-year variables that are correlated to $LGFV^d * PC_{2008}^c * LG_t$ and $Spread$. We exploit a granular instrumental variable method (GIV) first proposed by Gabaix and Koijen (2020). Next we explain our implementation of the GIV method step by step.

First, we run a regression of $Spread_t^d$ on all variables in equation 1.2 except the triple interaction term $LGFV^d * PC_{2008}^c * LG_t$:

$$Spread^d = Controls + City \times Year\ FE + Other\ FEs + \epsilon^d$$

Thus, ϵ_t^d measures the net effect of $LGFV^d * PC_{2008}^c * LG_t$ on $Spread^d$, ruling out all city-year effects, or any other deal-specific characteristics, such as area, supply method, land usage, that could affect spread.

Second, to construct the city-time aggregate GIV_t^c, we take the difference of area-weighted average of ϵ^d and equal weighted average of ϵ^d, which should be a function of $PC_{2008}^c * LG_t$ by our construction:

$$GIV_t^c = size\ weighted\ \epsilon^d - equal\ weighted\ \epsilon^d$$

Thus, with the predicted value of GIV_t^c, $G\hat{I}V_t^c$ by running the following time-series regression for each city, $G\hat{I}V_{ct}$ purely captures the effect of $PC_{2008}^c \times LG_t$ on spread in city c at time t.

$$GIV_t^c = \alpha + PC_{2008}^c \times LG_t + u_t^c$$

At the end, we test our hypothesis that land collateralization by LGFV firms promotes the growth of house prices after the fiscal stimulus starting in late 2008.

$$\Delta hp_t^c = G\hat{I}V_t^c + \beta' \mathbf{X_t^c} + city + province \times year + \epsilon_t^c \tag{1.3}$$

where $\mathbf{X_t^c}$ is the same as in equation 1.1 in section 1.5.1, which includes GDP growth, income growth, and population growth.

Table 1.8: GIV Method on House Price Changes

This table shows the regression result of $\Delta hp_t^c = G\hat{I}V_{ct} + Ctrls + city + year + \epsilon_{ct}$, where Δhp_t^c is the log growth rate of house price for city c of year t; $G\hat{I}V_{ct}$ is the fitted value of GIV, more details refer to part 1.5.3 of Section 1.5; $Ctrls$ include GDP growth, income growth, population growth. Standard errors are clustered at province level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	Distance, *meters*		
Δhp_t^c	500	1000	1500
$G\hat{I}V_{ct}$	0.0561**	0.0492*	0.0378*
	(0.0214)	(0.0250)	(0.0218)
GDP	-0.0979	-0.100	-0.155
	(0.123)	(0.123)	(0.109)
Income	-0.0124	-0.0837	-0.122
	(0.103)	(0.124)	(0.119)
Population	0.181	0.217	0.400
	(0.581)	(0.568)	(0.605)
Observations	305	325	339
R-squared	0.466	0.445	0.430
City FE	YES	YES	YES
Year FE	YES	YES	YES

The results of regression 1.3 are reported in Table 1.8. For samples with different distance between purchased land parcels and collateralized land parcels of the same entity, we all get significantly positive coefficients for $G\hat{I}V_{ct}$ at least at 10% statistical level. Due to the construction of $G\hat{I}V_{ct}$, it is not easy to interpret the coefficient with economic significance. But with the significantly positive coefficients on our $G\hat{I}V_{ct}$ measure for various samples, we could conclude that after ruling out all confounding factors, the lending growth channeling through LGFVs and political connections causally leads to higher housing price through the mechanism described in section 1.5.2.

1.6 Robustness Checks

1.6.1 Political Connection Measure

The definition of political connection in our measure could be correlated to the population in the city, which could affect the loan growth in turn. Although we have controlled the population growth rate in equation 1.1, it is possible that the level of population could affect the loan growth as well. So we scale the political connection with the population in 2008($PC_{2008}^c / Pop_{2008}^c$) [25] and run the regression equation 1.1. The result is shown in Table 1.9. As shown in the table, our results for the lending to all SOE firms, LGFV firms, other SOE firms and private sectors still hold after considering the effect of population on loan growth.

[25] Where Pop_{2008}^c stands for the population of city c in 2008.

Table 1.9: Robustness on Loan Growth

This table shows the regression results of $Outcome_t^c = \alpha \times PC_{2008}^c/Pop_{2008}^c \times LG_t + \beta'\mathbf{X_t^c} + city + province \times year$. The dependent variable $Outcome_t^c$ is the lending growth for city c in year t to all SOEs, LGFVs, other SOEs and private sectors. $PC_{2008}^c/Pop_{2008}^c \times LG_t$ is the product of political connections in the year 2008 for city c and country-wide loan growth in year t scaled by the Population in 2008, and then times 100. $\boldsymbol{X_t^c}$, includes GDP growth, income growth, population growth. $city$ fixed effect and $province \times year$ fixed effect are included. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Lending to ...	(1) All SOE	(2) LGFV	(3) Other SOE	(4) Private
$PC_{2008}^c/Pop_{2008}^c \times LG_t$	1.400**	1.735**	0.00745	0.104
	(0.471)	(0.538)	(0.416)	(0.268)
GDP per cap	0.139	0.260	0.0634	0.0822
	(0.131)	(0.140)	(0.0770)	(0.103)
Income	-0.00320	-0.0344	-0.107*	-0.112
	(0.0911)	(0.152)	(0.0402)	(0.0853)
Population	0.0606	0.202	0.0728	0.488*
	(0.120)	(0.175)	(0.106)	(0.179)
Observations	1,233	1,204	1,220	1,242
R-squared	0.556	0.657	0.499	0.386
City FE	YES	YES	YES	YES
Prov*Year FE	YES	YES	YES	YES

1.6.2 Relaxing the Criteria of Land Transactions

In order to make sure that we are comparing the same land parcel when it was sold and when it was collateralized, we restrict the time difference between the delivery date of purchased land parcels and collateralization start date to be less than half a year. But this strict criterion restricts our sample size as well. We relax the time difference to be less than 1 year and 1.5 year, and run regression equation 1.2 again. The results are shown in Table 1.10. All the coefficients on $LGFV^d \times PC_{2008}^c \times LG_t$ are still significant although with lower magnitude. This makes sense because as the time difference becomes larger, it is more possible that the land quality is higher when the land parcel is collateralized than the quality when it has been sold, so that the spread between the collateral valuation and purchase price is higher. This leads to less difference on the spread between LGFV firms and non-LGFV firms.

1.7 Conclusion

This paper explores the unintended consequence of the economic stimulus plan of 2009-2010 in China. Most existing literature emphasizes the funding structure of local government financial vehicles, which was issuing bonds or borrowing from banks to finance the local government's stimulus programs. We construct a model to illustrate that for a local government that is the monopolistic supplier of land in the local market, it is always optimal for the local government to use collateralization. Then we empirically show that to achieve the lending target, local governments sell land usufruct rights to local government financial vehicles through the primary land market at lower prices compared to firms that are not local government financing vehicles, and then local government financial vehicles post the land parcels as collateral to borrow from banks at higher valuation. Moreover, we also explore the heterogeneity of local government borrowing by showing that local cities with more political connections with the central government have gone through higher loan growth.

Overall, our results show that the collateral channel of local government financing

Table 1.10: Robustness on Land Transactions: Spread

This table shows the robustness regression results of $Outcome^d = \alpha \times LGFV^d \times PC_{2008}^{c(d)} \times LG_t + Ctrls + FEs + \epsilon^d$ where the dependent variable $Outcome_t^d$ is the spread, defined as log difference between collateral valuation and purchase price per square meter for transaction d. $LGFV^d$ is a dummy variable, which equals to 1 if the buyer(borrower) of transaction d is LGFV, 0 otherwise. PC_{2008}^c is the city-level political connection in 2008 for transaction d which is located in city c. LG_t represents the country-wide loan growth of the Big 4 Banks in year t when the collateralization happens. The lower order interaction terms $LGFV^d \times PC_{2008}^c$, $LGFV^d \times LG_t$, as well as each variable in the interaction term $LGFV^d$ are included, while $PC_{2008}^c \times LG_t$, PC_{2008}^c, LG_t are absorbed by the fixed effects. Control variables include the time difference between purchase and collateralization, loan maturity, log of area, floor-area-ratio and land quality(level). Fixed effects include city-year fixed effects, usage fixed effects, supply method fixed effects. The standard errors are clustered at province and year level. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	< 1 *year*			< 1.5 *year*		
Dependent Variable	Distance, *meters*			Distance, *meters*		
Spread	500	1000	1500	500	1000	1500
$LGFV^d$	-0.157	-0.0294	-0.0609	-0.129	-0.137	-0.0903
	(0.198)	(0.199)	(0.213)	(0.179)	(0.199)	(0.252)
$LGFV^d \times PC_{2008}^c \times LG_t$	1.835***	2.209***	2.283***	0.978***	1.320***	1.616***
	(0.0436)	(0.102)	(0.222)	(0.0384)	(0.0868)	(0.180)
$LGFV^d \times PC_{2008}^c$	-0.110	-0.170***	-0.174***	-0.0192	0.0227	-0.0138
	(0.0612)	(0.0233)	(0.0223)	(0.0444)	(0.117)	(0.115)
$LGFV^d \times LG_t$	-0.233***	-0.730***	-0.743***	0.308***	-0.177***	-0.632***
	(0.0435)	(0.103)	(0.157)	(0.0281)	(0.0111)	(0.0366)
Timediff	0.104	0.0969	0.0928*	0.0558	0.0432	0.0315
	(0.0748)	(0.0800)	(0.0412)	(0.0577)	(0.0454)	(0.0331)
Maturity	0.0602**	0.0587**	0.0552**	0.0606**	0.0589**	0.0539**
	(0.0209)	(0.0150)	(0.0173)	(0.0176)	(0.0132)	(0.0158)
Logarea	-0.179***	-0.194***	-0.194**	-0.196**	-0.217**	-0.210**
	(0.0372)	(0.0375)	(0.0423)	(0.0449)	(0.0583)	(0.0611)
LogFAR	-0.234	-0.283	-0.313*	-0.261	-0.312**	-0.323**
	(0.165)	(0.151)	(0.127)	(0.124)	(0.0926)	(0.0923)
Level	0.0136	0.0130	0.0148	0.0119	0.0139	0.0174*
	(0.0144)	(0.0124)	(0.00843)	(0.0126)	(0.00774)	(0.00734)
Observations	7,715	8,886	9,946	10,302	11,921	13,467
R-squared	0.480	0.468	0.460	0.447	0.437	0.427
Controls	YES	YES	YES	YES	YES	YES
CityYear FE	YES	YES	YES	YES	YES	YES
Other FEs	YES	YES	YES	YES	YES	YES

leads to the rising housing prices in China. We also find the underlying mechanism to accelerate the local credit supply in China by local governments' monopolistic power of the land market. Overall, our findings contribute to a better understanding of the role of local government financing, macro policies and potential effects on the asset market.

Chapter 2

IPO Prospectus Language and Long-Run Performance[1]

Abstract

Based on a textual analysis of IPO prospectuses of US firms from 1997 to 2016, we find that firms with higher (abnormal) negative sentiment significantly outperform those with lower levels over the long term, using both pooled data and calendar time portfolio approaches. A variety of sentiment and performance measures reinforces the above finding. The story that retail investors buy stocks with less negative sentiments in IPO initially and later realize lower returns could explain our result.

Key words: IPO, Textual Analysis, S-1 Filings, Long-run Performance
JEL codes: G14, G18, G24

2.1 Introduction

A growing literature of textual analysis on news or financial reports submitted to the U.S. Securities and Exchange Commission (SEC) shows that the sentiment revealed in these news or reports affects short-term stock returns. Loughran and McDonald (2013) demonstrates that the sentiment of uncertainty in form S-1 which is filed at the beginning of initial public offering (IPO) process is positively related to the first day underpricing. Li (2006) shows that risk sentiment in firms' annual reports predicts lower earnings in the next year and lower returns in twelve months after the annual report filing date, where the paper defines risk sentiment as the frequency of "risk" and "uncertain" in 10-K filings. Tetlock, Saar-Tsechansky, and Macskassy (2008) shows that the negative word frequency in news predicts lower earnings and returns on the following day. Although news are quite short-term, annual basis 10-K filings reflect firms comprehension of their future prospect as well as risks they may face, and S-1, a filing for SEC and investors to have more knowledge of the IPO firms, should be more long-term viewed. While first-day underpricing reflects the game among issuers, underwriters, existing investors and debtholders, it is the wealth growth in the long run that matters for investors who buy the stock after the first day, as

[1]This paper is co-authored with Rustam Abuzov from University of Virginia.

well as investors in the primary market who are not speculators. Moreover, the long-term return is also a reflection of mispricing of IPOs. In this paper, we analyze the effect of sentiments in S-1 filings on IPOs' long-term performance.

If the market is efficient, all stocks should be correctly priced at any time. However, the market efficiency is heterogeneous among stocks. If the negative or uncertain sentiment reflects the difficulty for investors to value the stock, the negative or uncertain sentiment should predict positive long-term return. For IPOs, less information is available for the market, and this positive association should be more significant as the market takes more time to absorb the existing information and obtain new information.

A firm files S-1 to SEC at the beginning of IPO process, then amends more information by submitting S-1/A, and if the IPO is successful, the firm submits a 424 filing at the day of IPO or within days after IPO. Thus, the 424 filings contain more information from SEC's comments, book-building process, while S-1 reflects more information of the manager's opinion on firms' business models and the perspective regarding the firm's future. Thus, our analysis is mainly based on S-1 filings, but in untabulated results, we also use 424 documents as a robustness check, and get similar results.

Our empirical analysis is based on a sample of completed IPOs in the United States during 1997 and 2016 with their S-1 filings available from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, returns available from CRSP, and financial data available from Compustat. To measure the sentiment of S-1 filings, we exploit the seven sentiment word lists specifically for financial statements from Loughran and McDonald (2011): positive, negative, uncertain, litigious, weak modal, strong modal and constrained. These word lists have been widely used in finance literature to measure the sentiment of news, S-1s, 10-Ks, and conference calls. We count the frequency for each word list, and divide it by the total word count of the S-1 filing to get the sentiment measure. We get seven sentiment measures for each S-1 filing. Then we test which sentiment measure has an effect on the long-term return of IPOs.

Our empirical analysis includes three parts. First, we run a cross-sectional OLS regression to test the correlation between each sentiment measure and the long-term return of IPOs, with a variety of controls that have been shown to have predictability of long-term returns. The results reveal that negative sentiment and litigious sentiment in S-1 are associated with higher long-term returns, and the predictability is statistically and economically significant. Our focus from now on is on negative sentiment. Second, we use quantile regression methodology to test the effect of sentiment on the long-term return since long-term returns are highly right skewed. The results of quantile regressions show that the positive correlation between negative sentiment and long-term return is concentrated on the right side of the long-term return distribution. At last, we form calendar time portfolios based on the quintiles of negative sentiment measures in the past 3 or 5 years, and hold the portfolio for 3, 6, 9 and 12 months. We then measure the difference between the top quintile (highest sentiment measure) and bottom quintile (lowest sentiment measure) using the Fama-French three-factor model augmented with Carhart (1997) momentum factor. The results show that a strategy that long the portfolio of IPOs with the highest negative sentiment and short the portfolio of IPOs with the lowest negative sentiment within 3 years after IPO yields adjusted returns ranging from 0.9% to 4.28% for holding periods of 3 to 12 months. For a short summary, we show that negative sentiment in S-1 filings significantly predicts higher IPO returns at long horizon, with the controls of a number of IPO-related and firm characteristics. Our long-term result complements Loughran and McDonald (2013) on the effect of uncertainty sentiment on first-day underpricing, where they get a positive link between the uncertain sentiment and first day return for IPOs.

Next, we discuss some explanations for our findings. First, the overperformance could be a reflection of higher risks taken by the firm. Since our measures of long-term returns are adjusted by benchmarks, the risks here are firm specific risks. The idiosyncratic risk is estimated from a 4-factor model. The sorting results show that idiosyncratic volatility increases with negative sentiments, but the regression results do not support the higher risking taking story once additional variables are controlled.

Second, the misvaluation of IPOs could come from that managers mislead investors to undervalue or overvalue their firms. If a firm going public is managed by an overconfident executive, it is more likely that its S-1 contains less negative words and more positive words. With the asymmetric information between managers and investors, investors are more likely to overvalue this stock in the short term, causing a positive prediction of negative sentiment on long-term returns. We sort managers' overconfidence by negative sentiment to test the story. We measure managers' overconfidence as the count of CEO-year of CEOs who had options more than 67% in-the-money during their tenure (Malmendier and Tate (2005)). The results show the opposite. Our data is limited in analyzing CEO overconfidence, and more detailed research is needed.

Third, it could be that stocks with higher negative sentiment are associated with less retail investor attention, and these stocks realize higher returns in the long run with more information being absorbed by the market. By analyzing the negative sentiments in each section on S-1 filings, we find that the negative sentiment in the *Summary* part is significantly correlated with long-run overperformance. Individual investors care more about the summary part, while disregard other more important parts, such as risk factors. This implies that individual investors are easily affected by sentiments in S-1 filings, and buy stocks with less negative sentiment, which later realize lower returns.

To get more insights of attention-based explanation, we exploit the search volume index from Google Trends (Da, Engelberg, and Gao (2011)). Specifically, we utilize the abnormal search volume index for each IPO as a proxy for retail investor attention and then run regressions with retail investor attention as the mediator between negative sentiment and long-term returns. The story is that there is a positive link between retail investor attention and retail investor sentiment, and retail investors' attention prior to IPO induces greater buying pressure which drives up the first-day return of IPO stocks, and the dissipation of the buying pressure generates long-run underperformance. Our results imply that the association between negative sentiments and long-term performance is fully through retail investor attention.

In summary, we find that the negative sentiment predicts higher long-term returns of IPO firms, and this association can be explained by retail investors' irrationality. Our evidence contributes to two strands of literature: textual analysis and explanation of IPO long-term underperformance.

Most of the existing literature on textual analysis focuses on short-term returns of stocks, normally 1 to 3 days. The text they use determines the horizon of returns: news are quite frequent and thus returns should be at high frequency as well; annual reports are in yearly basis, so the maximum return horizon is 1 year. However, S-1 is filed at the beginning of the IPO process, which reflects firms' long-term outlook for the future, and therefore, sentiments revealed in S-1 should have a long-term effect if the stock is not corrected priced originally. Our results contribute to the analysis of sentiments and returns by complementing the existing literature with a longer matched horizon between documents and returns. Secondly, literature explains the long-term underperformance of IPOs with market timing (Schultz (2003)), institutional ownership (Field and Lowry (2009)), venture capital backing (Brav and Gompers (1997)), and others. Ritter and Welch (2002) shows that the long-term underperformance may be caused by overconfidence of both entrepreneurs and investors. Our results complement this point by analyzing the effect of overconfidence of managers and investors on long-term performance of IPO firms with sentiments on S-1 filings.

The remainder of the paper is organized as follows. Section 2.2 describes the sample and variable construction process. Section 2.3 presents the empirical test of the link between document sentiments and long-term returns with different statistical methods. Section 2.4 provides some explanations of the above proved links. Section 2.5 concludes the paper.

2.2 Sample and Variable Construction

2.2.1 Sample Construction

Our focus is on IPOs which file S-1 documents in the United States. Our sample period is from 1997 to 2016 as the number of S-1 filings for IPO starts to stabilize in 1997, and the sample stops in 2016 because we need at least a 3-year period to calculate long-term returns. Our data is from several sources. The first is Thomson/Refinitiv Securities Data Company (SDC)'s New Issue database. We exclude ADRs, units, REITs, closed-end funds and IPOs by financial firms, IPOs with offer price less than 5 dollars. The sample is corrected according to Professor Jay Ritter's website. The second source is Securities Exchange Commission (SEC)'s Electronic Data Gathering, Analysis, and Retrieval system (EDGAR). We download all S-1, S-1/A, and 424B filings for the sample period. A detailed parsing and cleaning process for EDGAR filings is available in section 2.2.2. The third source is Center for Research in Security Prices (CRSP) which provides prices and returns of stocks listed in U.S. Finally, we get financial information of IPO firms from Compustat.

