Essays on Rural-Urban Migration and Firm Performance Differentials in China

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Yayun Pan, M.A.

Washington, DC
April 22, 2016

Essays on Rural-Urban Migration and Firm Performance Differentials in China

Yayun Pan, M.A.

Dissertation Advisor: James Albrecht, Ph.D. and Susan Vroman, Ph.D.

Abstract

This dissertation focuses on some important features of the Chinese economy, a key player in the global economy.

Chapter 2 looks at the rural-urban migration with particular emphasis on the interaction between urban labor and housing markets. The long-held perception of the seemingly unlimited "cheap labor" has started to shift in the past decade with two puzzling observations: rapid growth in migrant wages and substantial migrant labor shortages despite the presence of a still large rural labor reserve. The theoretical and empirical work presented in this chapter is one of the first attempts to quantitatively examine the impact of the rising urban housing cost on migration, migrant labor shortages and migrant wages. Simulation results show that quantitatively the increase in the urban housing cost from 2004 to 2008 accounts for approximately 34 percent of the migrant wage increase observed in that time period and that the migrant labor supply is 11 percent less than it would have been had the housing price been fixed at the 2004 level.

Chapter 3 investigates empirically how ownership type affects the performance, as measured by return on assets (ROA), of private (non-SOEs) versus state-owned enterprises (SOEs) before and after the state sector reform. The unconditional quantile regression results show that SOE ownership has a negative association with the ROA measure across the entire distribution in both pre- and post-reform periods. The effect is especially large at the upper and bottom quantiles. The decomposition analysis suggests that at the aggregate level, differences in the return to characteristics are relatively more important in "explaining" the ROA differentials at the bottom and upper quantiles in both pre- and post-reform periods, indicating that no matter whether before or after the reform the SOE ownership keeps firms in the upper quantiles from achieving top notch performance, and traps firms at the bottom quantiles from moving up the performance distribution. It is worth noting that although the SOE structure effect explains the majority of the ROA differential at the upper quantiles in both time periods, it becomes less of a problem after the reform.

INDEX WORDS: Rural-Urban Migration, Housing Market, Quantile Regression, Decomposition, Firm Performance, Chinese Economy, Dissertations, Theses (academic)

Dedication

The research and writing of this dissertation is dedicated to my dear husband Dr. Hua Chai and my two daughters Deborah Erya Chai and Norah Anya Chai.

Acknowledgements

I am deeply grateful to my main advisors, Professor James Albrecht and Professor Susan Vroman, for their invaluable advice and guidance, their patience, and continual support, inspiration and encouragement. I am also deeply indebted to Professor Axel Anderson, whose generous support and insightful comments have prepared me better for the job market. I also thank Professor Garance Genicot, Professor Roger Lagunoff and Professor Luca Flabbi for their extremely helpful comments and suggestions at various stages of this project.

I also thank my loving husband for being so considerate, tolerant and supportive. Without his help in every possible way, the completion of this dissertation would have been much more difficult. I appreciate the constant love, encouragement and support from my parents, Liqiang Pan and Ying Zhou. To my beloved daughters Erya and Anya, you have been the constant source of strength and inspiration for me throughout this endeavor. All remaining errors within this dissertation are my own.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation focuses on some important features of the Chinese economy, a key player in the global economy. China's rapid growth in the past decades has been largely driven by the industrialization utilizing its abundant supply of rural labor and the vibrant liberalization in the corporate sector, namely the movement from state to private ownership of firms. The respective chapters of my dissertation study these two important aspects of the Chinese economy.

The first chapter of the dissertation looks at the rural-urban migration with particular emphasis on the interaction between urban labor and housing markets. The abundance of the rural-urban migrant workers has been crucial to the emergence of China as the world's manufacturing hub. However, the long-held perception of the seemingly unlimited "cheap labor" has started to shift in the past decade with two puzzling observations: rapid growth in migrant wages and substantial migrant labor shortages despite the presence of a still large rural labor reserve. The theoretical and empirical work presented in this chapter is one of the first attempts to quantitatively examine the impact of the rising urban housing cost on migration, migrant labor shortages and migrant wages. To this end, the canonical Diamond-Mortensen-Pissarides search model is extended to a rural-urban two-sector setup that includes a "housing" good in the consumption basket with endogenously determined quantity and price. To address the quantitative question, I calibrate the model to match a variety of moments in the Rural-Urban Migration in China (2008) dataset. Simulation results show that higher housing cost discourages rural-urban migration and pushes up migrant wages. Quantitatively the model suggests that the increase in the urban housing cost from 2004 to 2008 accounts for approximately 34 percent of the migrant wage increase observed in that time period and that the migrant labor supply is 11 percent less than it would have been had the housing price been fixed at the 2004 level. Migrant labor shortages, however, increase very mildly because of an offsetting reduction in vacancies.

The second chapter of this dissertation investigates empirically how ownership type affects the performance, as measured by return on assets (ROA), of private (non-SOEs) versus state-owned enterprises (SOEs) before and after the state sector reform. I use the 1998-1999 and 2006-2007 data from China's Annual Survey of Industrial Production, an extensive firm-level dataset, and adopt the unconditional quantile regression method proposed by Firpo, Fortin, and Lemieux (2009). The regression results show that the SOE ownership has a negative association with the ROA measure across the entire distribution in both 1998-99 and 2006-07. The effect is especially large at the upper and bottom quantiles. In the decomposition analysis, the overall ROA differential is divided at different quantiles into a composition effect which reflects the extent to which the ROA differential is "explained" by differences in firm characteristics and an SOE structure effect which captures the portion of the ROA differential attributed to differences in the return to the characteristics. I find that at the aggregate level, the SOE structure effect is more important in "explaining" the ROA differentials at the bottom and upper quantiles in both pre- and post-reform periods, indicating that no matter whether before or after the reform the state ownership keeps firms in the upper quantiles from achieving top notch performance, and traps firms at the bottom quantiles from moving up the performance distribution. It is worth noting that although the SOE structure effect "explains" the majority of the ROA differential at the upper quantiles in both time periods, it becomes less of a problem after the reform. These empirical findings help shed light on the role ownership type plays in firm performance, and thus offer insights to policy makers interested in improving corporate governance systems in economies with a substantial presence of SOEs such as China.

Chapter 2

The Impact of Rising Housing Price on Rural to Urban Migration

2.1 Introduction

Massive rural-urban migration since 1980's has been one of the major driving forces of China's rapid industrialization. However, the long held perception of the unlimited cheap labor has started to shift in the past decade, with two observations attracting ever-growing attention: (i) the rapid growth in wages of migrant workers[1]; and (ii) the increasing media coverage of migrant labor shortages[2]. Rural migrant workers enjoyed an unprecedented rise in their wages in the past decade with the average nominal wage going up by 116 percent and average real wage by more than 77 percent. In the period from 2004 to 2008 their real wage rose by 26 percent, or about 6 percent per annum (Figure 2.1). Meanwhile, an increasing fraction of posted job vacancies was reported to be unfilled in the major destination cities. In Guangdong province, China's southern industrial base, migrant labor shortages were estimated at about two million at the beginning of 2004[3].

[1] Migrant workers are those who left their hometowns and worked in other places for more than six months in the year.

[2] Yingping Huang, "A Labor Shortage in China", The Wall Street Journal (August 6, 2004).

[3] http://www.people.com.cn/GB/news/1023/2695434.html

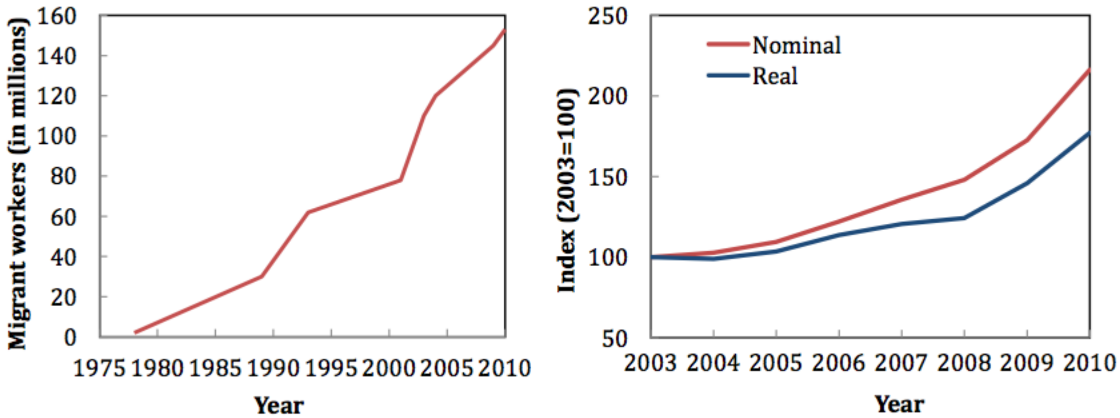

Figure 2.1: Left: Magnitude of Migration 1978-2010; Right: Nominal and Real Migrant Wages, 2003-2010
Source: Data on magnitude of migration are from Li (2008) and NBS. Data on nominal wage in 2003-2009 are from Deng, Knight and Li (2011). The migrant nominal wage in 2010 is from NBS, *Investigation Monitoring Report of Migrant Workers of China (2012)* , from www.stats.gov.cn/tjsj/zxfb/201305/t20130527_12978.html. Real wages for migrants are calculated by the author using the national consumer price index (2003=100) from NBS.