We start each matching process with the SDC database, and at the end use identifiers from other sources as cross checks. We match the SDC database and EDGAR filings with firms' central index key (CIK). We use CUSIP as the identifier to match the SDC database with CRSP price data, where we get the permno of the IPO firms. We match the SDC database with the financial data using CIK and CUSIP, and check the permno with CRSP price data. Table 2.1 shows the matching process and observations of each source and matched observations. Our sample starts with 6,206 new issues from SDC[2]. According to the tradition of cleaning the initial public offering data from SDC: 572 new issues are dropped due to the unavailability of founding year data from Professor Jay Ritter's website; 229 new issues are not in the United States; 379 issues belong to ADR, Units, REITs, or closed-end funds; and there are 90 new issues with offer price less that 5 dollars. 4,936 new issues are left for merging with financial information, returns and S-1 documents. 449 firms are dropped due to unavailability of financial information. We require the IPO firm to have return/price data within 7 days from the day of IPO, which removes 135 firms. The reason why 1,400 firms don't have S-1 filings is that in the U.S., S-1 is the most common filing for registration of securities, but other forms should be filed depending on the size or identity of the company. SB-2 is for small business, F-1 is for

[2]The original dataset contains 25,619 observations, and by dropping issues which are not initial public offerings, and duplicates where a firm gets listed on several stock exchanges, 6,206 new issues are left.

foreign companies, and N-2 is for investment company registration. At the end, we drop financial firms (SIC code: 6000-6999) following other papers. The final sample consists of 2,439 IPOs listed on NYSE, Nasdaq, AMEX during 1997-2016 period.

Table 2.1: Sample Construction

This table reports sample construction process of IPOs during 1997-2016. The sample is obtained from Thomson/Refinitiv Securities Data Company (SDC)'s New Issue database and corrected by Professor Jay Ritter. ADRs, units, REITs, closed-end funds, financial firms are excluded from the sample. IPOs also need to have offer price of at least $5 per share. We also require the IPOs to have founding year data from Jay Ritter's website, financial information from Compustat, price data from CRSP and S-1 documents.

	Observations removed	Remaining sample size
New issues from SDC		**6,206**
Founding year available	572	
Listed in US	229	
ADR, Units, REITs, Closed-end funds	379	
Offer price less than $5	90	
SDC new issues after cleaning		**4,936**
Financial information available	449	4,487
CRSP price/return available	135	4,352
S-1 filings available	1,400	2,952
Drop Financial firms	513	2,439
Sample size		**2,439**

2.2.2 Parsing the S-1 Documents and Word Lists

After downloading all the S-1 documents, we clean each individual document following the parsing procedure demonstrated in Loughran and McDonald (2011). Moreover, we remove "PART II: Information Not Required In The Prospectus" and the content thereafter. PART II in IPO prospectus mainly contains financial reports and notes, and even after removing tables, notes still contain negative words like "loss" from income statement item "Net (loss) income". These words affect the measure of frequencies of word lists, but do not reflect any sentiments. And previous research shows that the Risk factors, Management Discussion and Analysis (MD&A) sections are more informative than other parts[3]. The cleaned document is then parsed into words for the subsequent analysis.

Similar with Loughran and McDonald (2013) and Tetlock, Saar-Tsechansky, and Macskassy (2008), we measure document sentiment using the simple frequency of various word lists from Loughran and McDonald (2011). The word lists developed by Loughran and McDonald (2011) is specifically for financial documents, and they classify words into 7 lists: positive, negative, uncertain, litigious, weak modal, strong modal and constrained[4]. By analyzing SEC's 10-K filings in the U.S., Loughran and McDonald (2011) shows that their list of words with negative sentiment typically has a negative meaning in financial reports.

For each word list j, we count the word frequency in each of the cleaned document i and divide it by the total word count of the cleaned file i to get the proportion of the word list.

$$ratio_{i,j} = \frac{Wordcount_{i,j}}{Wordcount_i}$$

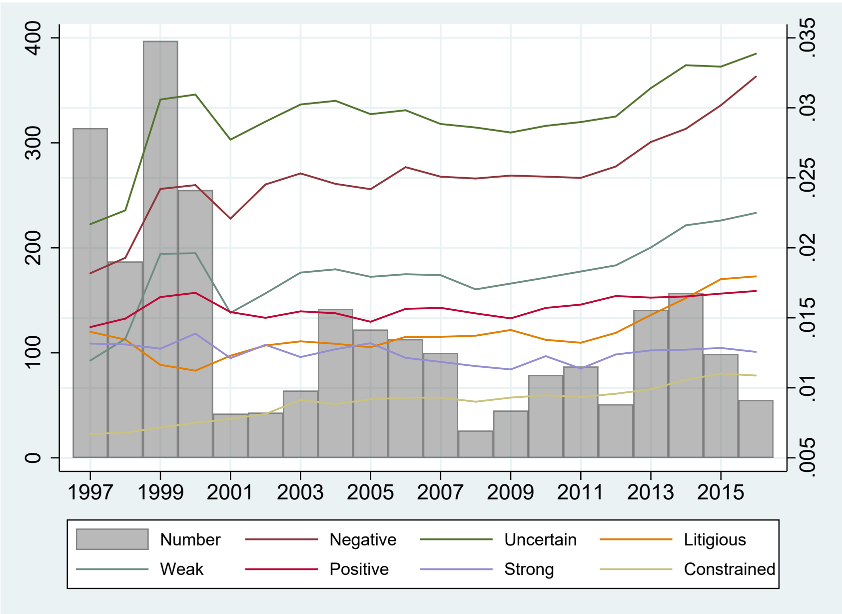
[3]See eg. Hanley and Hoberg (2010).

[4]The word lists are available on https://sraf.nd.edu/textual-analysis/resources/.

Figure 2.1 shows the time trend of initial public offerings along with different sentiments during the sample period. There is a boom of IPOs at the end of the 20th century due to the dot bubble, and immediately followed a bust, then the listing started to recover from the year 2004. IPO listings reached the bottom during the Global Financial Crisis and then rose again from the year 2010. Regarding the trend of different sentiments, the frequencies of positive or strong modal words have a flat or even downward trend, while other sentiments rise especially after the Global Financial Crisis. Negative, uncertain, and weak modal words rise the most among the 7 word lists. Loughran and McDonald (2013) argues that negative, uncertain, and weak modal words reflect the uncertainty about future cash flow projections, thus affecting the initial valuation and then predict first-day underpricing of IPO companies.

Figure 2.1: Time Trend of Sentiments

This graph plots the proportions of words contained in the S-1 filings and number of IPOs for 1997-2016. The left axis measures the number of IPOs *Number*. The right axis measures the average ratio for each word list: positive, negative, uncertain, litigious, weak modal, strong modal and constrained.



We exploit several methods of detrending to get the firm specific loading on the sentiment instead of reflecting a general fashion. For each IPO i, each word list j, we compute the following measures,

$$\begin{aligned} DF_{i,j,t} &= ratio_{i,j,t} - \text{average ratio}_{j,t-1} \\ SF_{i,j,t} &= \frac{ratio_{i,j,t} - \text{average ratio}_{j,t-1}}{\sigma_{j,t-1}} \\ DFI_{i,j,t} &= ratio_{i,j,t} - \text{average ratio}_{j,s,t-1} \end{aligned}$$

where s denotes the 48 industry classification developed by Fama and French (1997); t denotes the filing date of S-1 document[5]. So average ratio$_{j,t-1}$ represents the average ratio for word list j of all IPO firms listed in the past 365 calendar days. And average ratio$_{j,s,t-1}$ represents the average ratio for word list j of all IPO firms in industry s and get public in the past 365 calendar days. $\sigma_{j,t}$ denotes the variation of the corresponding ratio in the

[5]We get similar results if using IPO date.

past 365 days based on t[6]. Therefore, $DF_{i,j,t}$ and $DFI_{i,j,t}$ are the demeaned version of the corresponding sentiment measures, while standardizing the original sentiment measures, we get $SF_{i,j,t}$. It's worth mentioning that the standardization removes IPOs in the first year, and IPOs for which no IPO gets listed in the past 365 calendar days or in the same industry.

Following Field and Lowry (2009), we construct another measure of abnormal ratio by controlling both year and offer size effects with a fractional logit methodology. Specifically, for each year, we model the conditional mean of each of the sentiment measure as:

$$E[ratio_{i,j,t}|x] = \frac{exp(\beta_1 + \beta_2 * Proceeds_i)}{1 + exp(\beta_1 + \beta_2 * Proceeds_i)}$$

The we use the difference between the actual value and expected value to measure the unexpected sentiment:

$$Err_{i,j,t} = ratio_{i,j,t} - E[ratio_{i,j,t}|x]$$

In later analysis, we mainly show the results for $SF_{i,j,t}$, as it removes trend in the sentiment measures and keeps more sample observations for analysis. For robustness tests, we also show results for other measures listed above.

2.2.3 Long-Run Performance

Similar with Ritter (1991), we use wealth relative based on buy-and-hold returns as the measure of long-run performance of IPO companies.

$$BHR_i = \prod_{t=1}^{T}(1 + r_{i,t})$$

$$WR_i = \frac{BHR_i}{BHR_{benchmark}}$$

where the return accumulation starts from the second trading day of IPO i to eliminate the first-day underpricing effect; T equals either 36 trading months or 60 trading months, and each trading month consists of 21 trading days. To be comparable with past studies, we use 3 groups of benchmarks:

- Index: to account for the effect of market factors[7].
- Fama-French 48 industry portfolio returns: to account for the return trend for different industries[8].
- Fama-French size and B/M matched portfolio returns: Brav and Gompers (1997) shows that the long-run underperformance of IPOs is not specific to IPOs, firms with similar size and B/M ratio that have not issued equities also have poor performance. So we use this benchmark to account for the size and growth effects.

[6] We do not use $\sigma_{j,s,t}$, the variation of the corresponding ratio in the past 365 days and in industry s, because there are not enough observations for the calculation of standard deviation, which will generate extreme values.

[7] Value weighted, equal weighted NYSE/AMEX/NASDAQ index returns.

[8] Both value weighted and equal weighted returns are available from Professor Kenneth R. French's website.

In section 2.3, we mainly focus on Fama-French 48-industry adjusted returns, because index adjusted returns can not reflect firm specific characteristics, while missing B/M data for IPOs further makes the sample size smaller. We also report robust results where returns are adjusted by other benchmarks.

2.2.4 IPO Control Variables

We use issue-related and firm-specific variables prior to IPO taken from past literature to explain the long-run performance of IPOs. The issue related variables include: (1) Nasdaq buy-and-hold returns 30 days prior to IPO as a measure of hotness of IPO market; (2) *up revision*: 1 if the IPO firm has an positive change from the mid-point of the last filing range to the offer price, 0 otherwise; Loughran and McDonald (2013) argues that the uncertain/negative tone in S-1 may be positively related to an adjustment of offer prices, as the tone reflects a weak informational position of issuers relative to underwriters; (3) *Lead UW's $ market share*, share of proceeds raised of successful IPOs by the lead underwriter in the past calendar year; (4) *lead underwriter's ranking* as a measure of the underwriter's reputation taken from Loughran and Ritter (2004) and updated in Professor Jay R. Ritter's website; (5) *VC dummy*, 1 if the IPO is backed by a venture capital, 0 otherwise; (6) *Internet dummy*, 1 if the firm is classified as an internet firm, 0 otherwise, taken from Professor Jay R. Ritter's website; (7) *Days between S-1 and IPO*: the log value of the period from the S-1 filing of this IPO to the IPO date; (8) *First-day return*: first day underpricing measured as the return from the offer price to the first-day closing price; (9) *Log(proceeds)*: log of proceeds raised from the issue. Firm-specific variables include: (1) *Lagged Log(asset)*: log of assets at the fiscal year end prior to IPO; (2) *Lagged ROE*: return on equity at the fiscal year end prior to IPO; (3) *Log(age)*: log(1+age), where age is the difference between IPO year and founding year taken from Professor Jay R. Ritter's website.

2.2.5 Summary Statistics

Table 2.2 reports the summary statistics of financial information, IPO process, S-1 filings, and long-term performance measures for the IPO sample.

Panel A reports the characteristics of the IPO firms. The average firm size (asset) prior to the IPO process is 456 million dollars, which is relatively small compared to a typical firm listed in NYSE/Nasdaq/AMEX. Average ROE before IPO is -0.136 and more than half of the sample has negative trailing returns, which is consistent with the evidence shown in Loughran and McDonald (2013), and Gao, Ritter, and Zhu (2013). Firms' age when they go public varies from just established to more than 150 years old with an average of 18 years. About 53% of the IPOs are backed by venture capitals. 20% of the IPOs are categorized as Internet firms and their initial public offerings concentrated on the year 1999 and 2000.

Panel B reports summary statistics related to the IPO process. A typical IPO raised 162 million dollars from the IPO process. The number of days between S-1 and IPO date varies from 23 calendar days to 1266 calendar days with an average of 113 calendar days. More than 75% the IPOs have positive return on their first-day trading and experience an average first-day return of 29.8% in our sample, while the median return is 12.5%. Focusing on the same period as Loughran and McDonald (2013), 1997-2010, we get similar patterns for the first-day return, 34% for the mean value and 13% for the median value,

Table 2.2: Summary Statistics

This table reports summary statistics of 2439 U.S. listed IPO firms during 1997-2016. *Log(age)* equals *log(IPO year - Founding year + 1)*, where *Founding year* data is obtained from Jay Ritter's website. *Lagged Log(assets)*, *Lagged ROE* are taken from the financial statements of the last period before IPO. *VC dummy* is obtained from Jay Ritter's website, measuring whether the firms is venture capital backed or not. *Internet dummy* is obtained from Jay Ritter's website, measuring whether the firm is an internet firm or not. *Log(proceeds)* takes log of proceeds raised from IPO (in million of U.S. dollars). Days between S-1 and IPO (log) is the log of calendar date difference between IPO date and S-1 filing date. *Prior NASDAQ 30-day return*: Nasdaq buy-and-hold returns 30 calendar days prior to IPO. *up revision* is 1 if the IPO firm has an positive change from the mid-point of the last filing range to the offer price, 0 otherwise. *Lead UW's $ market share* is the share of proceeds raised of successful IPOs by the lead underwriter in the past calendar year. *Lead UW's rank* is a measure of underwriter reputation taken from Loughran and Ritter (2004) and updated in Professor Jay R. Ritter's website. Words in S-1 filings are classified as 7 categories according to Loughran and McDonald (2011): positive, negative, uncertain, legal, weak modal, strong modal and constrained, and then divided by the word counts in S-1 to get proportions for each word list. Panel D reports statistics for long-term performance, where long-term performance is measured as $WR_i = BHR_i/BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1 + r_{i,t})$, where T takes value of 36 months(*3-year*) or 60 months(*5-year*); *benchmark* includes market return (*index*), Fama-French 48 industry return(*industry*), size and B/M matched return(*size-B/M*).

	(1) N	(2) Mean	(3) SD	(4) Min	(5) P25	(6) Median	(7) P75	(8) Max
Panel A: Firm information								
Log(age)	2,439	2.392	0.938	0	1.792	2.303	2.890	5.112
Lagged Log(assets)	2,346	4.044	1.900	-6.908	2.854	3.815	5.162	11.82
Lagged ROE	2,306	-0.136	2.025	-5.724	-0.683	-0.0219	0.298	6.491
VC dummy	2,439	0.535	0.499	0	0	1	1	1
Internet dummy	2,439	0.203	0.402	0	0	0	0	1
Panel B: IPO process								
Log(proceeds)	2,439	4.430	0.955	1.386	3.807	4.323	4.886	9.681
Days between S-1 and IPO (log)	2,439	4.505	0.613	3.135	4.159	4.431	4.771	7.414
Prior NASDAQ 30-day return	2,438	0.00101	0.00258	-0.0117	-0.000441	0.00109	0.00250	0.0102
First-day return	2,439	0.298	0.513	-0.197	0.00417	0.125	0.340	2.719
Up revision	2,439	0.239	0.427	0	0	0	0	1
Lead UW's $ market share	2,195	0.0876	0.0766	0.00135	0.0285	0.0631	0.130	0.462
Lead UW's rank	2,430	8.038	1.340	1	8	8.500	9	9
Panel C: Form S-1								
Negative	2,439	0.0243	0.00569	0.00897	0.0206	0.0243	0.0276	0.0781
Positive	2,439	0.0160	0.00301	0.00724	0.0140	0.0159	0.0177	0.0418
Uncertain	2,439	0.0289	0.00544	0.0125	0.0254	0.0292	0.0325	0.0699
Litigious	2,439	0.0138	0.00352	0.00520	0.0113	0.0133	0.0159	0.0349
Strong	2,439	0.0131	0.00275	0.00524	0.0113	0.0128	0.0144	0.0290
Weak	2,439	0.0179	0.00458	0.00441	0.0146	0.0181	0.0210	0.0547
Constrained	2,439	0.00858	0.00212	0.00230	0.00703	0.00845	0.00993	0.0227
Panel D: Long-term performance (wealth relative)								
3-year WR, index	2,439	0.873	1.435	8.15e-06	0.158	0.491	1.155	27.64
5-year WR, index	2,110	0.838	1.464	8.15e-06	0.102	0.425	1.089	30.89
3-year WR, industry	2,427	0.879	1.263	1.26e-05	0.179	0.529	1.183	17.69
5-year WR, industry	2,098	0.855	1.371	9.90e-06	0.130	0.450	1.127	25.48
3-year WR, size-B/M	2,234	1.063	2.013	0.000151	0.164	0.547	1.340	39.03
5-year WR, size-B/M	1,936	1.031	1.895	9.64e-06	0.109	0.468	1.335	36.10

but the 599 IPOs during 2011-2016 only have average first-day return of 19%. Thus, our sample is comparable to previous research. Only 23.9% IPOs in the sample have positive change from the mid-point of the last filing range to the offer price. Similar as statistics shown in Loughran and McDonald (2013), more than 75% of the IPO firms in the sample are taken public by a top-tier underwriter (ranking at least 8), where the lead underwriter's market share prior to this IPO is 8.76% on average.

Panel C reports the summary statistics of sentiment word list fractions for S-1 filings. For each word list, the mean and median are very close, with little variation. For example, the mean of negative word ratio is 2.43%, which is almost identical to its median. The standard deviation is 0.569%, and the range from minimum to maximum is 0.9% to

7.8%. Negative, uncertain and weak modal words account for more proportion than other word lists. Our summary statistics are larger than those reported in Loughran and McDonald (2013), as we only focus on the main text of the prospectus and thus "PART II INFORMATION NOT REQUIRED IN THE PROSPECTUS" is dropped in our analysis. As a robustness, we replace our measure with sentiment measures based on the whole S-1 document. See more details in section 2.3.1.

Panel D reports the summary statistics of adjusted long-term returns. More than half of the IPOs underperform their benchmarks. All the medians are smaller than their averages, suggesting the wealth relative measures are right skewed. Our wealth relative measures with respect to market index or industry have averages smaller than 1, reflecting the long-run underperformance of IPOs. However, when we use size and B/M matched portfolio returns as the benchmark, IPO firms and their matched firms get almost the same return on average, with an average value of 1.063 for 3-year measure, and the median is smaller than the average, which is 0.547.

2.3 Empirical Results

In this section, we explore the link between the long-term performance and the different sentiments of S-1 documents. First, we use the simple OLS regression by treating each IPO equally. The long-term performance measures are skewed which may make statistical inference of the OLS result problematic with our sample size, so we use quantile regression to solve the problem. Another issue is that returns on individual IPOs especially for long-run buy-and-hold measures, overlap as we treat each IPO equally. Calendar time portfolio formation avoids this problem as it treats each month equally. In the OLS and quantile regressions, we winsorize the extreme 2% value of long-term performance measures, and the extreme 1% sentiment measures[9].

2.3.1 OLS Regression

In this part, we estimate cross-sectional regressions of cumulative three-year (adjusted) long-term returns on sentiment measures of S-1 filings, with a variety of variables controlled. Our specification is as follows:

$$WR_i = \alpha + \beta sentiment_i + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i \tag{2.1}$$

where the dependent variable, WR_i is the adjusted three-year buy-and-hold return of IPO firm i. The explanatory variable we are interested in is $sentiment_i$, which is one of the 7 different sentiment measures defined in section 2.2.2, and we run regression for each of the sentiment measures. The vector $\mathbf{X_i}$ contains a number of control variables that are described in section 2.2.4. IPOs issued in the same year may capture similar characteristics that will affect the correlation between the sentiment measures and long-term returns. For example, firms that are expected to perform worse are more inclined to go public in the "hot market" and show more optimistic outlook in their prospectuses, which makes β for negative word list positive. However, this negative β only reflects the effect of market timing and long-term poor performance. Therefore, the time fixed effect Year FE captures the market timing effect and makes us only focus on the within year differences of IPO firms. Firms in one industry may have better performance and their

[9]The results are similar for different winsorizing criteria, and also for raw data with no winsorization.

prospectuses contain more negative words due to industry-specific characteristics which we don't capture in the regression analysis, then β for negative word list actually shows the correlation between long-term performance and the industry-specific characteristics. So we include industry fixed effect FF48 FE, which denotes the Fama-French 48-industry dummies, to capture the above mentioned effect and focus on within-industry variations. The standard errors are clustered at calendar quarter and industry level[10].

Table 2.3 reports the regression results of equation 2.1. In this table, Fama-French 48 industry value-weighted portfolio return is used as the benchmark to calculate wealth relatives. In Table 2.4 we show results for other benchmarks. Table 2.5 presents results for various standardization methods on sentiment ratios.

Each column in column (1) and column (3)-(8) contains a different word list defined in Loughran and McDonald (2011). As a comparison, column (2) shows results for *Finneg* defined as negative word count divided by the word count of the whole S-1 document. *Finneg* is obtained from WRDS-SEC database. Most of the estimated coefficients for the seven word lists are insignificant, except the ratios for negative and litigious word lists. The results in column (1) and (5) indicate that negative sentiments and litigious sentiments in the prospectus predict better long-term performance. One standard deviation increase in negative sentiment predicts an increase in long-term performance of 5.88% compared to firms listed in the same industry, in the same year, which amounts to a 3-year buy and hold return of 6.64%[11]. One standard deviation increase in litigious sentiment predicts an increase in long-term performance of 6.83% compared to firms in the same industry, in the same year. More negative sentiment in the S-1 filings may reflect uncertainty of future cash flows, which makes investor hard to value the stock, or reflect the conservatism of managers. These cause the stock price lower than their fundamental value and in the long-run, when the market reveals its true value, the firm has better long-term performance. We try to provide more detailed evidence in section 2.4. Column (2) shows that the negative ratio utilized in previous research is positively related to long-term performance of IPOs, but insignificantly. Thus, including the "PART II" in the prospectus into analysis affects the results, and from the reasons mentioned in Section 2.2.2, we believe that our results are more convincing.

The negative coefficient of Prior NASDAQ 30-day return shows that firms go public during hot market perform worse, which is consistent with the market timing hypothesis. IPOs with higher ranking underwriters (Lead UW's rank) perform better in the long term. Venture capital-backed IPOs (VC dummy) show lower long-term performance, while in Brav and Gompers (1997), using value-weighted returns, ventured-backed IPOs and nonventure-backed IPOs have similar performance for 5-year returns. We presume two reasons that could explain this difference. First, there is no overlap between the two samples. The sample in Brav and Gompers (1997) ends in 1992 and consists of 934 venture-backed IPOs, 3407 nonventure-backed IPOs. Our sample includes 1,305 venture-backed IPOs, 1,134 nonventure-backed IPOs from 1997 to 2016. Venture-backed IPOs are more prevalent from 2007, so venture capitals may not select better firms as before. Second, we include sentiment measures which may reflect effects of venture capitalists. If we drop our sentiment measures from the regression, VC dummy is no longer significant which is consistent with Brav and Gompers (1997). The more mature (Log(age)) and

[10]We use calendar quarter to make sure we have enough clusters, and the results are similar if we cluster at the calendar year and industry level.

[11]Average 3-year Fama-French matched industry portfolio return is 112.86%, then one standard deviation increase in negative ratio will lead to 112.86%*0.0588=6.64% 3-year buy-and-hold return.

Table 2.3: Long-Term Performance and S-1 Sentiment Proportions, OLS

This table reports results for regression $WR_i = \alpha + \beta sentiment_i + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i$. The sample includes IPOs from 1997 to 2016. WR_i measures IPOs' long-term performance, $WR_i = BHR_i / BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1 + r_{i,t})$, where Fama-French 48 industry value-weighted return is used as the benchmark, T takes the value of 36 months. $sentiment_i$ takes proportions of one of the seven categories of word lists classified in Loughran and McDonald (2011): positive, negative, uncertain, legal, weak modal, strong modal and constrained, then standardized by the respective ratio 365 days before the filing date $SF_{i,j,t} = \frac{ratio_{i,j,t} - \text{average ratio}_{j,t-1}}{\sigma_{j,t-1}}$. $\mathbf{X_i}$ includes various control variables described in Table 2.2. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Negative	0.0588**							
	(0.0231)							
Finneg		0.0348						
		(0.0230)						
Positive			-0.00178					
			(0.0195)					
Uncertain				0.0120				
				(0.0169)				
Litigious					0.0683**			
					(0.0273)			
Strong						-0.0255		
						(0.0227)		
Weak							0.00548	
							(0.0188)	
Constrained								0.0197
								(0.0192)
Days between S-1	-0.0275	-0.0251	-0.0245	-0.0240	-0.0231	-0.0227	-0.0244	-0.0248
and IPO	(0.0443)	(0.0448)	(0.0438)	(0.0435)	(0.0447)	(0.0424)	(0.0438)	(0.0438)
Prior NASDAQ	-16.89*	-16.45*	-16.33*	-16.30*	-17.79*	-16.04*	-16.30*	-16.49*
30-day return	(9.254)	(9.239)	(9.169)	(9.153)	(8.865)	(9.256)	(8.994)	(9.172)
Up revision	0.0243	0.0186	0.0245	0.0239	0.0250	0.0255	0.0241	0.0257
	(0.0601)	(0.0587)	(0.0576)	(0.0574)	(0.0585)	(0.0580)	(0.0594)	(0.0571)
Lead UW's $	-0.0199	-0.0553	-0.0454	-0.0407	-0.0789	-0.0642	-0.0432	-0.0372
market share	(0.358)	(0.369)	(0.371)	(0.368)	(0.362)	(0.370)	(0.367)	(0.369)
Lead UW's rank	0.0820***	0.0815***	0.0838***	0.0840***	0.0858***	0.0845***	0.0836***	0.0837***
	(0.0283)	(0.0288)	(0.0299)	(0.0300)	(0.0308)	(0.0295)	(0.0295)	(0.0300)
VC dummy	-0.0817**	-0.0807**	-0.0767**	-0.0773**	-0.0643*	-0.0811**	-0.0774**	-0.0738**
	(0.0348)	(0.0363)	(0.0348)	(0.0350)	(0.0335)	(0.0342)	(0.0357)	(0.0344)
Internet dummy	-0.0942	-0.0832	-0.0741	-0.0762	-0.0797	-0.0744	-0.0757	-0.0742
	(0.0632)	(0.0634)	(0.0631)	(0.0630)	(0.0632)	(0.0634)	(0.0634)	(0.0630)
Log(age)	0.0666**	0.0643**	0.0646**	0.0663***	0.0664***	0.0601**	0.0654**	0.0664**
	(0.0257)	(0.0247)	(0.0246)	(0.0244)	(0.0240)	(0.0257)	(0.0251)	(0.0247)
First-day return	-0.154***	-0.153***	-0.148***	-0.149***	-0.143***	-0.148***	-0.149***	-0.148***
	(0.0357)	(0.0344)	(0.0339)	(0.0338)	(0.0334)	(0.0336)	(0.0337)	(0.0340)
Lagged Log(assets)	0.0860***	0.0840***	0.0816***	0.0829***	0.0823***	0.0786***	0.0822***	0.0811***
	(0.0215)	(0.0215)	(0.0226)	(0.0214)	(0.0226)	(0.0233)	(0.0208)	(0.0225)
Log(proceeds)	-0.136***	-0.140***	-0.142***	-0.140***	-0.146***	-0.138***	-0.141***	-0.143***
	(0.0393)	(0.0385)	(0.0367)	(0.0389)	(0.0367)	(0.0374)	(0.0397)	(0.0362)
Lagged ROE	0.0190*	0.0188*	0.0180*	0.0181*	0.0172	0.0184*	0.0181*	0.0179*
	(0.0100)	(0.0100)	(0.0101)	(0.0102)	(0.0105)	(0.0101)	(0.0101)	(0.00999)
Constant	0.504	0.532	0.526	0.503	0.512	0.519	0.517	0.524
	(0.335)	(0.330)	(0.324)	(0.331)	(0.335)	(0.317)	(0.336)	(0.325)
Observations	1,830	1,824	1,830	1,830	1,830	1,830	1,830	1,830
R-squared	0.128	0.127	0.125	0.126	0.130	0.126	0.125	0.126

larger the firm is (Lagged Log(asset)), the less first-day return is (First-day return), the less proceeds raised from IPO (Log(proceeds)) and the higher the return on equity is before IPO (Lagged ROE), the better is the long-term performance.

Table 2.4 reports the coefficients for negative sentiment in equation 2.1 for various return benchmarks listed in section 2.2.3. Panel A shows results for equal-weighted benchmark returns, while panel B shows those for value-weighted benchmark returns. Using size and B/M matched portfolio return as the benchmark, either equal-weighted or value-weighted, one standard deviation increase in negative ratio will lead to more increase in

Table 2.4: Negative Sentiment in OLS Regression for Various Return Benchmarks

This table reports β for $sentiment = negative$ in regression $WR_i = \alpha + \beta sentiment_i + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i$. The sample includes IPOs from 1997 to 2016. WR_i measures IPOs' long-term performance, $WR_i = BHR_i / BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1 + r_{i,t})$, where T takes value of 36 months or 60 months. Benchmark includes: index equal-weighted returns from CRSP (*Index-EW*), Fama-French 48 industry equal-weighted returns (*Indstry-EW*), size and B/M matched equal-weighted portfolio returns from Professor Kenneth R. French's data library (*Size, B/M-EW*), index value-weighted returns (*Index-VW*), Fama-French 48 industry value-weighted returns (*Indstry-VW*), size and B/M matched value-weighted portfolio returns (*Size, B/M-VW*). $sentiment_i$ takes proportions of negative word lists classified in Loughran and McDonald (2011), then standardized by the negative ratio 365 days before the filing date $SF_{i,j,t} = \frac{ratio_{i,j,t} - \text{average ratio}_{j,t-1}}{\sigma_{j,t-1}}$. $\mathbf{X_i}$ includes various control variables described in Table 2.2. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Equal weighted benchmark return

	Index-EW	Industry-EW	Size, B/M-EW
36-month	0.0461**	0.0497**	0.0664**
	(0.0178)	(0.0189)	(0.0294)
60-month	0.0649***	0.0589**	0.121***
	(0.0230)	(0.0218)	(0.0333)

Panel B: Value weighted benchmark return

	Index-VW	Industry-VW	Size,B/M-VW
36-month	0.0509**	0.0588**	0.0663**
	(0.0226)	(0.0231)	(0.0293)
60-month	0.0808**	0.0820***	0.121***
	(0.0315)	(0.0230)	(0.0333)

wealth relatives for 3-year and 5-year IPO returns. As each IPO firm is matched with firms in similar size and B/M ratio, there is less variation inside the benchmark portfolio. Thus, we get almost identical coefficients when we use size and B/M matched portfolio as the benchmark for panel A and panel B. However, for index and industry benchmarks, negative sentiments are more economically significant for value-weighted benchmark returns.

Table 2.5 reports the results for negative sentiment in equation 2.1 with various standardization methods listed in section 2.2.2, where we use Fama-French 48 industry value-weighted portfolio return as the benchmark. Column (1) shows results for negative ratio which is the unadjusted proportion of negative word list in S-1 document. The coefficient is significant at 5% statistical level, and one standard deviation increase (0.00569) leads to an increase of 3-year buy-and-hold return of 8.59%. Column (2) - (5) shows results for various standardization methods. The coefficients for negative sentiment are all significantly positive except for DFI_Neg. Although the coefficient for DFI_Neg is insignificant, the sign is still positive. We should note that including industry as another dimension makes the sample size smaller as some IPOs don't have comparable IPOs within one year, and in the "cold" market, the sentiment measures may be driven by extreme values due to less observations, making DFI_Neg not a real reflection of the deviation from the average value.

In summary, the regression results in Table 2.3, 2.4, 2.5 show that the negative sentiment in S-1 filings positively predicts long-term performance of IPO firms after controlling for a variety of underwriting, market, firm fundamental factors.

2.3.2 Quantile Regression

In this part, we exploit the quantile regression method to mitigate the high skewness of long-term return measures. Specifically, we run quantile regressions for 5 quantiles of

Table 2.5: Negative Sentiment in OLS Regression with Various Standardization Methods

This table reports β for $sentiment = negative$ in regression $WR_i = \alpha + \beta sentiment_i + \delta \mathbf{X_i} +$ Year FE + FF48 FE $+ \epsilon_i$. The sample includes IPOs from 1997 to 2016. WR_i measures IPOs' long-term performance, $WR_i = BHR_i/BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1+r_{i,t})$, where T takes value of 36 months. Benchmark is Fama-French 48 industry value-weighted returns. $sentiment_i$ takes proportions of negative word lists classified in Loughran and McDonald (2011). For each IPO i, $Neg_i = \frac{NegativeWordcount_i}{Wordcount_i}$, $DF\_Neg_{i,t} = Neg_{i,t} -$ average negative ratio$_{t-1}$ where average negative ratio$_{t-1}$ is the average negative ratio for IPOs listed within 1 year. $SF\_Neg_{i,t} = \frac{Neg_{i,t} - \text{average negative ratio}_{t-1}}{\sigma\ of\ negative\ ratio_{t-1}}$, where $\sigma\ of\ negative\ ratio_{t-1}$ is the standard deviation of negative ratio for IPOs listed within 1 year. $DFI\_Neg_{i,t} = Neg_{i,s,t} -$ average negative ratio$_{s,t-1}$, where average negative ratio$_{s,t-1}$ is the average negative ratio for IPOs listed in the same industry as IPO i within 1 year. $Err\_Neg_{i,t} = Neg_{i,t} - E[Negative_{i,t}|x]$, $E[Neg_{i,t}|x] = \frac{exp(\beta_1+\beta_2 * Proceeds)}{1+exp(\beta_1+\beta_2 * Proceeds)}$ for each year following Field and Lowry (2009). $\mathbf{X_i}$ includes various control variables described in Table 2.2. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)
Neg	13.37**				
	(6.374)				
DF_Neg		11.72**			
		(4.984)			
SF_Neg			0.0588**		
			(0.0231)		
DFI_Neg				9.684	
				(7.359)	
Err_Neg					15.43**
					(6.605)
Days between S-1 and IPO	-0.0242	-0.0270	-0.0275	-0.00376	-0.0248
	(0.0392)	(0.0440)	(0.0443)	(0.0506)	(0.0389)
Prior NASDAQ 30-day return	-19.11**	-16.96*	-16.89*	-16.18*	-19.30**
	(8.634)	(9.271)	(9.254)	(9.035)	(8.661)
Up revision	0.0329	0.0244	0.0243	0.0325	0.0334
	(0.0553)	(0.0599)	(0.0601)	(0.0624)	(0.0553)
Lead UW's $ market share	0.0640	-0.0230	-0.0199	-0.287	0.0635
	(0.363)	(0.359)	(0.358)	(0.392)	(0.360)
Lead UW's rank	0.0660**	0.0824***	0.0820***	0.0852***	0.0646**
	(0.0271)	(0.0285)	(0.0283)	(0.0302)	(0.0263)
VC dummy	-0.0408	-0.0815**	-0.0817**	-0.0748**	-0.0411
	(0.0501)	(0.0349)	(0.0348)	(0.0354)	(0.0504)
Internet dummy	-0.0562	-0.0928	-0.0942	-0.0881	-0.0592
	(0.0667)	(0.0629)	(0.0632)	(0.0673)	(0.0674)
Log(age)	0.0422	0.0664**	0.0666**	0.0675**	0.0425
	(0.0287)	(0.0257)	(0.0257)	(0.0256)	(0.0286)
First-day return	-0.158***	-0.154***	-0.154***	-0.154***	-0.159***
	(0.0322)	(0.0355)	(0.0357)	(0.0355)	(0.0329)
Lagged Log(assets)	0.0961***	0.0859***	0.0860***	0.0781***	0.0972***
	(0.0232)	(0.0215)	(0.0215)	(0.0209)	(0.0226)
Log(proceeds)	-0.133***	-0.137***	-0.136***	-0.127***	-0.156***
	(0.0381)	(0.0388)	(0.0393)	(0.0387)	(0.0333)
Lagged ROE	0.0159	0.0188*	0.0190*	0.0228**	0.0160
	(0.0130)	(0.0100)	(0.0100)	(0.00993)	(0.0129)
Constant	0.258	0.505	0.504	0.385	0.695***
	(0.405)	(0.334)	(0.335)	(0.352)	(0.250)
Observations	2,066	1,830	1,830	1,679	2,066
R-squared	0.113	0.128	0.128	0.126	0.113

the dependent variable, 10th, 25th, 50th, 75th, and 90th of equation 2.1. We only show results for negative sentiment. Table 2.6 reports the estimates of quantile regressions. Column (1) shows the regression result for negative sentiment in OLS setting. Column (2)-(5) show results for negative sentiment at 10th, 25th, 50th, 75th, and 90th quantiles. All the coefficients for negative sentiment are positive. The negative sentiment has a more significant effect on long-run performance of IPO firms on the right side of the long-term return distribution, with the coefficients ranging from 0.00715 (at the 10th percentile) to 0.142 (at the 90th percentile).

Table 2.6: Long-Term Performance and S-1 Sentiment Proportions, Quantile Regression

This table reports results for regression $WR_i = \alpha + \beta SF\_Neg_i + \delta \mathbf{X_i}$ + Year FE + FF48 FE + ϵ_i using quantile regression method. The sample includes IPOs from 1997 to 2016. WR_i measures IPOs' long-term performance, $WR_i = BHR_i/BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1+r_{i,t})$, where Fama-French 48 industry value-weighted return is used as the benchmark, T takes the value of 36 months. $SF\_Neg_{i,t} = \frac{Neg_{i,t} - \text{average negative ratio}_{t-1}}{\sigma\ of\ nagative\ ratio_{t-1}}$ where average negative ratio$_{t-1}$ is the average negative ratio for IPOs listed within the past 365 days, where $\sigma\ of\ negative\ ratio_{t-1}$ is the standard deviation of negative ratio for IPOs listed within the past 365 days. $\mathbf{X_i}$ includes various control variables described in Table 2.2. Column (1) shows the OLS results with robust standard errors clustered at industry and quarter level. Column (2)-(5) shows results at 10th, 25th, 50th 75th and 90th quantiles and standard errors are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) OLS	(2) 0.1	(3) 0.25	(4) 0.5	(5) 0.75	(6) 0.9
Negative	0.0588**	0.00715	0.0188	0.0416**	0.0851**	0.142*
	(0.0246)	(0.0280)	(0.0234)	(0.0210)	(0.0390)	(0.0742)
Days between S-1 and IPO	-0.0275	0.00951	0.00118	-0.0152	-0.0461	-0.0870
	(0.0361)	(0.0393)	(0.0329)	(0.0293)	(0.0544)	(0.104)
Prior NASDAQ 30-day return	-16.89**	-8.301	-10.24	-14.03**	-21.22*	-30.74
	(8.120)	(8.915)	(7.454)	(6.653)	(12.32)	(23.53)
Up revision	0.0243	-0.0169	-0.00763	0.0106	0.0451	0.0909
	(0.0508)	(0.0509)	(0.0425)	(0.0380)	(0.0703)	(0.134)
Lead UW's $ market share	-0.0199	0.646	0.496	0.202	-0.356	-1.093
	(0.429)	(0.486)	(0.406)	(0.363)	(0.672)	(1.283)
Lead UW's rank	0.0820***	0.00997	0.0262	0.0581***	0.118***	0.198***
	(0.0228)	(0.0232)	(0.0195)	(0.0175)	(0.0324)	(0.0618)
VC dummy	-0.0817	-0.0568	-0.0624	-0.0734*	-0.0942	-0.122
	(0.0500)	(0.0555)	(0.0464)	(0.0414)	(0.0766)	(0.146)
Internet dummy	-0.0942	-0.0733	-0.0780	-0.0872*	-0.105	-0.128
	(0.0610)	(0.0631)	(0.0527)	(0.0470)	(0.0871)	(0.166)
Log(age)	0.0666**	0.0599*	0.0614**	0.0644***	0.0699	0.0773
	(0.0294)	(0.0322)	(0.0269)	(0.0240)	(0.0445)	(0.0850)
First-day return	-0.154***	-0.0102	-0.0425	-0.106***	-0.226***	-0.385***
	(0.0451)	(0.0399)	(0.0334)	(0.0301)	(0.0557)	(0.106)
Lagged Log(assets)	0.0860***	0.0415**	0.0516***	0.0712***	0.108***	0.158***
	(0.0195)	(0.0201)	(0.0168)	(0.0151)	(0.0279)	(0.0533)
Log(proceeds)	-0.136***	0.00465	-0.0271	-0.0894***	-0.207***	-0.364***
	(0.0354)	(0.0354)	(0.0296)	(0.0268)	(0.0495)	(0.0945)
Lagged ROE	0.0190*	-5.98e-05	0.00422	0.0126	0.0285*	0.0496*
	(0.00967)	(0.0109)	(0.00913)	(0.00816)	(0.0151)	(0.0289)
Constant	0.504*					
	(0.263)					
Observations	1,830	1,832	1,832	1,832	1,832	1,832
R-squared	0.128					

2.3.3 Calendar Time Portfolio

In this section, we use the calendar time portfolio methodology to analyze the relationship between negative sentiments in S-1 and long-term returns of IPOs by estimating the returns of portfolios of IPO firms sorted by sentiments in S-1 documents. Specifically, we evaluate the difference in holding period returns between portfolios of IPO stocks with negative sentiment at the top quintile and portfolios of IPO stocks with negative sentiment at the bottom quintile.