The rapid increase in migrant wages and the reported migrant labor shortages have invited lively public discussions (Wang, Cai, Gao (2005); Tuñón (2006); Deng, Knight, and Li (2011); Liu (2015)). Many scholars view the two observations as simply natural results of depleting rural labor supply and reaching the Lewis turning point, namely the structural change of an economy from excess to limited labor supply. For example, Cai and Wang (2010) provide evidence of a reduced rural labor reserve and suggest that China is getting close to the Lewis turning point. However, other scholars argue that the abundant supply of low-cost labor in the rural areas has not yet been fully absorbed. For example, Golley and Meng (2011) point out that despite some evidence of rising nominal urban unskilled wages between 2000 and 2009, there is little in the data to suggest that this wage increase has been caused by unskilled labour shortages. They find that China still has abundant under-employed workers with very low income in the rural sector.

It is only recently that a few researchers noted that migrant wage increases and labor shortages have occurred at the same time as China experienced a dramatic increase in its urban housing prices. Lu, Zhang, and Liang (2015) propose an explanation that connects the observed wage increases to the rising housing cost in China. In the past decade, real prices of constant quality housing increased by about 225 percent across 35 major cities in China as recorded in Wu, Gyourko, and Deng (2010). According to national statistics, the real average selling price of newly constructed residential buildings[4] increased by 60 percent in the 11 major migrant destination cities from 2004 to 2008. Although published by the National Bureau of Statistics, this number may be subject to underestimation[5]. Figure 2.2 shows how the real housing indices evolve over time in the major destination cities from 1999 to 2013.

[4] The average selling price of newly constructed residential buildings (a.k.a. the Average Selling Price Index) is one of the most popular housing price indices available in China. It is published by the NBS since 1998.

[5] Due to concerns of underestimation, many global real estate companies, such as Centaline Group and DTZ, have built their own house price indices which show larger increase in urban housing prices.

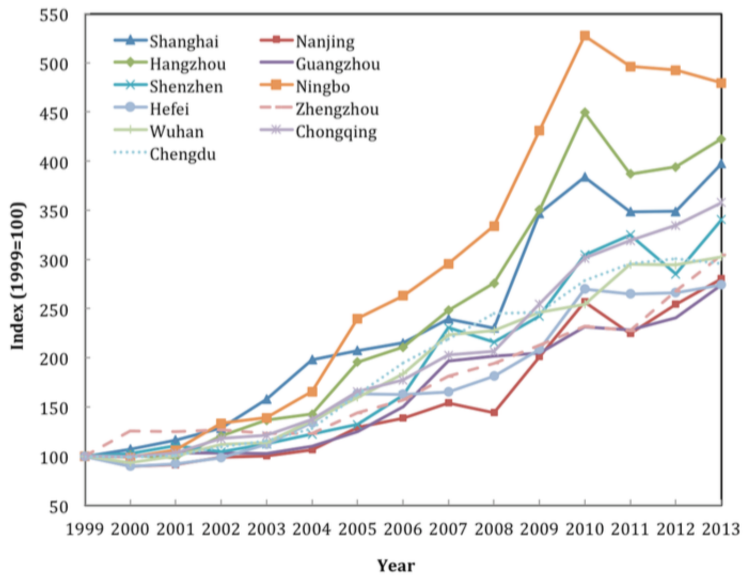

Figure 2.2: Real House Price Indices 1999-2013 (11 cities, Average Selling Price Index)
Source: CEIC China Premium Database; NBS data. Real house price indices are calculated by the author using city-level consumer price indices.

The rise in urban housing prices coincided with increased wages for rural migrant workers both across time and space. From 2003 to 2010, wages for rural migrant workers have shown a strong positive correlation (0.96) with the average residential house selling price index over time. The same relationship is also manifested in the cross-sectional migrant survey data in 2008 and 2009. Among the major destination cities for rural migrant workers, those with high housing prices also offer higher wages for migrant workers. (Figure 2.3)

Based on these observations, some scholars have proposed the hypothesis that rising urban housing prices explain both the increase in migrant wages and the shortage of migrant workers (e.g. Lu, Zhang and Liang 2015).

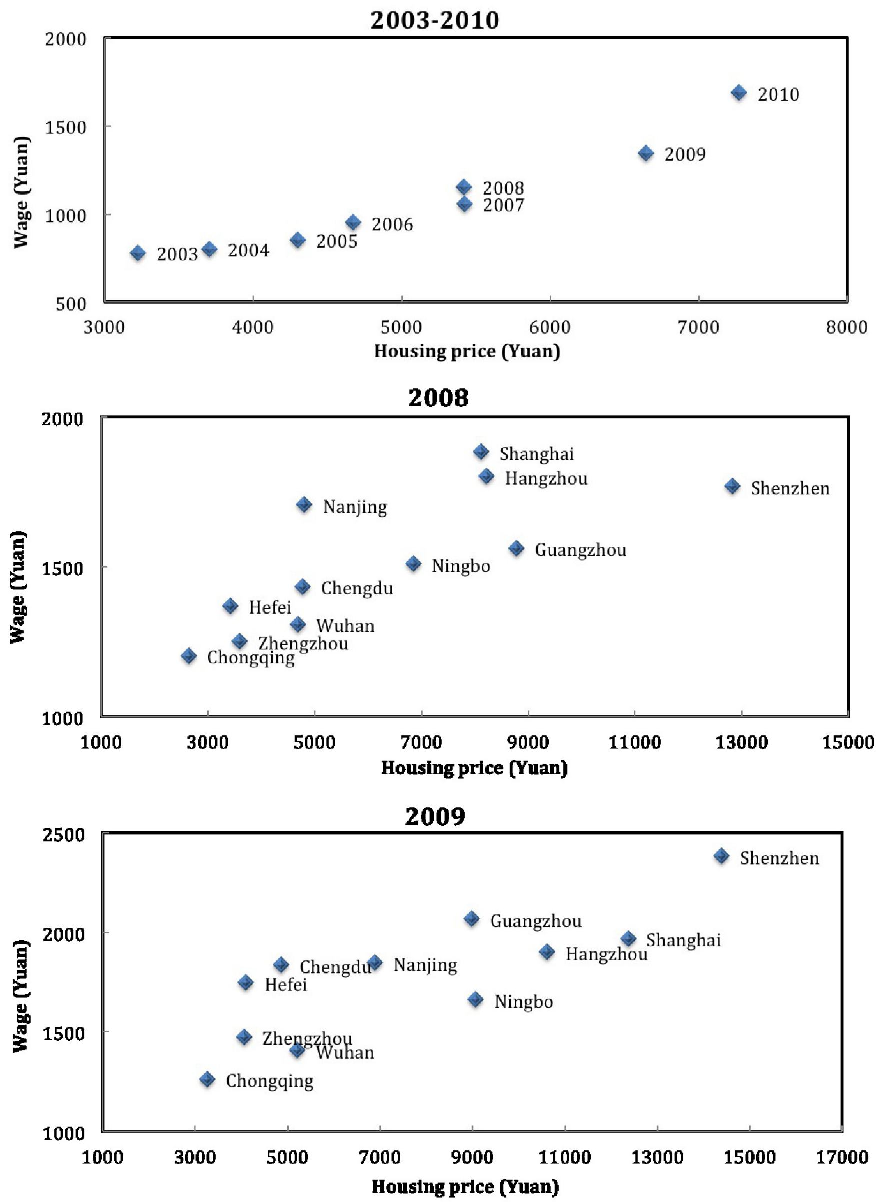

Figure 2.3: Wages and House Price Dynamics across Time and Space

Note: Both wages and house prices are in nominal terms.

Source: Aggregate wage data in the top graph are from the same source as in Figure 1 (right). 2002 city-level wage data in the middle and bottom graphs come from China Household Income Survey 2002 which covers seven migrant destination cities. 2008 and 2009 city-level wage data are from Rural Urban Migration in China (RUMiC 2008 and 2009). All house price data are from CEIC China Premium Database. The aggregate annual house price is the average residential building selling price index (35 city average). City-level house prices are the average residential selling price index for each migrant destination city.

Despite supporting evidence, this hypothesis thus far lacks formal studies. This paper attempts to fill this void by quantitatively examining the impact of rising housing prices on the size of migration, migrant labor shortages and migrant wages in a general equilibrium macro-labor framework. To do this, I extend the standard Diamond-Mortensen-Pissarides[6] search model of the labor market to a rural-urban two sector setup[7]. Rural residents can choose to either stay in the rural sector or to migrate to cities to seek employment there. Migrants and urban residents consume non-housing goods as well as housing services. In the model economy, variations of the housing price are generated by exogenous shocks to urban residents' preference over housing. A preference shock that leads to an increase in the housing price raises the cost of living for rural migrants and hence discourages rural to urban migration. With fewer (and relatively more productive) migrant workers left in the urban sector, their wages go up. As migrant workers become more costly, firms substitute migrant workers with urban workers. The shortage ratio, defined as the number of unfilled vacancies divided by total posted vacancies (filled and unfilled), responds to housing price increases very mildly even though the supply of migrant workers is reduced. The reason is that although there are fewer migrant workers available to fill these vacancies, which tends to raise the shortage ratio, the number of total posted job vacancies decreases as migrant labor becomes more expensive. The two offsetting effects together result in a stable shortage ratio.

The model is calibrated to match a variety of moments in the Rural-Urban Migration in China (2008) dataset. Quantitative results show that the rising housing price is an important factor in explaining the rapid migrant wage growth in 2004-2008. It accounts for about 34 percent of the increase in migrant wages observed in that

[6] Pissarides (2000)

[7] In this paper, the urban sector represents the major coastal provinces where rural workers migrate.

period. Migrant labor supply in the urban sector would be reduced by 10.6 percent in 2008 relative to the hypothetical scenario where the housing price was fixed at the 2004 level.

The rest of the chapter is organized as follows. Section II provides the background on rural-urban migration in China and reviews related literature. Section III develops the quantitative framework. Section IV presents the empirical application including data description, calibration, and results. Section V presents the sensitivity analysis and robustness checks. Section VI concludes.

2.2 Background on Rural-Urban Migration in China and Literature Review

2.2.1 Background on rural-urban migration

China has experienced significant socioeconomic transformations in the past thirty years. Since the late 1970s, by opening up to foreign trade and investment, the country's economic reform has led to a rapid growth in the demand for unskilled labor in urban areas. The slow natural growth in the urban labor force could not meet the increasing demand of the rapid market expansions in these areas. According to Chan and Hu (2003), the annual urban natural growth rates were stable at around 1 percent in 1991-2000. The World Bank estimates that the natural labor growth in China's urban population totaled only 9 percent in the past decade[8]. In response to these urban labor shortages, the household registration system[9], known as the

[8]Source: *Urban China: Toward Efficient, Inclusive, and Sustainable Urbanization - Supporting Report 1-3, Figure 1.4 p.87-88*, The World Bank Group.

[9]China's household registration system, a.k.a. the *hukou* system, was established in 1958 to control labor mobility between rural and urban areas. Individuals are identified as "rural *Hukou*" or "urban *Hukou*", based on their legal residence. There used to be very tight control over workers migrating from rural areas to urban areas. In the recent decades, with increasing labor demand in Chinese coastal cities, the government has gradually eased the

"*hukou*", implemented in the 1950s by the government to restrict internal migration from rural to urban areas, was gradually relaxed. The system has been increasingly relaxed in the 1990s. The accumulating effect of these developments spurred China's massive internal migration, the largest movement of labor force within a country. According to China's National Bureau of Statistics (NBS), the total number of rural migrants reached 166 million in 2013. Since 2002, migrant workers have accounted for more than 40 percent of urban employment (Figure 2.4). Approximately 60 percent of migrant workers are concentrated in manufacturing and construction industries and have been an important force in China's industrialization.

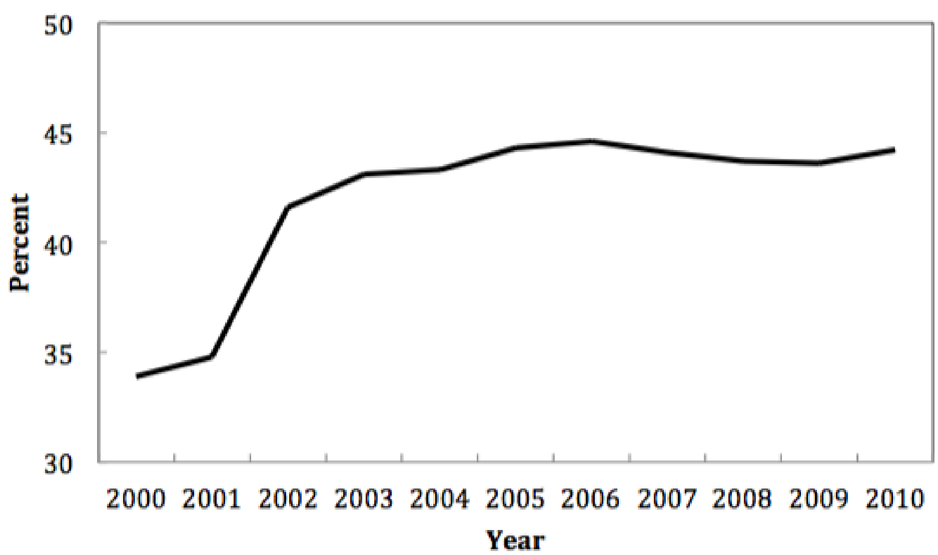

Figure 2.4: Migrant Workers and Urban Employment
Source: Data on migrant workers from 2000 to 2007 come from Table 2 in Cai and Zhao (2009). Data on migrant workers from 2008-2010 come from National Bureau of Statistics of China (NBS), *Investigation Monitoring Report of Migrant Workers of China*, from http://www.stats.gov.cn/tjsj/zxfb/201405/t20140512_551585.html. Data on urban employment are from China Statistic Yearbook (various years) published by NBS.

---

restriction. Since then more and more rural workers choose to migrate but they are still facing very limited access to social benefits in cities.

2.2.2 Literature review

The most relevant literature to this paper explores the causes and impacts of the recent wage increase and labor shortages of unskilled workers in China. Among this strand of literature, Lu, Zhang, and Liang (2015) is the only paper that explores the link between China's rising housing prices and wage increases. Their empirical analysis shows that rapid growth in housing prices deters migration and drives the upward-movements in wages. Different from their methodology, this paper uses a quantitative general equilibrium model capable of performing counterfactual analysis. This paper also explores the effect of housing prices on migrant labor shortages, which is absent in theirs. Knight, Deng, and Li (2011) produce evidence that urban housing cost is among the major factors that deter rural-urban migration and prevent migrant workers from bringing their families with them. But they do not estimate the quantitative impact of rising urban housing prices on the size of migration or labor shortages.

Although there is a large body of literature on the large-scale rural to urban migration in China, for example, Zhao 2003, Cai and Wang 2008, Li 2008, and Taylor 2011, only a few papers utilize structural models to investigate specific issues of interest. The closest one to this paper is Laing, Park and Wang (2005). They modify the basic Harris-Todaro model in a search equilibrium setting to study the impact of household registration (i.e., the *Hukou* system) on urban labor market outcomes. This paper differs from theirs in two respects. First, the focus of this paper is not the household registration system. Institutional restrictions such as the *hukou* system is captured by the transportation cost that rural workers pay when moving from the rural sector to the urban sector. Second, the main contribution of Laing et al. (2005) is that it lays a solid theoretical foundation to study the rural-urban migration in China. In this paper, I extend the standard search model to incorporate a housing market to explore the effect of rising housing prices on migrant labor market outcomes.

Despite the important role of housing to households, there has not been much work studying the relationship between housing prices and labor market outcomes. There is even less empirical research to study the connection between housing cost and labor market outcomes in the Chinese economy. Rupert and Wasmer (2012) studies the effect of housing market frictions on inter-city labor mobility and unemployment using US data. Gabriel et al. (1992) analyze migration flows during the 1980s housing market boom in the U.S. and found that high housing prices in the destination regions deterred migration. Charles, Hurst, and Notowidigdo (2014) find that positive demand shocks significantly increased wages and employment between 2000 and 2007, particularly for less-skilled workers in the U.S. Therefore this paper also contributes to this stream of literature.