At the beginning of each calendar month, all IPOs listed in the past 36 months are grouped into quintiles based on their negative sentiment measures, where the bottom quintile represents IPOs with the lowest sentiment measure and the top quintile represents IPOs with the highest sentiment measure. Consistent with Brav and Gompers (1997), the portfolio is equal weighted. Then the portfolios are held for 1 month, 3 months, 6 months, and 12 months. Therefore, for a holding period of n months, a fraction of $1/n$ of the portfolio is rebalanced every month. We adjust the returns with Fama and French (1993) 3-factor model augmented by the momentum factor from Carhart (1997). Table

2.7 reports the estimated intercepts from a four-factor model, including market factor, size factor, value factor and momentum factor. Consistent with Field and Lowry (2009), the regression is weighted by IPO numbers in the portfolio.

Table 2.7: Return Differentials for Negative Sentiment Portfolios

This table reports abnormal returns for portfolios of IPOs with negative sentiment at the top quintile, at the bottom quintile, and a strategy to long the top quintile and short the bottom quintile, in the past 36-month or 60-month periods. The portfolios are equal weighted, and the regressions are weighted by the number of IPOs in the portfolio. Holding periods are 1-12 months. Alphas estimated from the four-factor model are shown in the table. The four-factor model includes the three Fama and French (1993) factors, and the Carhart (1997) momentum factor. Estimates are reported for the holding period net return. Robust standard errors are in parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	Holding Period				
	1m	3m	6m	9m	12m
Panel A: 36-month IPOs					
Bottom Quintile	-0.00243	-0.00719	-0.0137**	-0.00868	-0.00880
	(0.00321)	(0.00460)	(0.00586)	(0.00753)	(0.00844)
Top Quintile	0.00669	0.0102	0.0129	0.0246*	0.0351*
	(0.00532)	(0.00820)	(0.0114)	(0.0148)	(0.0178)
Top - Bottom	0.00908**	0.0170**	0.0262***	0.0330***	0.0428***
	(0.00442)	(0.00705)	(0.00968)	(0.0122)	(0.0163)
Panel B: 60-month IPOs					
Bottom Quintile	0.000176	0.00179	0.00357	0.000120	-0.00669
	(0.00208)	(0.00364)	(0.00573)	(0.00663)	(0.00758)
Top Quintile	0.00852**	0.0140**	0.0199**	0.0176	0.0130
	(0.00417)	(0.00654)	(0.00986)	(0.0120)	(0.0144)
Top - Bottom	0.00829**	0.0119**	0.0161**	0.0170*	0.0189*
	(0.00386)	(0.00530)	(0.00754)	(0.00875)	(0.0106)

The results show that a strategy that buys stocks with the highest negative sentiment when it went public within 3 years and sells stocks with the lowest negative sentiment when it went public within 3 years yields an abnormal return between 0.9% and 4.28% for holding periods of 1 month to 12 months. The abnormal returns for the bottom quintile are negative even though not significant except for a holding period of 6 months. The abnormal returns for the top quintile are positive and significant for a holding period of 9 months or 12 months. When sorting IPOs based on a 5-year period, IPOs in the bottom and top quintile have better performance than those for a 3-year period, which could be explained by survivalship bias. However, the long-short portfolio still gets significant positive returns ranging from 0.83% to 1.89% depending on the holding period. The positive return differentials between the top and bottom quintile portfolios are mostly due to the significant positive returns of IPO stocks at the top quintile of negative sentiment. Therefore, there are missing factors in the 4-factor model that could explain the return differentials.

As argued by Mitchell and Stafford (2000), although the calendar time portfolio approach solves the dependence issue, it has several potential problems: constant factor loading over time, heteroskadasticity due to changes on portfolio composition, hard to detect abnormal performance. To solve these problems, I exploit the bootstrap procedure suggested by Mitchell and Stafford (2000). The bootstrapping sample is 500 observations generated from the original sample with replacement. Table A.1 reports the abnormal returns for top quintiles, bottom quintiles and the long-short strategies estimated from the four-factor model. Similar as the results shown in Table 2.7, for the IPOs listed within

36 months, the bottom quintile portfolio generates negative abnormal returns, while the top quintile portfolio has positive returns (although not significant) for a holding period up to 12 months. The strategy to short bottom quintile and long top quintile generates positive abnormal returns, and the results are statistically significant for a holding period of 6, 9 and 12 months. The economic magnitude is less than that without bootstrapping. For example, for a holding period of 12 months, the abnormal return is 3.22%, which is about 25%[12] less than that without bootstrapping. Similar patterns exist for IPOs listed within 60 months: better performance for both bottom and top quintile portfolios, and positive return differentials for the long-short portfolio. The return differentials range from 0.71% to 1.31%, less than those in Table 2.7. These results imply that a strategy based on the negative sentiment of prospectus yields positive abnormal returns. Moreover, it is more difficult to detect positive abnormal performance with calendar time portfolio methodology.

2.4 Explanation

Results in Section 2.3 indicate that the negative sentiment in S-1 filings explains the long-run return differentials of IPOs. This section explores the channel through which the sentiment affects the returns. We try to explain the correlation with two players of the IPO game: managers and investors.

2.4.1 Managers

From the classic finance theory, returns are compensation for risk taking. Thus IPOs with better performance may be a result of high risk taking unanticipated at IPO. We measure the misexpectation of risk with idiosyncratic volatility. Specifically, for each IPO firm, we run Carhart (1997) 4-factor model with daily returns for 3 years after IPO. If a firm delists within 3 years, we only use the daily returns until the delisting date.

$$R_{it} - R_{ft} = \alpha + \beta_1 RMRF_{it} + \beta_2 SMB_{it} + \beta_3 HML_{it} + \beta_4 MOM_{it} + \epsilon_{it} \tag{2.2}$$

$$ivol_i = variance(\epsilon_{it}) \tag{2.3}$$

where R_f denotes the risk free rate, and the four factors at the right hand side of equation 2.2 are market factor (RMRF), size factor (return on a portfolio of short big firms and long small firms, SMB), value factor (return on a portfolio of short growth firms and long value firms, HML), momentum factor (return on a portfolio of short past losers and long past winners). Then take the volatility of residuals from equation 2.2 to get idiosyncratic volatility.

We sort IPOs by negative sentiment, and Panel A of Table 2.8 shows the average level of idiosyncratic volatility for each quintile. We measure the idiosyncratic volatility for 3 years, 5 years, and also exclude the first quarter to avoid unusual trading activities (Loughran and McDonald (2013)). Idiosyncratic volatility increases with negative sentiment in S-1 document for various periods. Table 2.9 presents the results of OLS regression with (log) idiosyncratic volatility as the dependent variable. Results in column (2) and (4) suggest that when additional controls are included in the regression, the negative sentiment is no longer significantly correlated with 3-year idiosyncratic volatility. Thus the

[12] (3.22% - 4.28%)/4.28%.

positive link between negative sentiment in S-1 and long-term wealth relatives is not a reflection of higher risk taking by the firm.

Table 2.8: Idiosyncratic Volatility and Overconfidence Sorted by Negative Sentiment

This table reports sorting results for idiosyncratic volatility and overconfidence based on negative sentiment in S-1 filings. Idiosyncratic volatility is calculated from the four-factor model in Carhart (1997): $R_{it} - R_{ft} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + \epsilon_{it}$, $ivol_i = variance(\epsilon_{it})$. Panel A shows average idiosyncratic volatility for a 3-year, 5-year, 3-year excluding first quarter after IPO, 5-year excluding first quarter after IPO period. Overconfidence is measured as *Holder 67*, defined as CEO-years of CEOs who had options more than 67% in-the-money during their tenure. Panel B shows the average frequency of *Holder67>=2*, *Holder67>=3*, *Holder67>=4*. *Negative* takes proportions of negative word lists in S-1 document, then standardized by the negative ratio for all IPOs listed within 365 days before the filing date. *N* displays the sample size in each quintile.

Panel A: Idiosyncratic volatility

Q	N	Negative	3-year	5-year	3-year ex. first quarter	5-year ex. first quarter
1	449	-1.196	0.0431	0.0437	0.0431	0.0439
2	439	-0.440	0.0487	0.0487	0.0488	0.0488
3	439	0.0847	0.0508	0.0505	0.0505	0.0503
4	439	0.612	0.0545	0.0542	0.0545	0.0543
5	432	1.539	0.0563	0.0562	0.0562	0.0561

Panel A: Overconfidence

Q	N	Negative	*Holder67>=2*	*Holder67>=3*	*Holder67>=4*
1	77	-1.260	0.636	0.481	0.325
2	68	-0.517	0.706	0.471	0.294
3	67	-0.0144	0.716	0.552	0.433
4	68	0.488	0.691	0.515	0.382
5	62	1.493	0.742	0.565	0.419

Research shows that executives appear to be particularly inclined to display overconfidence, and overestimation of probability of positive states could trigger overconfidence (Malmendier and Tate (2005), Gervais, Heaton, and Odean (2011)). Thus it is possible that overconfident CEOs present less negative tone in their IPO filings, and later generate lower returns. We measure overconfidence of CEOs incumbent during IPO with *Holder 67*, which is defined as all CEO-years of CEOs who had options more than 67% in-the-money during their tenure (Malmendier and Tate (2005)). We match the overconfidence data with our negative sentiment measure and sort the negative sentiment into quintiles. Then we count the frequency of *Holder 67*. Panel B of Table 2.8 displays the sorting results of averages for negative sentiment, $Holder\ 67 >= 2$, $Holder\ 67 >= 3$, $Holder\ 67 >= 4$. Out of our expectation, CEOs in IPOs with less negative sentiment do not show more overconfidence, but the bottom quintile show higher overconfidence than the second quintile for $Holder\ 67 >= 3$, $Holder\ 67 >= 4$. Although the top quintiles of negative sentiment show more overconfidence, the middle quintiles also have high values. There is no clear trend with overconfidence and negative sentiment. The merged dataset has much less observations, thus this analysis is only suggestive. More detailed exploration is for future research.

2.4.2 Investors

Brav and Gompers (1997) conjectures that retail investor sentiment could explain the underperformance of small, low book-market firms. Field and Lowry (2009) shows evidence that by better interpreting public information, institutional investors invest in IPOs with better performance. Retail investors are prone to sentiments revealed in the prospectus. Therefore, retail investors buy IPOs with less negative sentiments or more positive tones,

Table 2.9: Idiosyncratic Volatility and Negative Sentiment in S-1

This table reports results for regression $ivol_i = \alpha + \beta Negative_i + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i$. The sample includes IPOs from 1997 to 2016. $ivol_i$ measures IPOs' 3-year idiosyncratic volatility calculated from the four-factor model in Carhart (1997): $R_{it} - R_{ft} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + \epsilon_{it}$, $ivol_i = variance(\epsilon_{it})$. $Negative_i$ takes proportions of negative word lists in S-1 document, then standardized by the negative ratio for all IPOs listed within 365 days before the filing date. $\mathbf{X_i}$ includes various control variables described in Table 2.2. The dependent variable in column (1) and (2) is $ivol$, and the dependent variable in column (3) and (4) is $log(ivol)$. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) ivol	(2) ivol	(3) log(ivol)	(4) log(ivol)
Negative	0.00114***	0.000147	0.0574***	0.00630
	(0.000408)	(0.000361)	(0.0180)	(0.0152)
Lead UW's rank	-0.00347***	-0.00116***	-0.141***	-0.0329***
	(0.000417)	(0.000249)	(0.0147)	(0.00838)
VC dummy	0.000982	9.13e-06	0.0808***	0.0214
	(0.000593)	(0.000693)	(0.0260)	(0.0237)
Internet dummy	0.00877***	0.00933***	0.269***	0.273***
	(0.00140)	(0.00132)	(0.0435)	(0.0396)
Log(age)	-0.00513***	-0.00301***	-0.234***	-0.134***
	(0.000818)	(0.000717)	(0.0268)	(0.0219)
First-day return		-0.00198***		-0.0243*
		(0.000369)		(0.0140)
Lagged Log(assets)		-0.00261***		-0.116***
		(0.000410)		(0.0147)
Log(proceeds)		-0.00239**		-0.139***
		(0.00105)		(0.0360)
Lagged ROE		-0.000275*		-0.0101**
		(0.000159)		(0.00488)
Constant	0.0878***	0.0872***	-4.658***	-4.599***
	(0.00361)	(0.00514)	(0.134)	(0.168)
Observations	2,115	2,001	2,115	2,001
R-squared	0.578	0.603	0.654	0.695

and in the long run realize lower returns. We test this proposition in two ways: analyzing the most significant part of S-1 filings that generates positive long run performance; exploiting the attention based explanation with a measure of retail investor attention.

We extract each section from S-1 documents, and parse each section according to section 2.2.2. We obtain the standardized negative sentiment SF_Neg for each section. We only consider 4 sections in S-1: Summary, Risk factors, Management Discussion and Analysis (MD&A) and Use of Proceeds. *Summary* section is a brief introduction of the prospectus, normally sketching the above listed 3 sections: Risk factors, Management Discussion and Analysis (MD&A) and Use of Proceeds. *Risk factors* section describes all the risks the firm bear now or in the future in detail. *Management Discussion and Analysis (MD&A)* part discusses the financial conditions, business operations, and analyzes their financial statements in detail. *Use of Proceed* section briefly illustrates the offer price, the usage of proceeds. A typical statement in *Use of Proceeds* is "We intend to use the net proceeds to us from this offering, plus cash on hand, to repay indebtedness and to pay fees and expenses."

Table 2.10 reports the regression results where the dependent variable is the 3-year wealth relative benchmarked by the value-weighted return of Fama-French 48 industry portfolio. Column (1) - (4) run regressions for each section's negative sentiment, and column (5) includes them in one regression. The negative sentiment in *Summary* is significantly related to long-run overperformance. This implies that retail investors may only read the summary part which is at the beginning of the prospectus, and be influenced by the sentiment shown there. Then retail investors buy the stocks with less negative words, which realize underperformance in the long run. This suggestive evidence supports the investor sentiment story in affecting long-run returns.

Table 2.10: Long-Term Performance and Negative Sentiment in Each Section of S-1

This table reports results for regression $WR_i = \alpha + \beta Negative_i + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i$. The sample includes IPOs from 1997 to 2016. WR_i measures IPOs' long-term performance, $WR_i = BHR_i/BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1+r_{i,t})$, where Fama-French 48 industry value-weighted return is used as the benchmark, T takes the value of 36 months. $Negative_i$ takes proportions of negative word lists in the following sections of S-1 filings: summary, risk factor, Management Discussion and Analysis (MD&A), use of proceeds, then standardized by the respective ratio 365 days before the filing date $SF_{i,j,t} = \frac{ratio_{i,j,t} - \text{average ratio}_{j,t-1}}{\sigma_{j,t-1}}$. $\mathbf{X_i}$ includes various control variables described in Table 2.2. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Summary	(2) Risk factor	(3) MD&A	(4) Proceed	(5) All
Negative Summary	0.0471**				0.0483**
	(0.0230)				(0.0221)
Negative Risk		0.0126			0.00851
		(0.0293)			(0.0317)
Negative MD&A			0.00320		-0.00279
			(0.0119)		(0.0176)
Negative Proceed				-0.0310**	-0.0374***
				(0.0127)	(0.0131)
Days between S-1 and IPO	-0.0292	-0.0266	-0.0258	-0.0305	-0.0330
	(0.0371)	(0.0403)	(0.0418)	(0.0399)	(0.0418)
Prior NASDAQ 30-day return	-16.21**	-18.04**	-12.84	-16.01**	-16.25**
	(6.730)	(8.441)	(8.605)	(7.402)	(6.586)
Up revision	0.0264	0.0322	0.0210	0.0202	-0.00407
	(0.0528)	(0.0624)	(0.0675)	(0.0550)	(0.0588)
Lead UW's $ market share	-0.0185	-0.131	-0.0440	-0.0971	-0.132
	(0.341)	(0.412)	(0.396)	(0.418)	(0.404)
Lead UW's rank	0.0847***	0.0860**	0.0807**	0.0779**	0.0817**
	(0.0276)	(0.0322)	(0.0329)	(0.0344)	(0.0317)
VC dummy	-0.0876**	-0.0616	-0.0548	-0.0687*	-0.0662*
	(0.0343)	(0.0428)	(0.0379)	(0.0373)	(0.0382)
Internet dummy	-0.0148	-0.0584	-0.0698	-0.0548	-0.0657
	(0.0532)	(0.0625)	(0.0612)	(0.0644)	(0.0531)
Log(age)	0.0611**	0.0730***	0.0745***	0.0608**	0.0635**
	(0.0275)	(0.0250)	(0.0249)	(0.0244)	(0.0250)
First-day return	-0.126***	-0.137***	-0.140***	-0.135***	-0.133***
	(0.0366)	(0.0350)	(0.0360)	(0.0307)	(0.0309)
Lagged Log(assets)	0.0931***	0.0837***	0.0833***	0.0882***	0.0800***
	(0.0250)	(0.0236)	(0.0237)	(0.0245)	(0.0231)
Log(proceeds)	-0.143***	-0.137***	-0.139***	-0.137***	-0.122***
	(0.0397)	(0.0381)	(0.0366)	(0.0372)	(0.0367)
Lagged ROE	0.0211	0.0188*	0.0194**	0.0174	0.0187*
	(0.0130)	(0.00974)	(0.00917)	(0.0104)	(0.0100)
Constant	0.460	0.459	0.494*	0.552*	0.488
	(0.275)	(0.297)	(0.287)	(0.309)	(0.311)
Observations	1,717	1,736	1,715	1,730	1,673
R-squared	0.153	0.124	0.128	0.126	0.132

Da, Engelberg, and Gao (2011) argues that there is a positive link between retail investor attention and retail investor sentiment. By measuring retail investor attention with the Google search volume index from Google Trends, they show that IPO stocks with high investor attention prior to IPO have larger first-day returns and larger return reversals in one year for IPOs from 2004 to 2007. The reason is that retail investors' attention prior to IPO induces greater buying pressure which drives up the first-day return of IPO stocks, and the dissipation of the buying pressure generates long-run underperformance. With our stock-specific sentiment measure, we explore the relation of retail investor attention and sentiment, and whether the demonstrated effect of negative sentiment on long-term return is via retail investor attention.