2.3 The Model

2.3.1 The Urban Sector

2.3.1.1 Workers

There is measure one of rural residents. A rural resident can choose to either work in the rural sector, i.e., be a rural worker, or seek job opportunities in the urban sector, i.e., be a rural-urban migrant. Rural workers are endowed with one unit of labor service and are homogeneous in performing agricultural work. Rural migrant workers, on the other hand, are heterogeneous with respect to productivity y in urban work, which follows the distribution $G(y)$. In both cases, they supply labor inelastically. There is a one-time cost of transportation $t > 0$ if a worker migrates. The measure of rural workers remaining in the rural sector is l and the measure of rural migrant workers in the urban sector is $1-l$.

In the urban sector, besides the rural migrant workers, there is a measure κ of urban residents who are homogeneous with productivity y_H. They always stay in the urban sector and are not allowed to migrate to the rural sector by assumption. The labor market in the urban sector is segmented. Migrant workers and urban workers produce two different types of intermediate goods. I name the goods produced by migrant workers low-skill goods denoted by q_L, and the goods produced by urban workers high-skill goods denoted by q_H. There is no labor flow between the two markets. The assumption of segmented markets is consistent with the Chinese data which show that migrant workers and urban workers usually work in very different industries. Market segmentation in urban China is also well-documented in the literature in papers such as Meng and Zhang (2001), Knight and Yueh (2009), and Demurger et al. (2009). Following Acemoglu (2001), workers consume a unique final good q, i.e., the non-housing good, which is composed of the two intermediate goods. The technology of production for the non-housing good is

$$q = \left[\mu^{\frac{1}{\eta}} q_L^{\frac{\eta-1}{\eta}} + (1-\mu)^{\frac{1}{\eta}} q_H^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}},$$

where μ is the weight on the low-skill good and η is the elasticity of substitution between the high-skill good q_H and the low-skill good q_L. The formation of the non-housing good allows for substitution between low-skill goods and high-skill goods or equivalently substitution between migrant and urban workers. The intermediate goods q_L and q_H are sold in a competitive market. The equilibrium prices are

$$\begin{aligned} p_L &= \mu^{\frac{1}{\eta}} \left[\frac{q_L}{q}\right]^{-\frac{1}{\eta}}, \\ p_H &= (1-\mu)^{\frac{1}{\eta}} \left[\frac{q_H}{q}\right]^{-\frac{1}{\eta}}. \end{aligned}$$

I normalize the price of the non-housing good to 1. Therefore, in equilibrium the following condition holds

$$\mu p_L^{1-\eta} + (1-\mu)\, p_H^{1-\eta} = 1.$$

The utility function represents constant elasticity of substitution (CES) preferences defined over the consumption of the non-housing good q and the housing good h. All workers share the same functional form except for the weight parameters. I use superscript M and C to denote migrant workers and urban residents, respectively.

$$\Gamma^j(q,h) = \left[\left(\delta^j\right)^{\frac{1}{\varepsilon}} q^{\frac{\varepsilon-1}{\varepsilon}} + \left(1-\delta^j\right)^{\frac{1}{\varepsilon}} h^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}, \qquad j = M, C;$$

where ε is the elasticity of substitution between the housing good and the non-housing good.

In the urban sector, migrants and urban residents conduct sector-specific search due to market segmentation. Job search is governed by a random matching process $M(u,v)$, where u and v are the sector-specific unemployment rate and vacancy rate[10] respectively (i.e., $u \in \left\{u^M, u^C\right\}$ and $v \in \left\{v^M, v^C\right\}$). The matching function follows three standard properties: increasing in both arguments, concave, and homogeneous of degree 1. Let θ be the sector-specific market tightness ($\theta \in \left\{\theta^M, \theta^C\right\}$), which is defined as the vacancy to unemployment ratio, i.e., $\frac{v}{u}$. Denote $m(\theta) = M(1,\theta) = \omega\theta^{1-\phi}$ the sector-specific job finding rate for unemployed workers, where ω is the matching efficiency and ϕ is the matching elasticity[11].

[10] To be consistent with the notation in Pissarides (2001), the rate here represents the measure of vacancies per worker in the labor force.

[11] I assume the two segmented labor markets share the same matching function because of lack of data to accurately calibrate these functions separately. Estimating different matching functions for the two labor markets could be a useful direction for future research. There is also an alternative approach as in Dollar and Jones (2013), where they assume the market for migrant workers is subject to search and matching frictions while the market for urban residents is assumed to be competitive. Qualitative results will not change if I adopt their approach.

All workers in the urban sector are either employed or unemployed. Upon a successful match, wage $w(\cdot)$ is negotiated and determined by Nash bargaining. Job separation is assumed to be exogenous at the poisson rate $\lambda \in \{\lambda^M, \lambda^C\}$. Unemployed workers receive type specific unemployment benefits z ($z \in \{z^M, z^C\}$). There is no on-the-job search. The total income m for workers is

$$m^j = \{ \begin{array}{ll} w^j(\cdot), & \text{if employed} \\ z^j, & \text{if unemployed} \end{array}, \qquad j = M, C.$$

Workers solve their utility maximization problem and the optimal consumption allocation can be derived as a function of goods prices and income. Let $c_D(\cdot)$, $c_D^M(\cdot)$, $c_D^C(\cdot)$, $h_D(\cdot)$ denote the demand functions for the non-housing good, the low-skill good, the high-skill good, and the housing good, respectively, i.e.,

$$\begin{aligned} c_D(p, m) &= \frac{\delta p^\varepsilon}{\delta p^\varepsilon + p(1-\delta)} m, \\ c_D^M(p, p_L, m) &= \mu [p_L]^{-\eta} c(p, m), \\ c_D^C(p, p_H, m) &= (1-\mu) [p_H]^{-\eta} c(p, m), \\ h_D(p, m) &= \frac{1-\delta}{\delta p^\varepsilon + p(1-\delta)} m. \end{aligned}$$

Let $B(p, m)$ denote the optimal consumption bundle of the non-housing good and the housing good, i.e.,

$$B(p, m) = [c_D(p, m),\ h_D(p, m)];$$

and the indirect utility function can be represented by

$$v(p, m) = \Gamma[B(p, m)].$$

Let U and W denote the present discounted value of the expected income stream of an unemployed and an employed worker, respectively. Let V and J be the present discounted value of an unfilled vacancy and a filled vacancy, respectively. All firms and workers in the economy discount the future at the rate $r > 0$. With an infinite time horizon, the steady-state value functions for workers in the urban sector can be written as

$$rW^M(y) = v^M\left(p, m^M\right) + \lambda^M\left[U^M(y) - W^M(y)\right], \tag{2.1}$$

$$rU^M(y) = v^M\left(p, m^M\right) + m\left(\theta^M\right)\left[W^M(y) - U^M(y)\right], \tag{2.2}$$

$$rW^C = v^C\left(p, m^C\right) + \lambda^C\left[U^C - W^C\right], \tag{2.3}$$

$$rU^C = v^C\left(p, m^C\right) + m\left(\theta^C\right)\left[W^C - U^C\right]. \tag{2.4}$$

The non-housing good is the numeraire with its price normalized to one and the relative price of the housing good is p. Equations (2.1) and (2.2) are steady-state value functions for rural migrant workers. Equation (2.1) equates the flow value of an employed migrant worker to the migrant worker's instantaneous utility plus the expected value of the change of state from employed to unemployed. Equation (2.2) indicates that the flow value of an unemployed migrant worker is equal to the migrant worker's flow utility plus the expected value of the change in state from unemployed to employed. Likewise, equations (2.3) and (2.4) are the corresponding steady-state value functions for urban workers.

2.3.1.2 Firms

The setup of firms follows the standard Diamond-Mortensen-Pissarides search model. Firms choose to post one of two types of job vacancies for migrants and urban workers, respectively. All vacancies are identical in every respect and they are filled at the rate $\frac{m(\theta)}{\theta}$ for $\theta \in \left\{\theta^M, \theta^C\right\}$. Each firm can only post one job vacancy and makes its choice so as to maximize the present discounted value of profits. It incurs a one-time cost k to post a job vacancy and the cost is job specific $(k \in \{k_L, k_H\})$. I use subscripts

L and H to denote variables related to the low-skill job vacancy and the high-skill job vacancy, respectively. Firms meet workers at the rate $\frac{m(\theta)}{\theta}$ for $\theta \in \left\{\theta^M, \theta^C\right\}$. Let V and J be the present discounted value of an unfilled vacancy and a filled vacancy, respectively.

The flow value of an unfilled job vacancy is equal to the expected capital gain from a successful match less the cost of job posting, i.e.,

$$rV_L = -k_L + \underset{y}{E}\left\{\frac{m\left(\theta^M\right)}{\theta^M}\max\left[J_L\left(y\right) - V_L, 0\right]\right\} \tag{2.5}$$

$$rV_H = -k_H + \frac{m\left(\theta^C\right)}{\theta^C}\max\left[J_H - V_H, 0\right]. \tag{2.6}$$

The flow value of a filled job is equal to the net return of the job plus the expected capital loss from a possible job break-up, i.e.,

$$rJ_L\left(y\right) = p_L y - w^M\left(y\right) + \lambda^M\left[V_L - J_L\left(y\right)\right], \tag{2.7}$$

$$rJ_H = p_H y_H - w^C + \lambda^C\left[V_H - J_H\right], \tag{2.8}$$

where $p_L y - w^M\left(y\right)$ and $p_H y_H - w^C$ are the net returns obtained from a job vacancy filled by a migrant worker with productivity y and a job vacancy filled by an urban worker, respectively.

2.3.1.3 Wage determination

Wages for migrant and urban workers are determined by Nash bargaining with parameter β (worker's bargaining power), i.e.,

$$\begin{aligned} w^M\left(y\right) &= \arg\max\left[W^M\left(y\right) - U^M\left(y\right)\right]^\beta\left[J_L\left(y\right) - V_L\right]^{1-\beta}, \\ w^C &= \arg\max\left[W^C - U^C\right]^\beta\left[J_H - V_H\right]^{1-\beta}. \end{aligned}$$

Profit maximizing firms post vacancies as long as it is profitable, so in equilibrium $V=0$. The maximization implies that the worker's share of match surplus is the constant β, i.e.,

$$\begin{aligned}(1-\beta)\left[W^M(y)-U^M(y)\right] &= \beta J_L(y), &(2.9)\\ (1-\beta)\left[W^C-U^C\right] &= \beta J_H. &(2.10)\end{aligned}$$

Using equations $(2.1)-(2.4)$, (2.9), and (2.10), simple algebra yields the wage functions. The wage function for a migrant worker with productivity y is

$$w^M(y)=\psi^M p_L y+\left(1-\psi^M\right)z^M, \text{ where}$$

$$\psi^M=\frac{\left[r+\lambda^M+m\left(\theta^M\right)\right]\frac{\beta}{1-\beta}\frac{1}{r+\lambda^M}}{\left[r+\lambda^M+m\left(\theta^M\right)\right]\frac{\beta}{1-\beta}\frac{1}{r+\lambda^M}+\left[\delta^M+\left(1-\delta^M\right)p^{1-\varepsilon}\right]^{\frac{1}{\varepsilon-1}}}.$$

The bargained wage for a migrant worker is a weighted average between the migrant worker's productivity and the worker's unemployment benefits. More productive migrant workers expect higher negotiated wages.

The corresponding wage function for an urban worker is

$$w^C=\psi^C p_H y_H+\left(1-\psi^C\right)z^C, \text{ where}$$

$$\psi^C=\frac{\left[r+\lambda^C+m\left(\theta^C\right)\right]\frac{\beta}{1-\beta}\frac{1}{r+\lambda^c}}{\left[r+\lambda^C+m\left(\theta^C\right)\right]\frac{\beta}{1-\beta}\frac{1}{r+\lambda^C}+\left[\delta^C+\left(1-\delta^C\right)p^{1-\varepsilon}\right]^{\frac{1}{\varepsilon-1}}}.$$

Applying the wage functions to the value functions, we can get the job creation condition for migrant workers and urban residents, respectively

$$\begin{aligned}k_L &= \frac{m\left(\theta^M\right)}{\theta^M}\int_{\underline{y}}^{\infty}\frac{p_L y-w^M(y)}{r+\lambda^M}\frac{g(y)}{1-G\left(\underline{y}\right)}dy; &(2.11)\\ k_H &= \frac{m\left(\theta^C\right)}{\theta^C}\frac{p_H y_H-w^C}{r+\lambda^C}. &(2.12)\end{aligned}$$

In θ, w space, equations (2.11) and (2.12) can be represented by downward-sloping curves, which means that labor demand is downward-sloping.

2.3.1.4 Housing market

The price of housing services is endogenously determined by demand and supply. Demand for housing comes from two sources: migrants and urban residents. Total demand (h_D) is given by

$$H_D = \int_{\underline{y}}^{\infty} h_D^M\left(p, w^M(y)\right)(1-l)\left(1-u^M\right)\frac{g(y)}{1-G\left(\underline{y}\right)}dy + h_D^M\left(p, z^M\right)(1-l)u^M \tag{2.13}$$

$$+h_D^C\left(p, w^C\right)\kappa\left(1-u^C\right) + h_D^C\left(p, z^C\right)\kappa u^C. \tag{2.14}$$

The right-hand side of the equation is the sum of the income of employed migrant workers, unemployed migrant workers, employed urban workers, and unemployed urban workers, respectively.

I assume there are infinitely many perfectly competitive firms in the housing market. They use two inputs, the non-housing good (q) and land (d), to produce the housing good through the following Cobb-Douglas production function

$$h_S = \rho^{-\frac{1}{\xi}} q^{\frac{1}{\xi}} d^{1-\frac{1}{\xi}}, \ \xi > 1.$$

Suppose total land supply is fixed and equal to one. At the aggregate level, the total housing supply is a concave function of the total input of the non-housing good, i.e.,

$$H_S = \rho^{-\frac{1}{\xi}} Q^{\frac{1}{\xi}}, \qquad \xi > 1$$

I can also derive the corresponding cost function as

$$C\left(H_S\right) \equiv Q = \rho H_S^{\xi}, \qquad \xi > 1.$$

The elasticity of housing supply is $\frac{1}{\xi-1}$. When $\xi \to \infty$, the elasticity of housing supply is zero which is equivalent to a simple model where housing supply is assumed fixed. When $\xi = 2$, the cost function is quadratic which means that the marginal cost of production rises linearly with the amount of housing supply. Given a perfectly competitive housing market, the price of housing services should be equal to the marginal cost of production

$$p = MC\left(H_S\right) = \rho \xi H_S^{\xi - 1}. \tag{2.15}$$

2.3.2 The Rural Sector

There is measure l of rural workers remaining in the rural sector. Although rural residents are heterogeneous in productivity once they become migrant workers, I do not assume differences in productivity in the rural sector. Relaxing this assumption does not alter the labor market properties in the urban sector. I use superscript R to denote variables or parameters related to rural workers. Let Y^R be the rural aggregate output and A^R be the rural production technology. Rural labor productivity exhibits diminishing returns to scale and the production function is given by

$$Y^R = A^R l^\gamma, \qquad \gamma \in (0, 1).$$

Assume the rural labor market is perfectly competitive and rent/profit is equally distributed among rural workers. Therefore, each rural worker gets the same share of income $\frac{Y^R}{l}$ to spend on the non-housing good. For simplicity, I abstract away from a meaningful housing market by assuming a unit housing endowment for each rural worker staying in the rural sector. I assume rural workers and migrant workers have the same weight parameter δ^M in their utility functions. The steady-state present discounted value of performing agricultural work in the rural sector is equal to the present discounted value of lifetime utility, i.e.,

$$rW^R = \left[\delta^{M\frac{1}{\varepsilon}} \left(A^R l^{\gamma-1}\right)^{\frac{\varepsilon-1}{\varepsilon}} + \left(1 - \delta^M\right)^{\frac{1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}. \tag{2.16}$$

2.3.3 Steady State Equilibrium

A steady state equilibrium is a collection of prices $\{p_L, p_H, p, w^M(\cdot), w^C\}$, allocations $\{l, \underline{y}, u^M, u^C, \theta^M, \theta^C, H_S\}$, and *value functions* $\{W^M(\cdot)$, W^C, W^R, $U^M(\cdot)$, U^C, $J^M(\cdot)$, J^C, V^M, $V^C\}$ *that satisfy:*

1. *Market tightness* $\{\theta^M, \theta^C\}$ *solve the firms' profit maximization problems in equations* (2.5)*,* (2.6)*,* (2.7)*, and* (2.8)*;*

2. *The housing price* p*, wages* $\{w^M(\cdot), w^C\}$*, market tightness* $\{\theta^M, \theta^C\}$*, and the measure of rural workers remaining in the rural sector* $\{l\}$ *together solve workers' utility maximization problems in equations* (2.1),(2.2),(2.3),(2.4) and (2.16);