Since the weekly search volume index provided by Google Trends starts from the year 2004, we focus on a subsample from 2004 to 2016. For each IPO, we start a web crawling program by searching the suggestions of the IPO name, then find the most proximate suggestion term and use this term to get the search volume index (SVI) for the IPO.[13] For each IPO, we get its weekly search volume index for 4 quarters before and after the IPO date.[14] Following Da, Engelberg, and Gao (2011), we calculate $ASVI$(abnormal search volume index) as a measure of retail investor attention, which is defined as the log of SVI during the week minus the log of average SVI during the previous 8 weeks.[15] For the regression, we use $ASVI$ one week before IPO as a proxy for retail investor attention. Finally, we get 574 IPOs with sufficient data for analysis.

Figure A.1 plots the mean and median of SVI around the IPO week (± 15 weeks). Figure A.2 plots the pattern for ASVI. Retail investor attention proxied by SVI stays stable at a low level until three weeks before the IPO. The index roars from three weeks before the IPO and reaches its maximum at the IPO week, then it declines quickly 1 week after the IPO. From 3 weeks after the IPO, the index has little change, but the average stable value is a little higher than that before IPO.

Table 2.11 demonstrates the regression results with retail investor attention as the mediator between negative sentiments and long-run returns for IPOs from 2004 to 2016. Column (1) shows estimated results of equation 2.1 in the subsample 2004 to 2016. The estimated coefficient on negative sentiment is 0.0670, positively significant at 5% level, which is comparable to the estimation in Table 2.3 for the full sample. In column (2), we replace the negative sentiment with $ASVI$. The negatively significant estimated coefficient implies that stocks that receive higher attention from retail investors prior to IPO experience severe long-term underperformance. In column (3), we analyze the link between negative sentiment and retail investor attention. The results show that IPOs which display more negative tones in their prospectuses receive less attention from retail investors. In column (4), we include $ASVI$ to equation 2.1 to explore whether retail investor attention is the mediator of the effect of negative sentiments on long-term returns of IPOs. The estimated coefficient on negative sentiment is still positive, but not significant. However, the coefficient on $ASVI$ is negative and significant at 5% level. These results suggest that negative sentiment in the prospectus is associated with better long-term performance because these IPOs obtain less retail investor attention and thus less buying pressure during IPO, and as more information is absorbed by the market, they

[13] As noted in Da, Engelberg, and Gao (2011), ticker is a more accurate search term as company names have multiple meanings (eg. "Amazon"). IPOs don't have tickers, however, so we can only use company names as search terms. Moreover, investors may search the same firm with variations of company names. A good feature of Google Trends is that it makes suggestions on the search term. For example, when inputting "Anadys Pharmaceuticals Inc" into the search box, Google Trends show the suggestion of "Anadys Pharmaceuticals" and label it as "Company" with search term "/m/02sg749". Normally searching the "company" gives more accurate data than searching the full company name as a "search term".

[14] To cite Google Trends: *Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term.* Therefore, for each firm, we get its worldwide popularity relative to the peak point for the ± 4 quarters around the quarter of IPO.

[15] Therefore, stocks listed in the first 8 weeks of 2004 are dropped. We don't use median SVI in Da, Engelberg, and Gao (2011) because Google Trends make SVI to 0 when the search volume is relatively small (< 1) compared to the peak point and thus many values are 0. For example, for the first 5 weeks, SVI is 0; for the remaining 3 weeks, SVI>0. Then $ASVI$ calculated based on median SVI is just SVI of the week and can't reflect the trend of search volume.

Table 2.11: Long-Term Performance and Retail Investor Attention

This table reports results for regression $Y_i = \alpha + \beta Negative_i(ASVI_i) + \delta \mathbf{X_i} + \text{Year FE} + \text{FF48 FE} + \epsilon_i$. The sample includes IPOs from week 8 of 2004 to 2016. For column (1), (2), (4), the dependent variable is WR_i, which measures IPOs' long-term performance, $WR_i = BHR_i/BHR_{benchmark}$, and $BHR_i = \prod_{t=1}^{T}(1+r_{i,t})$, where Fama-French 48 industry value-weighted return is used as the benchmark, T takes the value of 36 months. $Negative_i$ takes proportions of negative word lists in S-1 filings, then standardized by the respective ratio 365 days before the filing date $SF_{i,j,t} = \frac{ratio_{i,j,t} - \text{average ratio}_{j,t-1}}{\sigma_{j,t-1}}$. For column (3), the dependent variable is $ASVI_i$, the abnormal search volume one week before IPO, defined as the log of search volume during the week minus the log of average search volume during the previous 8 weeks. $\mathbf{X_i}$ includes various control variables described in Table 2.2. Robust standard errors clustered by Fama-French 48 industry and quarter are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) WR	(2) WR	(3) ASVI	(4) WR
Negative	0.0670**		-0.0404*	0.0294
	(0.0285)		(0.0234)	(0.0326)
ASVI		-0.180**		-0.176**
		(0.0817)		(0.0815)
Days between S-1 and IPO	-0.0659	-0.0615	-0.0382*	-0.0602
	(0.0425)	(0.0444)	(0.0207)	(0.0447)
Prior NASDAQ 30-day return	-18.02	-32.90	-4.957	-31.99
	(18.79)	(25.00)	(13.83)	(25.60)
Up revision	0.0332	0.0523	-0.0310	0.0521
	(0.0766)	(0.101)	(0.0552)	(0.102)
Lead UW's $ market share	-0.649	-0.571	0.278	-0.559
	(0.394)	(0.479)	(0.357)	(0.469)
Lead UW's rank	0.108**	0.126**	0.00345	0.125***
	(0.0400)	(0.0462)	(0.0336)	(0.0449)
VC dummy	-0.0345	-0.0220	0.0605	-0.0237
	(0.0821)	(0.0668)	(0.0498)	(0.0659)
Internet dummy	-0.0470	-0.0271	0.0437	-0.0310
	(0.0684)	(0.0645)	(0.0398)	(0.0631)
Log(age)	0.0983**	0.0592	-0.0268	0.0585
	(0.0463)	(0.0664)	(0.0210)	(0.0651)
First-day return	-0.0774	-0.0107		-0.0127
	(0.113)	(0.120)		(0.123)
Lagged Log(assets)	0.0499	0.0928**	5.32e-05	0.0930**
	(0.0487)	(0.0354)	(0.0167)	(0.0351)
Log(proceeds)	-0.0497	-0.112*	-0.00728	-0.106
	(0.0623)	(0.0648)	(0.0320)	(0.0685)
Lagged ROE	0.0277**	0.0339***	0.00886	0.0340***
	(0.0133)	(0.00809)	(0.00943)	(0.00827)
Constant	0.228	0.245	0.245	0.214
	(0.348)	(0.530)	(0.221)	(0.551)
Observations	922	574	574	574
R-squared	0.098	0.132	0.129	0.133

achieve higher returns[16].

2.5 Conclusion

Textual analysis in financial statements focuses on the market response in a short period. In this paper, we explore the relationship between sentiment revealed in S-1 filings and IPOs' long-run performance. With various measures on sentiments and a variety of long-term returns of IPOs, we find consistently positive links between negative sentiments and long-term returns. We also exploit quantile regression and calendar time portfolio methodologies to overcome the problems OLS may have, and the results still support the positive link. We try to explain the relationship by analyzing managers and investors

[16]The results still hold if we add institutional ownership before IPO as an explanatory variable. We drop *First-day return* in column (3) because $ASVI$ is defined one-week before IPO.

in the IPO process. Firms' taking more risk, or CEO overconfidence cannot explain our result. Our findings on sections in S-1 filings and search volume index from Google Trends imply that retail investors who are more subject to behavioral bias drive the positive link between the negative sentiment and long-run returns. Future research is needed for more convincing explanation about this link, especially for CEO overconfidence.

Chapter 3

Female Networks and Corporate Resource Allocation in China

Abstract

I present evidence that same-gender linkages between female corporate executives and female political leaders influence the location of production decisions for Chinese exports. The effect only exists in financially constrained firms. This effect is more significant if the firm is younger, smaller, more concentrated, and faces fiercer local competition. These results suggest that the most reasonable explanation is that female connection mitigates information friction.

Key words: Executives; Gender; Exports; Resource Allocation
JEL codes: G30; G34; J16; F14

3.1 Introduction

How do corporations make decisions about internal resource allocation? Literature shows that plant-level proximity to headquarters, social connections between the CEO and divisional managers (Giroud (2013), Duchin and Sosyura (2013)) could be possible factors. Would a factor as simple as gender commonality affect the internal resource allocation? The manifestation of female solidarity is women's inherent co-operativeness with each other (Cornwall (2007)). Political connections are important determinants of corporate outcomes, especially in China (Ding, Fan, and Lin (2018), Schoenherr (2019)). In this paper, I explore the impact of linkages between female corporate executives and local political leaders through gender commonality on decisions of production locations for export in the context of Chinese listed firms.

Females are minorities in both the business and political world. Fisman, Paravisini, and Vig (2017) presents results suggesting that the smaller the size of the group, the stronger the effect of proximity on loan outcomes. Moreover, Tate and Yang (2015) argues that having women leadership cultivates more female-friendly cultures. Thus female political leaders may prefer firms with women in the management team because of the recognition over each other or the pro-women environment the firm has. Therefore, my hypothesis is that corporations with female executives produce more in locations of which women are in charge as political leaders.

Instead of using plant-level data which is unavailable for China, I exploit the export data from China's General Administration of Customs to identify the production locations for Chinese listed firms from 2000 to 2006. This dataset provides all export transactions from China to other countries. It contains information about the export entity, location of production, quantity, value, product classification, unit, and other variables for each transaction. The prefecture city of production is determined by the location of production. I aggregate the transaction value into firm-prefecture city-month level to generate two variables for analysis: the fraction of export value, export value growth.

I define a firm-city-month as female connected if there are women in the top management team or board at the headquarter, and at the same time either the mayor or the secretary of the city (or both) in which the firm produces is female. The exogenous variation mainly comes from the rotation of local political leaders as most firms have female executives. The mayor is elected by Municipal People's Congress, while the secretary is assigned by the Provincial Party Committee[1]. In both cases, the rotation of municipal political leaders should not be affected by firms' preference. Moreover, gender is determined at the time of birth for corporate executives and political leaders. Therefore, the connection between the firm and city from gender commonality is plausibly exogenous to corporate decisions. With the firm-city-month level treatment variable, various fixed effects - firm-city fixed effect, city-month fixed effect, firm-month fixed effect - could be included to account for unobserved time-varying shocks to any firms or cities, or variations from supply or demand side for any firm-city pairs.

The empirical analysis contains three parts. For the first step, I analyze whether firms export more value in cities with female connection. The results show strong evidence of preferential female treatment. In the baseline results, I find that on average, the fraction of export value increases by 3% for cities with female connection. Having female connection also increases the absolute export value, and the growth on export value over last period. By adding another dimension of product classification into the analysis, I find that the increase in export value is from the rise of quantity, and there is no effect on the unit price. This female connection may be a reflection of links from education, or shared native or birth places as shown in Duchin and Sosyura (2013), Shen et al. (2019). Most of the data is missing for education and native/birth places of executives, although it is complete for political leaders. I construct the connection measures for education and native/birth places based on this incomplete data. Results show that neither connections through education or native/birth places have an effect. Moreover, the inclusion of the two connections does not affect statistical or economic significance of female connectedness. Consistent with Giroud and Mueller (2015), the effect of female connection is only significant for financially constrained firms.

For the second part, I analyze the possible source of this preference. It may come from favoritism or reduction in asymmetric information. Favoritism leads to misallocation of resources because favoring one over the other sacrifices efficiency. Conversely, a transaction creates value for both parties if preferences enable one party to learn more about the other (Fisman, Paravisini, and Vig (2017)). The results on ex-post performance and het-

[1]Administrative divisions in China: the highest level is provincial level, including provinces, autonomous regions, municipalities directly under the Central Government; the second level is city level, including prefecture city, autonomous prefecture (city level); the third level is county level, including county, autonomous county.

erogeneity analyses support the explanation of mitigating information friction, although the explanation of favoritism can not be ruled out. Firms with female connections have better ex-post performance which is measured by return on equity four quarters later at the firm level, and this positive effect remains even if the connection is not existing in four quarters.

At last, several heterogeneity analyses are conducted to have more insights on the effect as well as explanation. I find that connections with the secretary drive the effect on the fraction of export value while connections with the mayor experience a higher value growth. From the firm side, both CEOs and other executives' connections contribute to higher export value and growth. Then I try to find more evidence to support the information-based explanation by examining whether the effect of female connection changes with: internal concentration of production at the firm level, local competition the firm faces at the city level, firm size, firm age. The logic is that if favoritism outweighs information explanation, the effect of female connection should be more significant for firms which are more diversified in production, living in a more monopolistic market, larger, more mature, as these firms are essential for growth of local GDP, and thus suffering less asymmetric information problems. The estimated results suggest the opposite for the fraction of export values. The effect has a stronger impact on more concentrated firms, firms confronting more competitive environment, smaller and younger firms. However, my results on the value growth are significantly positive for the most diverse firms, largest firms, implying that should these firms have connection, they increase production in each city, causing an increase of value growth and no effect on the fraction of value in the connected city. This is an evidence for the positive externality of female connection, and also for the favoritism-based explanation.

This paper contributes to three strands of literature. First, the paper is related to research on connections, networks and social ties. Duchin and Sosyura (2013) looks at the link from education, previous employment, and nonprofit organizations between divisional managers and CEOs by hand-collecting data of S&P 500 firms . As emphasized in Becker (1996), these kinds of connections are dynamic, and subject to individual choice, which makes the empirical specification be trapped into an edogeneity problem. Endowments assigned at birth, however, avoid this problem. Some use shared birth or native places, shared language between two parties as the measure of connectedness (Shen et al. (2019)). However, too much information about native or birth places, education, language and others mentioned above is missing for executives during 2000-2006 in the context of China. Fisman, Paravisini, and Vig (2017) uses caste and religion between lenders and borrowers to define connections with Indian loan-level data. There is no caste in China, and most political leaders as well as ordinary people do not have any religious belief. Gender is determined at birth and no one changes the gender during the sample period. So identification based on gender is plausibly exogenous in China. This paper shows that a connection based on the simple measure of gender has an effect on internal resource allocation.

Second, there is a growing literature concerning the role that women play in finance. Adams and Ferreira (2009) finds that female directors have better attendance records than male directors, and women are more likely to join monitoring committee. Huang and Kisgen (2013) shows that female executives exhibit less overconfidence in corporate decision making compared to men. By comparing wage losses of workers from the same closing plant to the same new firm, Tate and Yang (2015) finds that women experience more wage losses than men, but this gap is significantly smaller if the new firm is with female leadership. They come to the conclusion that having women leadership cultivates more female-friendly cultures inside the firm. Stolper and Walter (2019) presents that the sameness on marital and parental status for female advisees increases the likelihood of following financial advice. And generally, they show that homophily - individuals' affinity for others like them - has a significant effect on their financial decisions. This paper complements the literature by showing the female executives affect internal decisions by reducing information friction due to their political connection based on gender. This creates value for firms.

Finally, my study is related to literature on internal market and soft information. Giroud (2013) finds that a reduction of travel time between headquarters and plants increases plant-level investment and efficiency because of better monitoring and gathering soft information. Duchin and Sosyura (2013) shows evidence that divisional managers with social connections with the CEO receive more capital, but the efficiency depends on the tradeoff between agency problem and information asymmetry. My identification with the exogenous rotation of prefecture city leaders could pin down the effect of soft information than the proxies which may suffer edogeneity problems.

The remainder of this paper is organized as follows. Section 3.2 presents the data. Section 3.3 describes the empirical specification. Section 3.4 shows the main empirical results. Section 3.5 distinguishes the explanations. Section 3.6 provides several heterogeneity analyses. Section 3.7 concludes the paper.

3.2 Data

My analysis of female connection on internal resource allocation relies on three data sources: transaction-level export data, figure information of executives of listed firms, and figure information for municipal political leaders. The data of plant information of each firm is unavailable in Chinese context. I use the transaction-level export data from China's General Administration of Customs to obtain information of cities in which firms have production. Not every firm produces for export, so the analysis based on this data is trying to understand how firms allocate internal resources for export. The figure information for political leaders is elaborate. Since unlisted firms do not disclose their executives information to public, my focus is on listed firms. The data of figure information is from China Stock Market & Accounting Research Database (CSMAR) and Chinese Research Data Services Platform (CNRDS), which are frequently used to explore China's corporate finance problems.

The transaction-level export data from China's General Administration of Customs covers all trade transactions from February 2000 to December 2006. Each transaction includes information about export entity, product classification (HS 8-digit code), unit, quantity, value, destination country, transportation method (flight, freight), port, custom's regime ("Processing and Assembling", "Processing with Imported Materials"), location the product is export from, and shipping route. Since I am interested in the export value fraction or growth for the connected and unconnected companies, I aggregate the trade data into firm-city-month level with the following process.

First, I extract the official city name from location data which reflects the producing place. Second, I match the listed firm names, their used names and subsidiary parts with the company name in exxport data to generate listed firm name for each transaction. Third, I aggregate the values and quantities to listed firm-city-month-product level. With this data, the unit value is computed by dividing values by physical quantities. In the main analysis, the data is further aggregated to firm-city-month level where only value is available as comparing quantities and unit prices for different products is meaningless. For extra tests, the export destination variation is calculated as the number of countries the listed firm exports to for a given city-month, and the product variation is defined as the number of different product classifications for export in firm-city-month level. We require each listed firm to have at least 2 cities for production in each month.

China Stock Market & Accounting Research Database (CSMAR) provides demographic information for executives of listed firms, and political leaders for each prefecture-level city as well as province-level municipality[2]. I complete this dataset with Chinese Research Data Services Platform (CNRDS). The political leader data includes name, position, age, gender, tenure, native/birth place, educational background. Besides this variables, the executive data also provides information about the appointment in other listed firms, salary and some governance measures. However, except position, age and gender, the executive data contains a large proportion of missing inputs, which makes analysis based on native preference, education commonality, or shared dialect unreliable. Political leaders contain the mayor and secretary for each prefecture city. Executives include presidents, managers, board directors, supervisors and other executives shown in the annual reports. Out of 1503 listed firm, there are 1449 listed firms have female executives during the sample. In firm-month level, 83.28% observations have female executives. For the 332 cities, females work in only 57 cities. And female political leaders only account for 5.76% of the city-month data. The variable of interest $Female$ is constructed as 1 if one or both of the political leaders are female, and the firm has at least one female as executives in a particular month. Due to the scarcity of women in the political world, this measure is mainly driven by the rotation of political leaders, which is determined by the top leaders in China and thus exogenous to firm or city-level characteristics.

At the end, we have 68,725 firm-city-month observations for 543 listed firms at 194 cities from February 2000 to December 2006. 3,858 observations have female political leaders and female executives in the firm-city-month level. Table 3.1 reports the descriptive statistics for various summary level. On average, listed firms export 11.8 (median is 2) products with value 1.4 million CNY (median is 0.094 million CNY) in each city-month level, to about 6 (median is 2) destinations. Each firm has on average 5.5 cities for production and the median is 3 cities. Average tenure of a political leader is 28.65 months during 2000 and 2006. There are 482 rotations of secretaries, 553 rotations of mayors during the sample period.

3.3 Empirical Specification

The baseline empirical specification identifies the effect of female leader-executive links on export outcomes. The specification takes the following form:

$$Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict} \tag{3.1}$$

The dependent variable is the export outcomes for a listed firm i, at city c, in month t. Using the trade data, I construct the following variables: $Fraction\ of\ Export\ Value_{ict}$, the proportion of export value of firm i in month t at city c relative to the overall export

[2]For simplicity, I use city, prefecture city interchangeably, and it includes prefecture-level cities and province-level municipalities.