3. *No net migration flows between the urban and the rural sector, i.e.,*

$$W^R = -t + U^M(\underline{y}), \tag{2.17}$$

where $\underline{y}$ *is the cutoff productivity that makes the rural worker indifferent between staying in the rural sector and being unemployed in the urban sector. This equation ensures that the present discounted value of staying in the rural sector is equal to the difference between the present discounted value of being unemployed in the urban sector and the transportation cost* t*.*

4. *No net flows of workers between employment and unemployment in the urban sector, i.e.,*

$$\begin{aligned} u^M m(\theta^M) &= (1-u^M)\lambda^M && (2.18)\\ u^C m(\theta^C) &= (1-u^C)\lambda^C && (2.19)\end{aligned}$$

where κ *is the measure of urban residents.*

5. *The housing market clears, i.e.,*

$$H_D = H_S, \tag{2.20}$$

where H_D *and* H_S *are given by equations* (2.13) *and* (2.15)*, respectively.*

6. *Goods markets clear for the non-housing good, the low-skill good, and the high-skill good*[12]*.*

7. *The labor market for rural residents clears. That is, the measure of rural residents working in the rural and urban sectors should sum up to one.*

$$G\left(\underline{y}\right) + (1 - l) = 1, \tag{2.21}$$

$$\text{i.e. } G\left(\underline{y}\right) = l. \tag{2.22}$$

2.4 Calibration

2.4.1 Data

The data are from the longitudinal survey on Rural Urban Migration in China (RUMiC). The project, started in 2008, aims to collect and provide longitudinal information on China's labor markets. It is supported by IZA and conducted by researchers at the Australian National University, the University of Queensland and Beijing Normal University. There are three independent surveys included in RUMiC: the Urban Household Survey (UHS), the Rural Household Survey (RHS) and the Migrant Household Survey (MHS). These surveys record detailed information on individual demographic characteristics, income and expenditures, and work and employment. One of the advantages of RUMiC data is that it covers major migrant destination cities. I use the UHS and MHS data from RUMiC 2008 for calibration in this paper.

[12] *For detailed on the market clearing conditions, please see Appendix 7.3.*

The Urban Household Survey 2008 surveyed 5,005 rural households and 14,695 household members in 19 cities[13]. The Migrant Household Survey 2008 surveyed 5,007 households and 8,446 household members in 15 cities[14]. Urban residents are individuals who possess urban *hukou*. A rural-urban migrant is defined as an individual who has rural *hukou*, but is living in a city at the time of the survey. The sample used for this paper is restricted to the 15 cities common to both urban residents and migrants. All individuals in the sample are aged 16–60. The sample contains 6,380 observations for urban residents and 6,634 observations for rural-urban migrants. Urban workers are on average more educated than migrants. The income gap between migrant workers and urban workers is significant and the average monthly wage of migrant workers in 2008 is only two thirds of that of urban workers. For unemployment duration, the survey asks two retrospective questions about the end of the previous job and the beginning of the new job. Migrant workers on average have a shorter unemployment spell than urban workers. The survey also provides information on how long workers have been working for their current employer. I can thus calculate the employment durations and find a much longer employment duration for urban workers than migrant workers. Table 2.1 provides some descriptive statistics.

	Migrant workers	Urban workers
Sample size	6,634	6,380
Ave. years of schooling	9.1	12.2
Rural *hukou*	Yes	No
Ave. monthly income (Yuan)	1597	2355
Ave. employment spell (Years)	3.7	12.7
Ave. unemployment spell (Years)	0.4	1.0

Table 2.1: Descriptive Statistics

[13] Anyang, Bengbu, Chengdu, Chongqing, Dongguan, Guangzhou, Hangzhou, Hefei, Jiande, Leshan, Luoyang, Mianyang, Nanjing, Ningbo, Shanghai, Shenzhen, Wuhan, Wuxi, Zhengzhou

[14] Bengbu, Chengdu, Chongqing, Dongguan, Guangzhou, Hangzhou, Hefei, Luoyang, Nanjing, Ningbo, Shanghai, Shenzhen, Wuhan, Wuxi, Zhengzhou

2.4.2 Calibration

The parameters to calibrate are the CES utility function parameters (δ^M,δ^C, ε), CES aggregator parameters (μ, η), production function parameters (A^R, γ), matching function parameters (ϕ, ω), worker's bargaining power β, measure of urban residents κ, job separation rates (λ^M, λ^C), unemployment benefits (z^M, z^C), parameters in the productivity distribution function (α, $y_{\min}$), cost of posting vacancies (k_L, k_H), transportation cost t, productivity of urban residents y_H, and housing cost function parameter (ρ, ξ), and discount rate r.

Denote Λ as the parameter space which is defined over the above 24 parameters:

$$\Lambda = \left\{\varepsilon, \mu, \eta, \gamma, \phi, \omega, \beta, \kappa, \lambda^M, \lambda^C, z^M, z^C, r, \alpha, y_{\min}, \xi, k_L, k_H, t, A^R, y_H, \delta^M, \delta^C, \rho\right\}$$

The first 16 parameters are calibrated based on previous studies in the literature or are directly calculated using the RUMiC 2008 dataset. The last eight parameters will be calibrated to match a variety of data moments in the RUMiC 2008 dataset. Due to limited data and the relative lack of previous studies, some parameters can not be calibrated to fit the Chinese economy. In this case, I pick a benchmark standard in the literature for the U.S. economy and then run robustness check using reasonable values of the parameters to make sure that they do not alter the results. Finally, all the parameters here are calibrated with a year as the implicit unit of time.

- $\{\varepsilon, \mu, \eta, \gamma, \phi, \omega, \beta, \kappa, \lambda^M, \lambda^C, z^M, z^C, r, \alpha, y_{\min}, \xi\}$

1. The elasticity of substitution between the housing good and the non-housing good (ε). The literature does not provide a clear standard to calibrate this parameter. Some papers suggest that this elasticity should be close to one (Davis and Heathcote (2005); Piazzesi et al. (2007); Kahn (2008)). Other papers argue that the intratemporal substitutability between the housing and the non-housing good is substantially less than perfect (Flavin and Nakagawa (2004);

Davidoff and Yoshida (2008); Gete (2010)). I set $\varepsilon = 0.5$ and also check $\varepsilon = 0.25$ and $\varepsilon = 0.75$ in the sensitivity analysis.

2. The weight parameter (μ) in the CES aggregator affects the estimated spending share on high-skill goods and low-skill goods. Lacking micro-level data on the composition of consumer expenditure, I set (μ) to 0.5 and run sensitivity analysis on this parameter in Section V.

3. The elasticity of substitution between low-skill goods and high-skill goods (η) essentially captures the substitution between urban workers and migrant workers. Literature studying the U.S. labor market shows that the elasticity of substitution between the skilled labor and the unskilled labor lies between 1 and 2 (Johnson (1997), Autor et al. (1997)). Katz and Murphy (1992) use Current Population Survey (CPS) data and estimate an elasticity of substitution between college and high school labor of about 1.41. Krussel et al. (2000) used macro data and got an estimate of the substitution elasticity between unskilled labor and skilled labor of 1.67 . I set $\eta = 1.5$ in the baseline model and check $\eta = 0.25$ and $\eta = 0.75$ for robustness.

4. The output elasticity in the rural production function (γ). Cao and Birchenall (2013) uses microeconomic farm-level data to estimate factor shares using an aggregate Cobb-Douglas production function. They arrive at an estimated labor share of 23 percent. I follow their work and set $\gamma = 0.23$.

5. Consistent with the search literature on other countries, the matching function takes the most common specification which is Cobb-Douglas with constant returns to scale, i.e., $M(u, v) = \omega u^{\phi} v^{1-\phi}$ (Petrongolo and Pissarides, 2001). It can also be written in terms of market tightness as $m(\theta) = \omega \theta^{1-\phi}$, where ϕ

is the matching elasticity and ω is the matching efficiency. The matching elasticity ϕ is usually estimated to be between 0.4 and 0.6 for the U.S. economy. For example, Mortensen and Nagypal (2007) estimates ϕ at 0.45. Albrecht et al. (2009) uses a matching elasticity of 0.5 for a model economy with an informal sector. In the literature studying the Chinese labor market, Liu (2013) estimates the aggregate matching function using Chinese data. She finds that the matching elasticity of job seekers with respect to unemployment is between 0.5 and 0.8 and an estimated value of 1.07 for the aggregate matching efficiency parameter. I follow Liu (2013) and set $\omega = 1.07$ and $\phi = 0.5$.