Table 3.1: Summary Statistics

This table presents the summary statistics for: firm-city-month level, export values in thousand, different number of products for export, and number of destinations export to; firm-month level, number of cities in which a firm has production; city-person-position level, the tenure of each political leader at a position in a city.

	N	mean	sd	p5	p50	p95
Firm-city-month level						
Export value, in thousand	68,725	1,419	5,617	4.211	93.85	6,726
Number of products	68,725	11.80	39.79	1	2	47
Number of destination	68,725	5.878	10.92	1	2	28
Firm-month level						
Number of cities	13,537	5.485	6.079	2	3	19
City-person-position level						
Tenure of political leader	1,676	28.65	18.09	3	26	62

value of firm i in month t,

$$Fraction\ of\ Export\ Value_{ict} = \frac{\text{Export value}_{ict}}{\text{Total export value}_{it}}$$

which shows the relative importance for export of city c; $log(value)$, the logarithm of export value of firm i in month t at city c, which reflects the absolute level of export value; $Value\ Growth_{ict}$, the export value of firm i at city c in month t relative to the export value of firm i at city c in the last period $t-1$,

$$Value\ Growth_{ict} = \frac{\text{Export value}_{ict}}{\text{Export value}_{ic,t-1}}$$

which demonstrates the relative change in value for city c over last month; destination variation, the number of countries to export at firm-city-month level; production variation, defined as the number of different product classifications for a firm-city-month. The analysis on the last two variables are complements for the main analysis for export value, reflecting the external margin of connected firms to enter new geographic or product market.

$Female_{ict}$ is a dummy variable denoting whether the political leaders in city c and executives in listed firm i have females in month t. The coefficient on $Female$ is a difference-in-difference estimate of the effect of female connection on export outcomes. $Female = 0$ represents three cases: the executives are all male, and the political leaders are both male as well; the firm has at least one female executives, while the political leaders are both men; the executives are all male, and there is at least one woman in the political position. Therefore, the coefficient on $Female$ is the weighted average of (1) the difference between export outcomes for cities with female leaders, while firms change at least one executive to female; (2) the difference between export outcomes for firms keeping the lineup with women, while at least one female political leader is appointed to a city where the firm is producing goods for export; (3) originally no women in the firm management or supervising team and the city, then the firm changes at least one executive to female, and at the same time, at least one woman is appointed to one city where the firm has production.

Most empirical specifications include firm-city (γ_{ic}), city-month(τ_{ct}) and firm-month (θ_{it}) fixed effects. The firm-city fixed effect captures time-invariant characteristics of each firm in each city. An example would be that CEOs have hometown bias, so that firms allocate more resource in native cities, less in other cities. The inclusion of firm-city fixed effect ensures that the estimation of β reflects the change in Y from the rotation of either political leaders or executives. The city-month fixed effect captures demand or supply shocks in a particular prefecture city, which affect the decision making for internal investments, and thus confound the variation in Y besides the female connection. The firm-month fixed effect accounts for any shocks to the listed firm during a particular period. An example is the industry specific shocks. These fixed effects ensure that the estimation of β reflects the effect of female connection on export outcomes, not from reverse causality driven by the possibility that female political leaders are allocated to a thriving export location which is also more attractive for female entrepreneurs. The error term ϵ_{ict} is clustered at city level to account for the correlation across firms in the same city and serial correlation of export in one city[3].

3.4 Empirical Results

In this section, I present estimated results for the specification 3.1 with a variety of value measures. Then I try to single out the effect of female connection from other confounding stories.

3.4.1 Main Result

Table 3.2 presents the effect of having female connection on export values using the specification of equation 3.1. Column (1) and (2) show the results for the fraction of export value in city c over the total export value for firm i in month t. The estimated coefficient indicates that the fraction of export value in cities with female connection increases by 3 percentage point after controlling for a range of fixed effects. I also use the interaction of share of female executives in listed firms and female political leaders in column (2) as a continuous measure of female connection. The result implies that in a firm with 10 executives, one female executive's replacement of a male executive will increase the fraction of export value in cities with female leaders by 0.77 percentage points [4]. The dependent variable in column (3) is the logarithm of absolute export value at firm-city-month level. Firm i exports 16.9% more value from female connected cities than unconnected cities. Column (4) and (5) report results for export value growth. The dependent variable for column (4) is a dummy variable indicating whether the city experience an increase in export value. The dependent variable for column (5) is the simple growth rate of export value over last month, and then it is winsorized at 1% level on both sides. The estimated coefficients show that export value from female connected cities are more likely to experience an growth, and the value growth is significant compared to cities staying unconnected[5]. In Table B.1, I show that the female connection has no effect on the destination variation or production variation.

[3] I get similar results if the error term is clustered at the firm level.

[4] 1/10*0.0773=0.00773.

[5] Column (1) and (2) also control the lagged fraction of export value in city c for firm i. In column (3), (4), and (5), the lagged logarithm of export value in city c for firm i is controlled.

Table 3.2: Effects of Female Connection on Export Value

This table reports the regression results for $Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. Then dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1) and (2)); $log(value)$, the logarithm of export value of firm i in month t at city c (column (3)); whether the export value of firm i in city c month t increases over the last month t (column (4)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (5)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Fraction of Export Value	(3) Log(value)	(4) Dummy = 1 if Value Increase	(5) Value Growth
Female	0.0301***		0.169*	0.0742***	0.997**
	(0.00903)		(0.0993)	(0.0269)	(0.430)
Share Female		0.0773**			
		(0.0364)			
Firm-Month FE	Yes	Yes	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes	Yes	Yes
Cluster	City	City	City	City	City
# Clusters	158	158	158	158	158
Observations	43,736	43,736	43,736	43,736	43,736
R-squared	0.945	0.945	0.876	0.511	0.614

3.4.2 Confounding with Other Connection Factors

The female preference could be a reflection of other cultural or demographic proximity between political leaders and executives. Shen et al. (2019) shows that the hometown ties between CEOs and local government officials significantly predict a positive impact on tax avoidance for Chinese private firms. Think about the following story: a city thinks highly of female population, and less of males, therefore women accept more education, and thus become entrepreneurs, political leaders in another city. The entrepreneurs get more economic resource because of the hometown ties, not because of the gender similarity. Then the female connection is a reflection of hometown connection. Another story is that some university may get subsidies by sending graduates to one place, and then some graduates grow up to executives, some to political leaders in that place. If the university has more female students, my $Female$ measure will be a mirror of the university connection. To ensure that the female connection itself affects the internal decisions of listed firms, I construct two measures based on the above stories: shared native place or birth place ($Native$), university-related ($University$). As mentioned in Section 3.2, most of the executives' native places or educational background are missing. Based on this limited data, I construct $Native$ as a dummy variable, equal to 1 if at least one of the executives in the listed firm and political leaders are born in or native from the same city. Irrespective of the degrees (undergraduate, graduate, MBA, PhD) and the graduation year, I define $University$ as 1 as long as at least one executive in the listed firm, and one political leader have been to the same university for education.

Column (1) and (2) of Table 3.3 shows the estimates of coefficients when $Native$ and $University$ are included in specification 3.1. Neither the statistical or economic magnitude of $Female$ changes after controlling the two confounding factors. There is no effect of native connection or university connection on export values and growths. Again, I emphasize that due to the limited data, these results are only suggestive. More elaborate analysis should be taken if data is available.

Table 3.3: Confounding Factors

This table reports the regression results for $Y_{ict} = \beta_1 Female_{ict} + \beta_2 Native_{ict} + \beta_3 University_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. $Native$ is a dummy variable equal to 1 if at least one political leader and at least one executive are from the same native place, 0 otherwise. $University$ is a dummy variable equal to 1 if at least one political leader and at least one executive are educated in the same university, 0 otherwise. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
Female	0.0300***	0.986**
	(0.00906)	(0.429)
Native	-0.00611	-0.476
	(0.00918)	(0.290)
University	-0.00442	-0.00230
	(0.00454)	(0.196)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	158	158
Observations	43,736	43,736
R-squared	0.945	0.614

3.4.3 Financial Constraint

As stated in Stein (1997) and Giroud and Mueller (2015), for a firm with efficient internal capital market, the headquarters should reallocate resources from other plants to the treated plants if the firm is financially constrained. To test this hypothesis, I separate the whole sample of firms in each year into two subsamples based on the financial constraint measures in the previous year. If a firm's measure of financial constraints is above the median in the previous year, it belongs to the financial-constrained subsample (FC). If a firm's measure of financial constraints is below the median in the previous year, it belongs to the non-financial-constrained subsample (NFC). I use SA index from Hadlock and Pierce (2010) and leverage as measures of financial constraints. The SA index is defined as -0.737*Size + 0.043*Size^2 - 0.040*Age, where Size = log(asset), and age is defined as current year minus the IPO year. Literature also use KZ index (Kaplan and Zingales (1997)) or WW index (Whited and Wu (2006)) for measures of financing constraints. However, the construction of these two indexes includes financial information (cash flow, Q, dividend, leverage, cash/capital, sales growth, industry sales growth) that could be specific to the U.S. market. When applying these index to the Chinese firms, scholars get different estimates (Cull et al. (2015)). Although Hadlock and Pierce (2010)'s analysis is based on U.S. firms as well, the SA index only contains firm size and age, which makes the index less sensitive to the market characteristics. Moreover, I additionally use firm leverage as a measure of financial constraints. Leverage is defined as the the ratio between total liability and total assets.

Table 3.4 reports the estimated results for subsamples based on SA index. The effect of female connection only exists in financially constrained firms. The estimated coefficients for fraction of export value and value growth for the financially constrained subsample are comparable to those estimated on the whole sample. Therefore, consistent with Giroud and Mueller (2015), Chinese firms reallocated resources to connected cities only if they are financially constrained. For financially unconstrained firms, the two coefficients are positive but not significant. One possibility is that when one city become connected, the firm allocates more resources to the city but not significantly more than others, and thus this city experiences a relatively small growth on export value. The other possible explanation is that there is positive externality of the connected cities, so other unconnected cities also have an increase in value which causes no relative change for the connected cities against other cities. In both cases, there is no negative spillover from the connected city to the unconnected cities.

Table 3.4: Financial Constraint: SA Index

This table reports the estimated results for $Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1) and (2)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (3) and (4)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. The sample is separated into non-financially constrained (NFC) and financially constrained (FC) subsamples. Financial constraint is measured by SA index: -0.737*Size + 0.043*Size^2 - 0.040*Age, where Size = log(asset), age is defined as current year minus IPO year. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	Fraction of Export Value		Value Growth	
	NFC	FC	NFC	FC
	(1)	(2)	(3)	(4)
Female	0.0219	0.0339***	0.450	1.051**
	(0.0180)	(0.0126)	(0.630)	(0.484)
Firm-Month FE	Yes	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes	Yes
Cluster	City	City	City	City
# Clusters	104	109	104	109
Observations	16,928	18,480	16,928	18,480
R-squared	0.948	0.952	0.691	0.637

I get similar results for subsamples separated by firm leverage. The results are shown in Table 3.5. The two coefficients for financially constrained firms are large than those when using SA index to measure financial constraints. For the unconstrained subsample, the coefficient on female connection is slightly negative when the fraction of export value is the dependent variable, and for value growth, the coefficient is in the similar magnitude as in Table 3.4 where the SA index is a proxy for financial constraints.

3.5 Favoritism or Information

Fisman, Paravisini, and Vig (2017) emphasizes that the positive outcome between two parties in transactions can be explained by two prominent theories: favoritism and/or reduction in information friction. Favoritism leads to resource misallocation, and therefore harms both parties at the end. In the context of bank lending to borrowers, favoritism causes higher default rate and less repayment conditional on default. In terms of export of listed firms, favoritism results in producing goods of poor quality, and thus worse performance in the long run. By contrast, if the preference between two parties is due to less asymmetric information, both parties will benefit from these transactions. In my case, firms with females will be less discriminated than without the connection, and then the connection creates more value both for the firm and for the city. I distinguish the two prominent explanations for the impact by analyzing firms' ex-post performance, and product-level unit prices and quantities.

Table 3.5: Financial Constraint: Firm Leverage

This table reports the estimated results for $Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1) and (2)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (3) and (4)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. The sample is separated into non-financially constrained (NFC) and financially constrained (FC) subsamples. Financial constraint is measured by firm-level leverage: total liability divided by total asset. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	Fraction of Export Value		Value Growth	
	NFC	FC	NFC	FC
	(1)	(2)	(3)	(4)
Female	-0.00173	0.0625***	0.402	1.836**
	(0.0209)	(0.0146)	(0.851)	(0.765)
Firm-Month FE	Yes	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes	Yes
Cluster	City	City	City	City
# Clusters	112	106	112	106
Observations	17,330	18,531	17,330	18,531
R-squared	0.952	0.950	0.696	0.642

3.5.1 Ex-post Performance

Borrowing from Fisman, Paravisini, and Vig (2017), I test the impact of female connection on firms' performance in the long run. I use firms' return on equity ratio as the measure of performance. Financial statements are at most at a frequency of quarter. Therefore, the analysis is on firm-quarter level. The specification is

$$ROE_{it+4} = \alpha + \beta Female_{it} + \delta \mathbf{X_{it}} + \gamma_i + \theta_t + \epsilon_{it} \tag{3.2}$$

The dependent variable is the return on equity ratio in the quarter $t+4$. The female connection is aggregated into firm-quarter level and still a dummy variable. It equals 1 if during the quarter, the firm has at least one female connection in cities in which it produces goods for export. $\mathbf{X_{it}}$ is time-varying firm characteristics, including the logarithm of total assets, leverage ratio defined as total liability over total assets. Firm and quarter fixed effects are included to control for time-invariant firm characteristics, and national demand or supply shocks. Standard errors are clustered at industry level to control for the correlation between firms in the same industry and serial correlation for firms in the the industry.

Column (1) and (2) in Table 3.6 presents the estimated results. Female connection results in around 2.76 percentage point increase in ROE in $t+4$, and the coefficient is similar when extra controls are included in the analysis.

The average tenure of a political leader is 28.65 months. So the estimation could be a reflection of female connection on contemporary performance, instead of the long run impact. I test this possibility by including two interaction terms: a firm has female connection in t and also female connection in 4 quarters ($Female_t \times Female_{t+4}$); a firm has female connection in t and become unconnected in 4 quarters ($Female_t \times (1-Female_{t+4})$). The coefficient on the latter one captures the effect on ex-post performance if a firm goes from connected to unconnected group. Column (3) in Table 3.6 shows that both coefficients on the interaction terms are positive and significant at 5% level. Thus firms with female connections have better performance in 4 quarters even if this connection does not exist at that time. After the controls are included, the coefficients stay significant

Table 3.6: Effect of Female Connection on Ex-post Performance

This table reports the estimated results for female connection on ex-post firm performance in $t+4$. t denotes quarter. The unit of analysis is firm-quarter. *Female* is a dummy variable, defined as 1 if during quarter t, firm i has female connection with any cities in which the firm produces goods for export. Log(asset) is the logarithm of firm's asset at the end of quarter t. Leverage is defined as total liability over total asset at the end of quarter t. Firm and quarter fixed effects are controlled. Robust standard errors clustered by industry are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) ROE in $t+4$	(2) ROE in $t+4$	(3) ROE in $t+4$	(4) ROE in $t+4$
Female	0.0276**	0.0262**		
	(0.0116)	(0.0116)		
$Female \times Female_{t+4}$			0.0236**	0.0212*
			(0.0102)	(0.0110)
$Female \times (1 - Female_{t+4})$			0.0320**	0.0315**
			(0.0151)	(0.0148)
Log(asset)		0.0664		0.0665
		(0.0517)		(0.0517)
Leverage		-0.201		-0.201
		(0.147)		(0.147)
Constant	0.00698	-1.313	0.00738*	-1.314
	(0.00432)	(1.044)	(0.00404)	(1.043)
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Cluster	Industry	Industry	Industry	Industry
# Clusters	68	68	68	68
Observations	3,405	3,264	3,405	3,264
R-squared	0.138	0.141	0.138	0.141

and the statistical significance does not change much in column (4). As suggested in Fisman, Paravisini, and Vig (2017), these positive coefficients indicate that it is not direct supervision or monitoring which requires the presence of political leaders that causes this increase in performance. The results support the theory that female connection reduces information friction and creates more value for listed firms.

3.5.2 Unit Price and Quantity

Either the growth on unit price or quantity or both could generate a rise in export value. Higher unit price may suggest better quality or more bargaining power on the export market. If the local government has the right to set quotas for firms to export based on the information it has, the local government may allocate less resource for firms where asymmetric information is prevailing. The increase in quantity supports the explanation of reducing information friction either because listed firms move the inventory to connected cities for export or allocate more resources to connected cities for production.

To examine whether it is unit price or quantity that drives up the export value, I exploit the following specification

$$Growth\ of\ Y_{ihct} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \delta_h + \epsilon_{ihct} \tag{3.3}$$

To make the unit price comparable across products, I include product h as another dimension in the analysis. The unit of analysis is firm-product-city-month level. I use HS 8-digit code to classify export goods. The results are similar if I use HS 6-digit code as the classification as in other papers, such as Fan, Li, and Yeaple (2015). Three dependent variables are analyzed separately: the growth rate of unit export price, of export quantity, of export value. The growth rate is computed as $\frac{Y_{ihct}}{Y_{ihc,t-1}}$.

Table 3.7 presents the estimated results for specification 3.3. Female connection causes an increase in growth rate of quantity of 0.388, similar to the magnitude as the impact on the growth in export value 0.350. Female connection has no effect on change of export price. Therefore, the positive impact of female connection on export value is through quantity increase, providing another evidence for the explanation of less asymmetric information between local government and listed firms.

Table 3.7: Effect of Female Connection on Unit Price and Quantity

This table reports estimated results for $Growth\ of\ Y_{ihct} = \beta Female_{ihct} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \delta_h + \epsilon_{ict}$. The dependent variables are: the growth of export price, export quantity and export value. I define growth as the simple growth from $t-1$ to t for product h in firm i at city c. $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Growth of Export Price	(2) Growth of Export Quantity	(3) Growth of Export Value
Female	0.00447	0.388**	0.350***
	(0.0230)	(0.152)	(0.0964)
Firm-Month FE	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes
HS-digit FE	Yes	Yes	Yes
Cluster	City	City	City
# Clusters	158	158	158
Observations	574,734	574,984	575,436
R-squared	0.075	0.105	0.106

3.6 Heterogeneity Analysis

This section presents robustness tests by doing several heterogeneity analyses across: positions of executives, positions of political leaders, concentration of firms' production, local competition, firm size and firm age. These analyses help to get more insights on the explanation of the positive relation between female connection and export value.

3.6.1 Heterogeneity by Position

As shown in section 3.2, political leaders contain mayors and secretaries. Executives could be partitioned into two groups: CEOs (chairman, vice chairman, CEO, vice CEO), others (directors, supervisors and others). Then it is natural to ask whether the effect of female connection differs by the positions of political leaders and executives. In Table 3.8, I generate four new variables: $Female\ Mayor$, denoted to 1 if at least one executive and the mayor in the city are female; $Female\ Secretary$, denoted to 1 if at least one executive and the secretary in the city are female; $Female\ CEO$, at least one of the CEOs and political leaders are female; $Female\ Other$, CEOs are not female, while at least one other executive (except CEOs) and one of the political leaders are female. In column (1) and (2), I include $Female\ Mayor$ and $Female\ Secretary$ in the regression to explore which political leader matters in the female connection on export value. Column (3) and (4) answer the question which type of executives plays a role in the result, and thus the independent variables are $Female\ CEO$, and $Female\ Other$. Connections with secretary drive the effect on share of export value, while connections with mayors lead to more value growth. In different settings, I get significant coefficient for both $Female\ Mayor$ and $Female\ Secretary$, but the relative magnitude does not change. An explanation is that the secretary is the chief leader in a prefecture city and has more control power over personnel affairs, thus the connection with the secretary makes firms to shift their production to the connected cities. The mayor controls the finances, and therefore the production process, so cities with mayor connections see a large value growth. The significant coefficients in column (3) and (4) indicate that both connections from CEOs and other types of executives contribute to higher export value and growth.

Table 3.8: Heterogeneity by Position of Political Leaders and Executives

This table reports the regression results for $Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1), (3)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2), (4)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. $Female\ Mayor_{ict}$ is a dummy variable denoting whether the mayor in city c is female and executives in listed firm i have females in month t. $Female\ Secretary_{ict}$ is a dummy variable denoting whether the secretary in city c is female and executives in listed firm i have females in month t. $Female\ CEO_{ict}$ is 1 if at least one CEO and political leader are female in city c, firm i, month t. $Female\ Other_{ict}$ is 1 if CEOs are not female, while at least one other executive (except CEOs) and political leader are female in city c, firm i, month t. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth	(3) Fraction of Export Value	(4) Value Growth
Female Mayor	0.0159	1.442**		
	(0.0121)	(0.656)		
Female Secretary	0.0392***	0.710		
	(0.0141)	(0.508)		
Female CEO			0.0196*	1.010**
			(0.00997)	(0.475)
Female Other			0.0367***	0.988**
			(0.00961)	(0.436)
Firm-Month FE	Yes	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes	Yes
Cluster	City	City	City	City
# Clusters	158	158	158	158
Observations	43,736	43,736	43,736	43,736
R-squared	0.945	0.614	0.945	0.614

3.6.2 Heterogeneity by Concentration of Production

The effect of female connection on export value may be different for firms with various degrees of concentration on production. Borrowing from the Herfindahl Index, I construct the concentration of production for each firm-month as:

$$concentration_{it} = \sum_c \left(\frac{\text{Export value}_{ict}}{\text{Total export value}_{it}}\right)^2 \tag{3.4}$$

where $\frac{\text{Export value}_{ict}}{\text{Total export value}_{it}}$ represents the share of firm i's export at city c in month t over the total export of firm i in month t. The less the *concentration* is, the more diversified the firm is. Since I do not distinguish different products in this measure, $concentration_{it}$ could be a measure of diversified business lines, or the firm producing the same product in various places to hedge against unexpected supply or demand shocks. The number of cities that listed firms export from ranges from 2 to 50 with an average of 5.5. And the concentration varies from 0.09 to 0.99 with an average of 0.62. For each month, firms are partitioned into 4 groups according to the degree of concentration in production. Quartile 1 includes the most diversified firms, and the concentration level increases along the quartiles. To examine the effect of female connection on various concentration levels, $Female$ is interacted with quartiles of concentration in specification 3.1:

$$Y_{ict} = \sum_g \beta_g Female_{ict} \times quartile_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict} \quad (3.5)$$

where g stands for the quartile the firm-city-month observation belongs to, from 1 to 4. The estimated results are presented in Table 3.9. Column (1) shows that female connection has no significant effect on most diversified firms, even though the coefficient is positive. However, coefficients for quartile 2 to quartile 4 are positive and significant at 1% level. This implies that the effect of female connection is more significant for more concentrated firms. Regarding the value growth in column (2), coefficients in all quartiles are positive and significant. While the coefficient in quartile 1 is significant at 10% level, the economic magnitude is less than the other three quartiles. It is possible that when a diversified firm has the connection, it increases production in all plants not just in the connected one. This leads to the fraction of export for the connected city unchanged, and at the same time the connected city experiences a growth in value.

Table 3.9: Heterogeneity by Concentration of Production

This table reports the estimated results for $Y_{ict} = \sum_g \beta_g Female_{ict} \times concentration_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. The concentration of each listed firm i in each month t is calculated as $concentration_{it} = \sum_c (\frac{\text{Export value}_{ict}}{\text{Total export value}_{it}})^2$. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
$Female \times concentration = quartile\ 1$	0.0140	0.871*
	(0.00975)	(0.445)
$Female \times concentration = quartile\ 2$	0.0501***	1.199***
	(0.0124)	(0.402)
$Female \times concentration = quartile\ 3$	0.0333***	0.970*
	(0.0123)	(0.534)
$Female \times concentration = quartile\ 4$	0.0446***	1.119**
	(0.0171)	(0.510)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	158	158
Observations	43,736	43,736
R-squared	0.945	0.614

Firms with less concentration have plants in multiple cities and normally, their market share in each individual city is quite high. The firm whose production is concentrated in a limited number of cities accounts for a less proportion in local cities. Therefore, those firms in quartile 4 should suffer more from the asymmetric information problems. From the result on column (2), we can not rule out the explanation of favoritism, but the results for fraction of export value do imply that mitigation of information friction plays a more essential role.

3.6.3 Heterogeneity by Local Competition

This part evaluates how the effect of female connection on export value differs in more competitive and more concentrated market. I measure the local competition in a city during a specified period with all export transactions, not just the transactions from listed firms.

$$Competition_{ct} = \sum_i \left(\frac{\text{Export value}_{ict}}{\text{Total export value}_{ct}}\right)^2$$

where i denotes all firms having export business in city c during month t. $\frac{\text{Export value}_{ict}}{\text{Total export value}_{ct}}$ measures firm i's export share in city c in month t over total export of all firms in city c in month t. The higher the measure of $Competition$ is, the less competitive the local market is. For each month, cities are partitioned into 4 groups according to their local competition level.

Table 3.10 reports the estimated results by exploiting specification 3.5 with competition quartiles as the interaction term. Column (1) shows that the effect of female connection on fraction of export value is not significant for either most competitive (quartile = 1) nor most concentrated (quartile = 1) market. However, the average local competition measure for the first three quartiles are pretty close, 0.001, 0.003 and 0.009, and that for quartile 4 is 0.06. Combining with the results in column (2), where quartile 1 and quartile 3 experience most significant increase in value growth, I conclude that the effect is more important in cities with more competition. The results shown in Table B.2 are similar if only listed firms are included in computing local competition, and in this case, the competition level is much higher with the average of 0.23, 0.27, 0.36 and 0.60 from quartile 1 to quartile 4.

Table 3.10: Heterogeneity by Local Competition

This table reports the estimated results for $Y_{ict} = \sum_g \beta_g Female_{ict} \times competition_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. I calculate the local competition in each city-month ct as $competition_{ct} = \sum_i \left(\frac{\text{Export value}_{ict}}{\text{Total export value}_{ct}}\right)^2$. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
$Female \times competition = quartile\ 1$	0.00890	1.827**
	(0.0173)	(0.806)
$Female \times competition = quartile\ 2$	0.0376***	0.828
	(0.00865)	(0.734)
$Female \times competition = quartile\ 3$	0.0304**	1.176***
	(0.0146)	(0.427)
$Female \times competition = quartile\ 4$	0.0219	0.250
	(0.0260)	(1.187)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	158	158
Observations	43,736	43,736
R-squared	0.945	0.614

This evidence is hard to reconcile with the favoritism theory. Suppose there are two firms in one city, then a woman is assigned as the secretary of the city. One of the firm is connected, and the other is not. It is more reasonable that the local government has enough knowledge about both firms as there are only two players in the market. Any preference of one firm over the other one should be out of favoritism. The one with connection will crowd out the unconnected one from this market, and thus see an increase of production of 100% if assuming the market share of either firm is 50% before. In this case, the coefficient for firms with less competition should see a huge growth of production. But the results in Table 3.10 present that the effect of female connection is not significant and the coefficients are relatively small in more concentrated market compared to less competitive market.

3.6.4 Heterogeneity by Firm Size

The effect of female connection on export value may vary by the size of listed firms. Big firms attract more attention from the public and have higher trading volume in security markets. In contrast, small firms suffer from more severe problems due to asymmetric information. If it is favoritism plays the role, the effect of female connection on export value should be more significant for big firms. While if the female connection mitigates the information friction, small firms should benefit more from the connection with local political leaders.

Table 3.11: Heterogeneity by Firm Size

This table reports the estimated results for $Y_{ict} = \sum_g \beta_g Female_{ict} \times size_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. I use the lagged asset as the size measure. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
$Female \times size = quartile\ 1$	0.0501***	1.446**
	(0.0132)	(0.642)
$Female \times size = quartile\ 2$	0.0298**	0.563
	(0.0130)	(0.638)
$Female \times size = quartile\ 3$	0.0303**	1.015**
	(0.0123)	(0.456)
$Female \times size = quartile\ 4$	0.0191	0.878**
	(0.0116)	(0.389)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	158	158
Observations	41,410	41,410
R-squared	0.943	0.620

The size of listed firms is measured as the latest available asset value in logarithm. Then for each month, each firm is partitioned into quartiles based on the size. Table 3.11 presents results with specification 3.5 by allowing $Female$ to vary across size quartiles. In column (1), where the dependent variable is the fraction of export value, the effect is strongest for small firms, and the magnitude declines along the quartiles. The effect is not significant for big firms. Column (2) shows results for value growth. The effect is strongest for smallest firms as well. But now for the largest firms, the coefficient is significant at 5% level. It is possible that when big firms have the connection, they increase production in all the cities, which causes that the fraction of export value of the connected city stays the same, but the absolute value in the connected city increases. This is a positive externality of female connection. Similar as the conclusion from the analysis on concentration, the favoritism-based explanation cannot be ruled out, but my results do favor the explanation about information.

3.6.5 Heterogeneity by Firm Age

In Table 3.12, I conduct an analysis of heterogeneity with firm age. Similar as the above results, I divide the firms into quartiles for each month based on their age and then interact the quartiles with the female connection measure. The result in column (1) shows that the effect of female connection is significantly positive for all age quartiles, although the coefficient is smaller for older firms (quartile = 4). For the value growth in column (2), however, the effect of $Female$ is only significant for younger firms. As young firms suffer more severe information problems, the results are also inclined to the information-based explanation.

Table 3.12: Heterogeneity by Firm Age

This table reports the estimated results for $Y_{ict} = \sum_g \beta_g Female_{ict} \times age_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. Firm age is defined as the current year minus the IPO year. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
$Female \times age = quartile\ 1$	0.0362***	1.079*
	(0.0126)	(0.621)
$Female \times age = quartile\ 2$	0.0304**	1.144***
	(0.0130)	(0.404)
$Female \times age = quartile\ 3$	0.0351**	0.817
	(0.0162)	(0.506)
$Female \times age = quartile\ 4$	0.0235**	0.802
	(0.0117)	(0.584)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	157	157
Observations	41,010	41,010
R-squared	0.942	0.619

3.7 Conclusion

This paper provides evidence that female connection between executives in listed firms and local political leaders has a positive effect on the fraction of export value and value growth in the connected city. The identification is based on the exogenous rotation of political leaders in the prefecture-city level. By showing that firms with connectedness outperform ex-post, and that the above effect is stronger for firms with more financial constraints, higher concentration, smaller size, younger age and facing fiercer local competition, the results favor the explanation of mitigating information friction. But I can not rule out the favoritism-based explanation. More rigorous work should be done on this topic.

How does the female connection affect the internal decision making is unclear. It could be due to the sameness on marital and parental status among women as emphasized in Stolper and Walter (2019). The other possible explanation is that female is the minority in the political world, and women are discriminated everywhere. Thus a sympathetic chord makes female political leaders favor more female-friendly companies. This is in line with women's solidarity, women's inherent co-operativeness with each other. To examine these possibilities, survey data is necessary. The mechanism of how the local political leaders allocate more resources to the connected firms is ambiguous as well. Possible channels could be through credit allocation or trade quota allocation. Rigorous work is needed. I leave these for future work.

Conclusion

In this dissertation, I analyze several separate topics in corporate finance studies.

In chapter 1, we explore the unintended consequence and the mechanism of the fiscal stimulus plan undertook during 2009-2010 in China. By constructing a model where the local government is the monopolistic supplier of land in the local market, we show that it is always optimal for the local government to use land as a collateral to borrow from banks instead of selling land in the primary market to private sector. Our empirical findings show that local governments sell land usufruct rights to local government financing vehicles through the primary land market at lower prices, and then the local government financing vehicles post land parcels as collateral at higher valuation to borrow from banks. We also show that local cities with more political connections with the central government experience higher loan growth. As a result of the land supply to local government financing vehicles, housing prices rise due to less residential land supply to private sectors. Overall, our findings contribute to a better understanding of the role of local government financing, macro policies and potential effects on the asset market.

In chapter 2, we concentrate on the IPO stocks in the United States. The prospectuses submitted to the Securities and Exchange Commission at the beginning of the IPO process (S-1 filings) reflect the outlook of managers, and thus should have long-run impacts on firm performance if the stocks are not correctly priced during IPO. In this paper, we explore the relationship between sentiments revealed in S-1 filings and IPOs' long-run performance. We find a positive link between negative sentiment and long-term overperformance. This link is significant for different measures of negative sentiment and long-term return with various benchmarks. The quantile regression results and calendar time portfolio analysis reinforce the link. This link may come from higher risk taking, CEOs' overconfidence, or irrationality of retail investors. The sorting and regression results do not support the explanation with higher risks or CEOs' overconfidence. With search volume index from Google Trends as a proxy for retail investor attention, we find that the positive link between negative sentiment and long-run return channels through the retail investor attention.

In chapter 3, my focus is on the role women play in firms' internal resource allocation decisions. By analyzing the export transaction data for China from 2000 to 2006, this paper provides evidence that female connection between executives in listed firms and local political leaders has a positive effect on the fraction of export value and value growth in the connected city. The identification is based on the exogenous rotation of political leaders in the prefecture-city level. There are two prominent explanations on this topic typically: favoritism and information. The results favor the explanation of mitigating information friction. I find that firms with connectedness outperform ex-post, and the heterogeneity analyses show that the above effect is stronger for firms with financial constraints, higher concentration, larger size, younger age and facing fiercer local competition. But I can not rule out the favoritism-based explanation. The mechanism behind is unclear though. How does the female connection affect the internal decision making? It could be due to the sameness on marital and parental status among women, or the sympathetic chord between females. How do the local political leaders allocate more resources to the connected firms? It is possible that there is an export quota for a city and the local political leaders have the control over allocating the quota to different firms.

All my papers have some shortcomings. Some of them is due to the limitation of data. And even if with the data, the empirical specification is always an issue. So research never ends.

Appendices

Appendix A

Appendix to IPO Prospectus Language and Long-Run Performance

A.1 Additional Tables and Figures

Table A.1: Return Differentials for Negative Sentiment Portfolios with Bootstrap

This table reports abnormal returns for portfolios of IPOs with negative sentiment at the top quintile, at the bottom quintile, and a strategy to long the top quintile and short the bottom quintile, for IPOs listed within the past 36-month or 60-month periods. The portfolios are equal weighted. Holding periods are 1-12 months. Alphas estimated from the four-factor model are shown in the table. The four-factor model includes the three Fama and French (1993) factors, and the Carhart (1997) momentum factor. Estimates are reported for the holding period net return. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	Holding Period				
	1m	3m	6m	9m	12m
Panel A: 36-month IPOs					
Bottom Quintile	-0.00160	-0.00517	-0.00908	-0.00756	-0.0122
	(0.00225)	(0.00391)	(0.00574)	(0.00746)	(0.00892)
Top Quintile	0.00424	0.00452	0.00780	0.0139	0.0200
	(0.00394)	(0.00701)	(0.0116)	(0.0146)	(0.0170)
Top - Bottom	0.00584	0.00969	0.0169*	0.0214*	0.0322**
	(0.00389)	(0.00625)	(0.00910)	(0.0124)	(0.0162)
Panel B: 60-month IPOs					
Bottom Quintile	8.30e-05	0.00129	0.00242	-0.00119	-0.00628
	(0.00182)	(0.00358)	(0.00597)	(0.00680)	(0.00757)
Top Quintile	0.00720*	0.0103*	0.0145	0.0110	0.00682
	(0.00372)	(0.00622)	(0.0101)	(0.0121)	(0.0134)
Top - Bottom	0.00711**	0.00905*	0.0121	0.0122	0.0131
	(0.00360)	(0.00526)	(0.00785)	(0.00903)	(0.0102)

Figure A.1: Search Volume Index around IPO

The figure plots the mean and median weekly search volume index (SVI) for U.S. listed IPO stocks over the period 2004 to 2016. The x-axis represents the week around IPO; the y-axis shows the raw index.

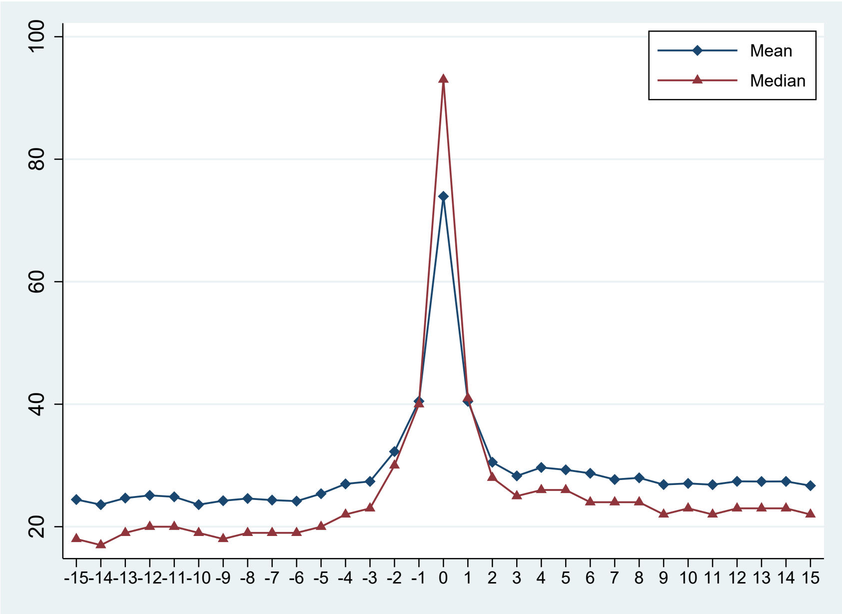
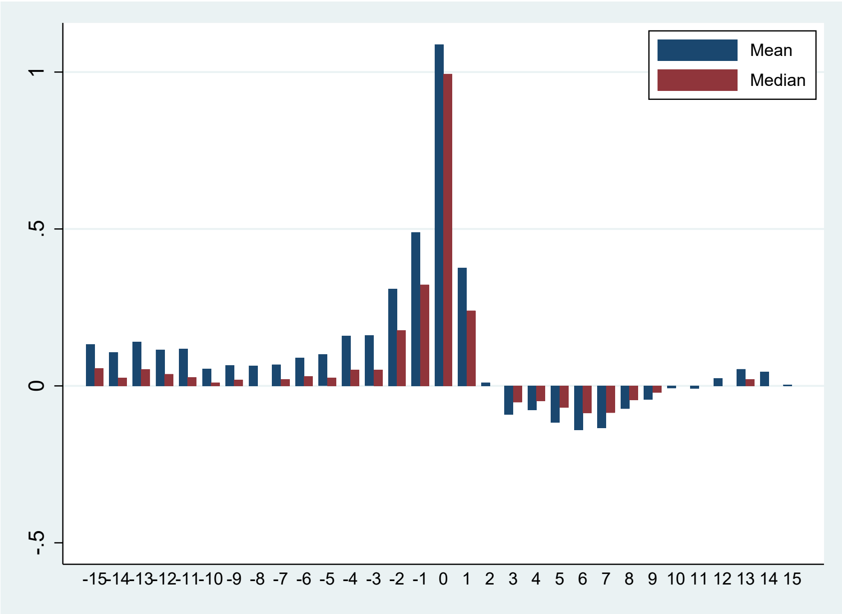


Figure A.2: Abnormal Search Volume Index around IPO

The figure plots the mean and median weekly abnormal search volume index (ASVI) for U.S. listed IPO stocks over the period 2004 to 2016. The x-axis represents the week around IPO. *ASVI* is defined as the log of SVI (search volume index) during the week minus the log of average SVI during the previous 8 weeks.