6. Wages are set by Nash bargaining. The Hosios condition (Hosios (1990)) requires that for efficiency the value of worker bargaining power is equal to the elasticity of the matching function with respect to unemployment. Since the matching function is Cobb-Douglas, I choose $\beta = 0.5$, i.e., workers and firms equally split the total surplus.

7. According to a recent study by China Data Center (CDC) at Tsinghua University[15], the urban hukou population has been stable around 27.6 percent of the country's total population. Since the measure of rural workers is normalized to one, I back out the measure of urban population (κ) to be 0.38.

8. Job separation rates $\left(\lambda^M, \lambda^C\right)$. Assume that durations of spells in both unemployment (t_U) and employment states (t_E) are exponentially distributed such that the expected duration of an employment spell equals the reciprocal of the separation rate, i.e., $E\left(t_E\right) = \frac{1}{\lambda}$. With data on employment spells, separation rates can be backed out using $\lambda^* = \frac{1}{E(t_E)}$. RUMiC 2008 reports an average

[15] http://zqb.cyol.com/html/2013-11/05/nw.D110000zgqnb_20131105_1-07.htm

employment spell of 3.7 years for migrant workers and 12.7 for urban residents, which implies $\lambda^M = 0.268$ and $\lambda^C = 0.079$.

9. Unemployment benefits $\left(z^M, z^C\right)$. These are the only two parameters that have nominal values. Due to data limitation, I normalize z^M to 1 and set z^C to 1.5 to be consistent with the income ratio between the urban workers and migrant workers. I also check the case in which migrants and urban workers have equal unemployment benefits in the sensitivity analysis.

10. Average annual interest rate (r) is 0.0513[16].

11. I assume that the productivity of migrant workers is Pareto distributed, i.e., $G(y) = 1 - \left(\frac{y_{\min}}{y}\right)^{\alpha}$. Following Guerrieri (2007), I set the shape parameter of Pareto distribution (α) to 3 and the scale parameter $(y_{\min})$ to 0.4.

12. China had a land supply reform in 2004 which established the government-monopoly land supply system. After that, the housing supply elasticity experienced a significant decline. Yan and Wu (2014) use panel data from twenty major Chinese cities and arrive at an estimate of the average housing supply elasticity for China to be 0.5. I use the results from their paper and set $\frac{1}{\xi-1} = 0.5$, i.e., $\xi = 3$.

[16] China Statistic Yearbook published by the National Bureau of Statistics

Parameters	Values	Source
ε	0.5	
μ	0.5	
η	1.5	Johnson (1997)
γ	0.23	Cao and Birchenall (2013)
ϕ	0.5	Liu (2013)
ω	1.07	Liu (2013)
β	0.5	Hosios (1990)
κ	0.37	NBS
λ^M	0.268	RUMiC 2008
λ^C	0.079	RUMiC 2008
z^M	0	RUMiC 2008
z^C	0	RUMiC 2008
r	0.0513	NBS
α	3	Guerrieri (2007)
y_{min}	0.4	Guerrieri (2007)
ξ	3	Yan and Wu (2014)

Table 2.2: Calibrated Parameters

- $\{k_L, k_H, y_H, A^R, t, \delta^M, \delta^C, \rho\}$

I use the method of moments to estimate the remaining seven parameters, i.e., costs of job posting (k_L, k_H), productivity of high-skill workers (urban residents) y_H, rural production technology parameter A^R, transportation cost t, CES utility function weight parameters (δ^M, δ^C) and marginal cost of housing supply ρ. The algorithm consists of the following steps:

Step 1: Guess the vector of parameters $\{k_L, k_H, y_H, A^R, t, \delta^M, \delta^C, \rho\}$;

Step 2: Solve the model and calculate moments from the model;

Step 3: Compare the moments from the model to those from the data;

Step 4: Update the parameter vector based on the above;

Step 5: Iterate until data and model moments converge.

I use RUMiC 2008 to generate the target data moments. The dataset provides detailed information on income and employment which allows me to calculate the unemployment rates for urban residents and migrants, the income ratio between urban residents and rural residents, the income ratio between urban residents and migrants, unemployment durations for urban residents and migrants, and housing expenditure shares for urban residents and migrants. Table 2.3 summarizes the calibrated parameters. The calibrated cost of job posting for urban workers is five times of that for migrant workers. The gap reflects the fact that migrants and urban residents work in different industries and the hiring process for urban workers is usually more sophisticated and costly than that for migrants. The calibrated value of productivity of urban workers is consistent with the wage premium observed in the dataset. The transportation cost in the model works as a relocation fee which captures all the expenditures associated with migration. The calibrated value accounts for about half of the mean wage of migrant workers.

Parameters	Description	Values
k^M	Cost of job posting for migrant workers	0.41
k^C	Cost of job posting for urban workers	2.10
y_H	Productivity of urban workers	7.02
A^R	Rural production technology	0.76
t	Transportation cost	1.12
δ^M	Weight parameter in the utility function	0.78
δ^C	Weight parameter in the utility function	0.86
ρ	Housing supply cost function coefficient	3.22

Table 2.3: Moment Estimation

To check the goodness of fit of the model, I compare a series of moments generated by the model and those observed in the data. The moments include unemployment rates, market tightness, unemployment durations, the measure of rural workers, wage ratios, and housing expenditure shares. The model fits well overall and the results are summarized in Table 2.4.

Variable	Model	Data
Unemployment rates		
u	0.09	0.05
u^M	0.13	0.10
u^C	0.07	0.09
Market tightness		
θ	1.85	0.85
Unemployment durations		
t_U^M	0.57	0.38
t_U^C	1.04	1.00
Measure of rural workers		
l	0.74	0.75
Wage ratios		
$\frac{w^C}{w^M}$	1.48	1.47
$\frac{w^M}{w^R}$	2.41	2.41
Housing expenditure shares		
$\frac{1-\delta^M}{\delta^M p^{\varepsilon-1}+1-\delta^M}$	0.21	0.21
$\frac{1-\delta^C}{\delta^C p^{\varepsilon-1}+1-\delta^C}$	0.13	0.13

Table 2.4: Model vs. Data Moments

2.4.3 Simulation Result

A supply-side shock that reduces housing supply or a demand-side shock that increases housing demand could be a potential candidate to generate changes in the housing price. However, a decreasing housing supply does not match the Chinese data as shown in Figure 2.5. The urban housing supply measured by the floor space completed each year has been increasing at both the national level and in the eastern coastal area.

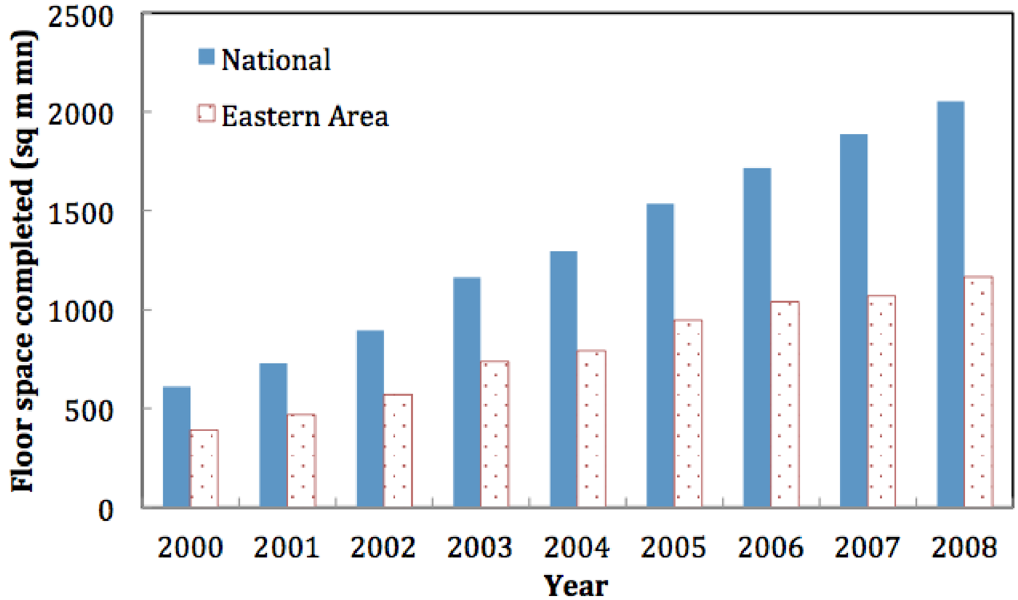

Figure 2.5: Floor Space Completed in 2008-2008
Source: CEIC China Premium Database; NBS data

The use of a demand-side shock is more plausible. In particular, I feed into the model an exogenous preference shock to affect housing prices. The assumption of an increase in the preference towards housing captures the two driving forces of China's housing boom in the past decade, i.e., the real-estate bubbles and the increasing cultural preferences towards housing in the marriage market. The preference shock works through the weight parameter in urban workers' utility function. A low preference towards housing (δ^C) weakens housing demand and brings down housing prices; a high preference towards housing leads to higher housing prices. Alternatively, a demand side shock on the migrant workers such as an increase in the transportation cost would also generate increases in the urban housing price. However, this is against the Chinese data as the housing boom was mainly driven by the increasing demand from the urban residents. In the simulations, I compare only the steady states between the baseline and the simulated scenario.