Appendix B

Appendix to Female Networks and Corporate Resource Allocation in China

B.1 Additional Tables

Table B.1: Effects of Female Connection on Product and Destination Variation

This table reports the regression results for $Y_{ict} = \beta Female_{ict} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. Then dependent variables are: production number, defined as the number of different product classifications for a firm-city-month (Column (1)); destination number, the number of countries to export at firm-city-month level (Column (2)); production variation, defined as the change of number of different product classifications for a firm-city-month (Column (3)); destination variation, the change of number of countries to export at firm-city-month level (Column (4)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) # Product	(2) # Destination	(3) Change Product	(4) Change Destination
Female	0.929	-0.0528	-0.0314	0.0396
	(1.276)	(0.471)	(0.0969)	(0.0526)
Constant	11.75***	5.881***	1.217***	1.144***
	(0.0716)	(0.0265)	(0.00580)	(0.00315)
Firm-Month FE	Yes	Yes	Yes	Yes
City-Month FE	Yes	Yes	Yes	Yes
Firm-City FE	Yes	Yes	Yes	Yes
Cluster	City	City	City	City
# Clusters	158	158	158	158
Observations	68,725	68,725	43,736	43,736
R-squared	0.937	0.945	0.383	0.400

Table B.2: Heterogeneity by Local Competition, Only Listed Firms

This table reports the estimated results for $Y_{ict} = \sum_g \beta_g Female_{ict} \times competition_{itg} + \gamma_{ic} + \tau_{ct} + \theta_{it} + \epsilon_{ict}$. The dependent variables are: the export value of firm i in month t at city c relative to the overall export value of firm i in month t (column (1)); the export value growth of firm i at city c in month t relative to the export value of firm i in the last period $t-1$ at city c (column (2)). $Female_{ict}$ is a dummy variable denoting whether both the political leaders in city c and executives in listed firm i have females in month t. I calculate the local competition in each city-month ct as $competition_{ct} = \sum_i (\frac{\text{Export value}_{ict}}{\text{Total export value}_{ct}})^2$. Only Listed firms' export is included. Robust standard errors clustered at city level are shown in the parentheses. *, **, *** indicates statistical significance at the 10%, 5%, and 1% level, respectively.

	(1) Fraction of Export Value	(2) Value Growth
$Female \times competition = quartile$ 1	0.0439**	0.731
	(0.0174)	(0.956)
$Female \times competition = quartile$ 2	0.0117	0.989*
	(0.00780)	(0.552)
$Female \times competition = quartile$ 3	0.0506***	2.405**
	(0.0176)	(1.101)
$Female \times competition = quartile$ 4	-0.00575	-0.502
	(0.0333)	(1.402)
Firm-Month FE	Yes	Yes
City-Month FE	Yes	Yes
Firm-City FE	Yes	Yes
Cluster	City	City
# Clusters	158	158
Observations	43,736	43,736
R-squared	0.945	0.614

List of Tables

List of Figures

Bibliography

Adams, Renée B and Daniel Ferreira. 2009. "Women in the boardroom and their impact on governance and performance." *Journal of Financial Economics* 94 (2):291–309.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2012. "Credit supply and house prices: evidence from mortgage market segmentation." Tech. rep., National Bureau of Economic Research.

Allen, Franklin and Ana Babus. 2009. "Networks in finance." *The network challenge: strategy, profit, and risk in an interlinked world* 367.

Altı, Aydoğan. 2005. "IPO market timing." *The Review of Financial Studies* 18 (3):1105–1138.

Ambrose, Brent W, Yongheng Deng, and Jing Wu. 2015. "Understanding the risk of China's local government debts and its linkage with property markets." *Available at SSRN 2557031* .

Ang, Andrew, Jennie Bai, and Hao Zhou. 2016. "The great wall of debt: Real estate, political risk, and Chinese local government credit spreads." *Columbia Business School Research Paper* (15-57).

Bai, Chong-En, Chang-Tai Hsieh, and Zheng Michael Song. 2016. "The long shadow of China's fiscal expansion." *Brookings Papers on Economic Activity* 2016 (2):129–181.

Bai, Jennie and Hao Zhou. 2018. "The funding cost of Chinese local government debt." *Georgetown McDonough School of Business Research Paper* (2603022):15–02.

Baker, Malcolm and Jeffrey Wurgler. 2006. "Investor sentiment and the cross-section of stock returns." *The Journal of Finance* 61 (4):1645–1680.

Becker, Bo and Victoria Ivashina. 2018. "Financial repression in the European sovereign debt crisis." *Review of Finance* 22 (1):83–115.

Becker, Gary. 1996. "Preferences and values." *Accounting for Taste, Harvard University Press: Cambridge* .

Bertrand, Marianne, Francis Kramarz, Antoinette Schoar, and David Thesmar. 2018. "The cost of political connections." *Review of Finance* 22 (3):849–876.

Betzer, André, Inga van den Bongard, Erik Theissen, and Christine Volkmann. 2017. "All is not lost that is delayed: overconfidence and investment failure." *SSRN Working Paper Series* .

Brav, Alon and Paul A Gompers. 1997. "Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and nonventure capital-backed companies." *The Journal of Finance* 52 (5):1791–1821.

Brueckner, Jan K, Shihe Fu, Yizhen Gu, and Junfu Zhang. 2017. "Measuring the stringency of land use regulation: The case of China's building height limits." *Review of Economics and Statistics* 99 (4):663–677.

Brunnermeier, Markus, Michael Sockin, and Wei Xiong. 2017. "China's Gradualistic Economic Approach and Financial Markets." *American Economic Review Papers Proceedings* 107 (5):608–613.

Burchardi, Konrad B and Tarek A Hassan. 2013. "The economic impact of social ties: Evidence from German reunification." *The Quarterly Journal of Economics* 128 (3):1219–1271.

Cai, Hongbin, J Vernon Henderson, and Qinghua Zhang. 2013. "China's land market auctions: evidence of corruption?" *The Rand journal of economics* 44 (3):488–521.

Carhart, Mark M. 1997. "On persistence in mutual fund performance." *The Journal of Finance* 52 (1):57–82.

Chen, Kaiji and Yi Wen. 2017. "The great housing boom of China." *American Economic Journal: Macroeconomics* 9 (2):73–114.

Chen, Ting and James Kai-sing Kung. 2018. "Busting the "Princelings": The campaign against corruption in China's primary land market." *The Quarterly Journal of Economics* 134 (1):185–226.

Chen, Ting and JK-S Kung. 2016. "Do land revenue windfalls create a political resource curse? Evidence from China." *Journal of Development Economics* 123:86–106.

Chen, Zhuo, Zhiguo He, and Chun Liu. 2020. "The financing of local government in China: Stimulus loan wanes and shadow banking waxes." *Journal of Financial Economics* .

Claessens, Stijn, Erik Feijen, and Luc Laeven. 2008. "Political connections and preferential access to finance: The role of campaign contributions." *Journal of Financial Economics* 88 (3):554–580.

Cohen, Lauren, Joshua Coval, and Christopher Malloy. 2011. "Do powerful politicians cause corporate downsizing?" *Journal of Political Economy* 119 (6):1015–1060.

Cohen, Lauren, Umit G Gurun, and Christopher Malloy. 2017. "Resident networks and corporate connections: Evidence from World War II internment camps." *The Journal of Finance* 72 (1):207–248.

Cong, Lin William, Haoyu Gao, Jacopo Ponticelli, and Xiaoguang Yang. 2019. "Credit allocation under economic stimulus: Evidence from China." *The Review of Financial Studies* 32 (9):3412–3460.

Cornelli, Francesca, David Goldreich, and Alexander Ljungqvist. 2006. "Investor sentiment and pre-IPO markets." *The Journal of Finance* 61 (3):1187–1216.

Cornwall, Andrea. 2007. "Myths to live by? Female solidarity and female autonomy reconsidered." *Development and Change* 38 (1):149–168.

Cull, Robert, Wei Li, Bo Sun, and Lixin Colin Xu. 2015. "Government connections and financial constraints: Evidence from a large representative sample of Chinese firms." *Journal of Corporate Finance* 32:271–294.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. "In search of attention." *The Journal of Finance* 66 (5):1461–1499.

Dasgupta, Amil, Andrea Prat, and Michela Verardo. 2011. "Institutional trade persistence and long-term equity returns." *The Journal of Finance* 66 (2):635–653.

Demirci, Irem, Jennifer Huang, and Clemens Sialm. 2019. "Government Debt and Corporate Leverage: International Evidence." *Journal of Financial Economics* 133 (2):337–356.

Deng, Yongheng, Randall Morck, Jing Wu, and Bernard Yeung. 2014. "China's pseudo-monetary policy." *Review of Finance* 19 (1):55–93.

Derrien, Franois. 2005. "IPO pricing in "hot" market conditions: who leaves money on the table?" *The Journal of Finance* 60 (1):487–521.

Di Maggio, Marco and Amir Kermani. 2017. "Credit-induced boom and bust." *The Review of Financial Studies* 30 (11):3711–3758.

Ding, Haoyuan, Haichao Fan, and Shu Lin. 2018. "Connect to trade." *Journal of International Economics* 110:50–62.

Duchin, Ran and Denis Sosyura. 2013. "Divisional managers and internal capital markets." *The Journal of Finance* 68 (2):387–429.

Engelberg, Joseph, Pengjie Gao, and Christopher A Parsons. 2012. "Friends with money." *Journal of Financial Economics* 103 (1):169–188.

Faccio, Mara, Ronald W Masulis, and John J McConnell. 2006. "Political connections and corporate bailouts." *The journal of Finance* 61 (6):2597–2635.

Fama, Eugene F and Kenneth R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33 (1):3–56.

Fama, Eugene F and Kenneth R French. 1997. "Industry costs of equity." *Journal of Financial Economics* 43 (2):153–193.

Fama, Eugene F and Kenneth R. French. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116 (1):1–22.

Fan, Haichao, Yao Amber Li, and Stephen R Yeaple. 2015. "Trade liberalization, quality, and export prices." *Review of Economics and Statistics* 97 (5):1033–1051.

Fan, Joseph PH, Tak Jun Wong, and Tianyu Zhang. 2007. "Politically connected CEOs, corporate governance, and Post-IPO performance of China's newly partially privatized firms." *Journal of Financial Economics* 84 (2):330–357.

Fang, Hanming, Quanlin Gu, Wei Xiong, and Li-An Zhou. 2016. "Demystifying the Chinese housing boom." *NBER macroeconomics annual* 30 (1):105–166.

Fang, Lily Hua and Sterling Huang. 2017. "Gender and connections among Wall Street analysts." *The Review of Financial Studies* 30 (9):3305–3335.

Favara, Giovanni and Jean Imbs. 2015. "Credit supply and the price of housing." *American Economic Review* 105 (3):958–92.

Favilukis, Jack, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. 2017. "The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium." *Journal of Political Economy* 125 (1):140–223.

Field, Laura Casares and Michelle Lowry. 2009. "Institutional versus individual investment in IPOs: The importance of firm fundamentals." *Journal of Financial and Quantitative Analysis* :489–516.

Finkelstein, Amy, James Poterba, and Casey Rothschild. 2009. "Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities." *Journal of Financial Economics* 91 (1):38–58.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. 2017. "Cultural proximity and loan outcomes." *American Economic Review* 107 (2):457–92.

Fracassi, Cesare. 2017. "Corporate finance policies and social networks." *Management Science* 63 (8):2420–2438.

Gabaix, Xavier and Ralph SJ Koijen. 2020. "Granular instrumental variables." Tech. rep., National Bureau of Economic Research.

Gao, Haoyu, Hong Ru, and Dragon Yongjun Tang. 2020. "Subnational debt of China: The politics-finance nexus." *Available at SSRN 2801878* .

Gao, Xiaohui, Jay R Ritter, and Zhongyan Zhu. 2013. "Where have all the IPOs gone?" *Journal of Financial and Quantitative Analysis* 48 (6):1663–1692.

Gervais, Simon, James B Heaton, and Terrance Odean. 2011. "Overconfidence, compensation contracts, and capital budgeting." *The Journal of Finance* 66 (5):1735–1777.

Gilje, Erik P, Elena Loutskina, and Philip E Strahan. 2016. "Exporting liquidity: Branch banking and financial integration." *The Journal of Finance* 71 (3):1159–1184.

Giroud, Xavier. 2013. "Proximity and investment: Evidence from plant-level data." *The Quarterly Journal of Economics* 128 (2):861–915.

Giroud, Xavier and Holger M Mueller. 2015. "Capital and labor reallocation within firms." *The Journal of Finance* 70 (4):1767–1804.

———. 2019. "Firms' internal networks and local economic shocks." *American Economic Review* 109 (10):3617–49.

Glaeser, Edward, Wei Huang, Yueran Ma, and Andrei Shleifer. 2017. "A real estate boom with Chinese characteristics." *Journal of Economic Perspectives* 31 (1):93–116.

Graham, John, Mark T Leary, and Michael R Roberts. 2014. "How does government borrowing affect corporate financing and investment?" Tech. rep., National Bureau of Economic Research.

Greenwood, Robin, Samuel Hanson, and Jeremy C Stein. 2010. "A gap-filling theory of corporate debt maturity choice." *The Journal of Finance* 65 (3):993–1028.

Greenwood, Robin, Samuel G Hanson, and Jeremy C Stein. 2015. "A comparative-advantage approach to government debt maturity." *The Journal of Finance* 70 (4):1683–1722.

Hadlock, Charles J and Joshua R Pierce. 2010. "New evidence on measuring financial constraints: Moving beyond the KZ index." *The Review of Financial Studies* 23 (5):1909–1940.

Hanley, Kathleen Weiss and Gerard Hoberg. 2010. "The information content of IPO prospectuses." *The Review of Financial Studies* 23 (7):2821–2864.

———. 2012. "Litigation risk, strategic disclosure and the underpricing of initial public offerings." *Journal of Financial Economics* 103 (2):235–254.

Haselmann, Rainer, David Schoenherr, and Vikrant Vig. 2013. "Lending in social networks." In *University of Warwick Working Paper*. Citeseer.

———. 2018. "Rent seeking in elite networks." *Journal of Political Economy* 126 (4):1638–1690.

Hsieh, Chang-Tai and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly journal of economics* 124 (4):1403–1448.

Huang, Jiekun and Darren J Kisgen. 2013. "Gender and corporate finance: Are male executives overconfident relative to female executives?" *Journal of financial Economics* 108 (3):822–839.

Huang, Yi, Marco Pagano, and Ugo Panizza. 2016. "Public debt and private firm funding: Evidence from Chinese cities." Tech. rep., Graduate Institute of International and Development Studies Working Paper.

———. 2020. "Local Crowding-Out in China." *The Journal of Finance* 75 (6):2855–2898.

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2017. "The housing boom and bust: Model meets evidence." Tech. rep., National Bureau of Economic Research.

Kaplan, Steven N and Luigi Zingales. 1997. "Do investment-cash flow sensitivities provide useful measures of financing constraints?" *The Quarterly Journal of Economics* 112 (1):169–215.

Karolyi, Stephen Adam. 2018. "Personal lending relationships." *The Journal of Finance* 73 (1):5–49.

Kaustia, Markku and Samuli Knüpfer. 2008. "Do investors overweight personal experience? Evidence from IPO subscriptions." *The Journal of Finance* 63 (6):2679–2702.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen. 2011. “The effects of quantitative easing on interest rates: channels and implications for policy.” Tech. rep., National Bureau of Economic Research.

———. 2015. “The impact of treasury supply on financial sector lending and stability.” *Journal of Financial Economics* 118 (3):571–600.

Kumar, Alok and Charles MC Lee. 2006. “Retail investor sentiment and return comovements.” *The Journal of Finance* 61 (5):2451–2486.

Lamont, Owen, Christopher Polk, and Jesús Saaá-Requejo. 2001. “Financial constraints and stock returns.” *The Review of Financial Studies* 14 (2):529–554.

Li, Feng. 2006. “Do stock market investors understand the risk sentiment of corporate annual reports?” *Available at SSRN 898181* .

Li, Hongbin, Lingsheng Meng, Qian Wang, and Li-An Zhou. 2008. “Political connections, financing and firm performance: Evidence from Chinese private firms.” *Journal of Development Economics* 87 (2):283–299.

Liu, Chang and Wei Xiong. 2018. “China’s Real Estate Market.” Working Paper 25297, National Bureau of Economic Research. URL http://www.nber.org/papers/w25297.

Liu, Laura Xiaolei, Yuanzhen Lyu, and Fan Yu. 2017. “Implicit government guarantee and the pricing of Chinese LGFV debt.” *Claremont McKenna College Robert Day School of Economics and Finance Research Paper* (2922946).

Loughran, Tim and Bill McDonald. 2011. “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks.” *The Journal of finance* 66 (1):35–65.

———. 2013. “IPO first-day returns, offer price revisions, volatility, and form S-1 language.” *Journal of Financial Economics* 109 (2):307–326.

Loughran, Tim and Jay Ritter. 2004. “Why has IPO underpricing changed over time?” *Financial Management* :5–37.

Lowry, Michelle, Roni Michaely, and Ekaterina Volkova. 2017. “Initial public offerings: A synthesis of the literature and directions for future research.” *Forthcoming Foundations and Trends in Finance* .

———. 2020. “Information Revealed through the Regulatory Process: Interactions between the SEC and Companies ahead of Their IPO.” *The Review of Financial Studies* 33 (12):5510–5554.

Lowry, Michelle, Micah S Officer, and G William Schwert. 2010. “The variability of IPO initial returns.” *The Journal of Finance* 65 (2):425–465.

Malmendier, Ulrike and Geoffrey Tate. 2005. “CEO overconfidence and corporate investment.” *The Journal of Finance* 60 (6):2661–2700.

Matsa, David A and Amalia R Miller. 2013. “A female style in corporate leadership? Evidence from quotas.” *American Economic Journal: Applied Economics* 5 (3):136–69.

Mitchell, Mark L and Erik Stafford. 2000. "Managerial decisions and long-term stock price performance." *The Journal of Business* 73 (3):287–329.

Ritter, Jay R. 1991. "The long-run performance of initial public offerings." *The Journal of Finance* 46 (1):3–27.

Ritter, Jay R and Ivo Welch. 2002. "A review of IPO activity, pricing, and allocations." *The Journal of Finance* 57 (4):1795–1828.

Ru, Hong. 2018. "Government Credit, a Double-Edged Sword: Evidence from the China Development Bank." *The Journal of Finance* 73 (1):275–316.

Schoenherr, David. 2019. "Political connections and allocative distortions." *The Journal of Finance* 74 (2):543–586.

Schultz, Paul. 2003. "Pseudo market timing and the long-run underperformance of IPOs." *The Journal of Finance* 58 (2):483–517.

Schwert, Michael. 2018. "Bank capital and lending relationships." *The Journal of Finance* 73 (2):787–830.

Shen, Yu, Di Gao, Di Bu, Lina Yan, and Ping Chen. 2019. "CEO hometown ties and tax avoidance-evidence from China's listed firms." *Accounting & Finance* 58 (5):1549–1580.

Shi, Kang, Xie Jin, and Xu Juanyi (Jenny). 2018. "Internal Capital Markets and Export Pricing: Evidence from Chinese Business Groups." .

Sicherman, Nachum, George Loewenstein, Duane J Seppi, and Stephen P Utkus. 2016. "Financial attention." *The Review of Financial Studies* 29 (4):863–897.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing like china." *American economic review* 101 (1):196–233.

Stein, Jeremy C. 1997. "Internal capital markets and the competition for corporate resources." *The Journal of Finance* 52 (1):111–133.

Stolper, Oscar and Andreas Walter. 2019. "Birds of a feather: The impact of homophily on the propensity to follow financial advice." *The Review of Financial Studies* 32 (2):524–563.

Sun, Rui and Ganna Zou. 2021. "Political connection, CEO gender, and firm performance." *Journal of Corporate Finance* 71:101918.

Tate, Geoffrey and Liu Yang. 2015. "Female leadership and gender equity: Evidence from plant closure." *Journal of Financial Economics* 117 (1):77–97.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. "More than words: Quantifying language to measure firms' fundamentals." *The Journal of Finance* 63 (3):1437–1467.

Wang, Hao, Honglin Wang, Lisheng Wang, and Hao Zhou. 2016. "Shadow Banking: China's Dual-track Interest Rate Liberalization." *Working Paper* .

Whited, Toni M and Guojun Wu. 2006. "Financial constraints risk." *The Review of Financial Studies* 19 (2):531–559.

Wu, Jing, Joseph Gyourko, and Yongheng Deng. 2012. "Evaluating conditions in major Chinese housing markets." *Regional Science and Urban Economics* 42 (3):531–543.

———. 2016. "Evaluating the risk of Chinese housing markets: What we know and what we need to know." *China Economic Review* 39:91–114.

Zhou, Yifan. 2020. "Political Leadership and Governance Structure." *Available at SSRN 3692883* .