2.4.4 Migration and migrant wages

Given the exogenous increase in urban residents' preference towards housing, housing prices increase, which reduces the utility of migrant workers staying in the urban sector. Fewer rural workers choose to migrate. This leads to a decline in total production of low-skill goods. Consequently, the price for low-skill goods increases whereas the price for high-skill goods falls. Firms and workers then renegotiate wages through Nash bargaining, resulting in an increase in wages for migrant workers and a decrease for urban workers. Analytically, what will happen to the unemployment rate for migrant workers as a result of rising housing cost is uncertain because it has two offsetting effects. First, a smaller size of migrant labor supply in the urban sector contributes to a higher job finding rate and job matches are more likely to form, which tends to reduce the unemployment rate. Second, as migrant workers become more costly, firms substitute out migrant workers with urban residents in producing the non-housing good. Fewer vacancies for migrant workers lead to a smaller job finding rate and hence tend to raise the unemployment rate. The net effect depends on which of the two effects is stronger.

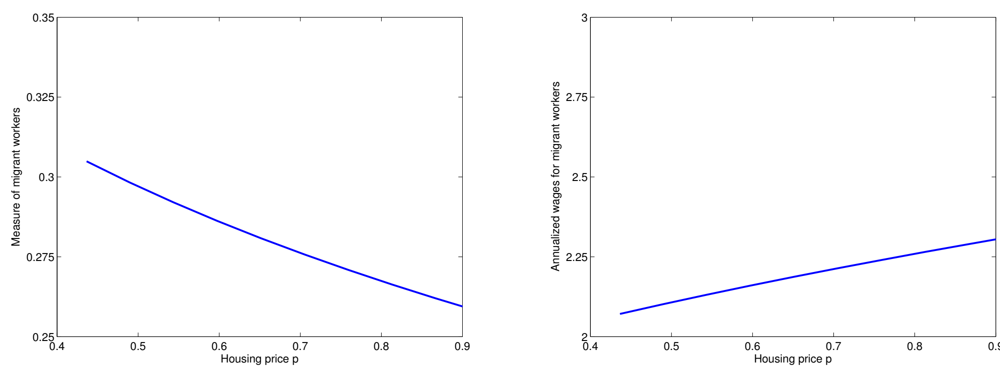

Figure 2.6: Responses of Size of Migration (left) and Migrant Wages (right) to Housing Price Changes
Note: The number of migrant workers is multiplied by 10 to be consistent with the scale in the migrant wage chart.

2.4.5 Migrant labor shortages

Migrant labor shortages are measured by the shortage ratio of rural migrant workers (ϑ). It is defined as the measure of unfilled migrant-specific job vacancies to the measure of total migrant jobs, i.e.,

$$\vartheta \equiv \frac{v^M (1-l)}{v^M (1-l) + (1-u^M)(1-l)} = \frac{v^M}{v^M + 1 - u^M}.$$

The shortage ratio is affected by the housing price through two channels: the pool of available migrant workers (migrant labor supply in the urban sector) and the demand for migrant workers (vacancies). The former raises the shortage ratio as the rising housing cost leads to fewer migrant workers in the urban sector. The vacancy channel works in the opposite direction. Firms adjust their vacancy postings as migrant workers become more costly and harder to find. Quantitatively, these two effects offset each other, resulting in a mild increase in the shortage ratio. (Figure 2.7)

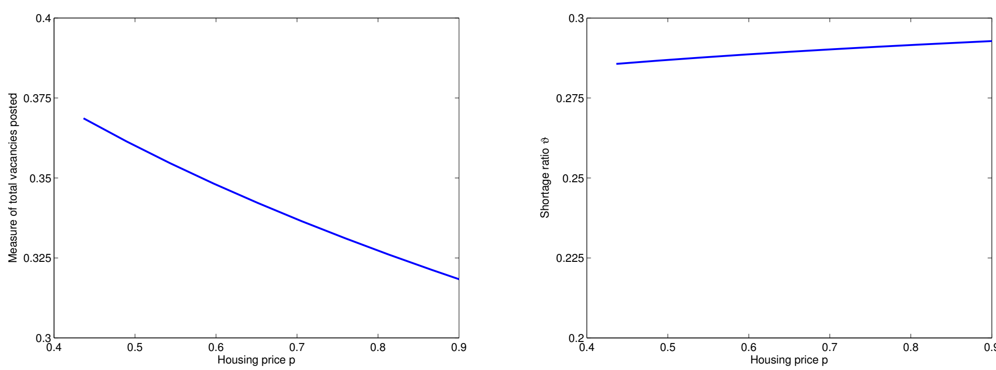

Figure 2.7: Responses of Total Job Vacancies Posted (left) and Shortage Ratio to Housing Price Changes
Note: Total job vacancies posted include both unfilled and filled.

According to China's National Bureau of Statistics, the real average residential house selling price increased by approximately 60 percent for the 15 migrant destination cities[17] from 2004 to 2008[18]. I use a preference shock that produces the same amount of increase in the housing price as observed in the data from 2004 to 2008. The model indicates that there would be 10.6 percent fewer (approximately 15 million) rural migrant workers in the urban sector relative to the hypothetical case in which the housing price stays at the 2004 level. This is because higher housing prices push up the cost of living and discourage rural residents from migrating to the city. As the supply of migrant workers dwindles, and since their average productivity improves as less productive workers leave the urban sector, the average wage for migrant workers would increase by about 7.7 percent, which accounts for 34 percent of the migrant wage growth observed in data in that time period.[19] This finding suggests that the rising housing price is an important driver of migrant wage growth. Because of a strong response from total vacancies posted, the shortage ratio, increases only by 1.7 percentage points. The results are reported in Table 2.5 below.

Variables	Price increased by 60%
$1-l$	-10.6%
w^M	7.7%
ϑ	1.7%

Table 2.5: Simulation Results

2.5 Sensitivity Analysis

This section checks the robustness of the quantitative results with respect to the parameters, i.e., elasticity of substitution between the housing good and the non-housing good ε, weight parameter μ, the elasticity of substitution between the low-skill good and the high-skill good η, and the unemployment benefits z^M and z^C.

[17]The same 15 cities as in RUMiC 2008

[18]Sources: CEIS; National Bureau of Statistics of China

[19]Real wages for migrant workers increased by about 22 percent from 2004 to 2008.

In each scenario, a preference shock is used to produce a 60 percent increase in the housing price which is the same amount of increase as in the baseline model.[20]

I demonstrate two alternative cases where the elasticity of substitution between the housing and the non-housing good is increased and decreased by 50 percent. In the baseline model with ε of 0.5, a 60 percent increase in housing price would raise migrant wages by 7.7 percent, reduce migrant labor supply by 10.6 percent and increase the shortage ratio by 1.7 percentage points as shown in the middle column in Table 2.6. The change in ε does not alter the sign on variables for migrant workers. All migrant labor market variables change monotonically and the magnitude of the variables increases as ε increases. In other words, the impact of rising housing cost on migrant labor market outcome is larger when the housing and the non-housing good become more substitutable.

ε	0.25	0.5	0.75
$1-l$	-9.6%	-10.6%	-11.7%
w^M	6.8%	7.7%	8.6%
ϑ	1.6%	1.7%	1.8%

Table 2.6: Varying ε

I check $\eta = 1.41$ and $\eta = 1.67$ which are used in Katz and Murphy (1992) and Krussel et al. (2000), respectively. Changing weight parameter μ yields moderate changes in the quantitative results. All three variables of interest exhibit monotonic changes as the elasticity of substitution between the low-skill good and the high-skill good increases. Larger effects are observed when the low-skill good and the high-skill good are more substitutable. (Table 2.7)

[20]Other values of these parameters have also been checked by the author and the results are available upon request.

η	1.41	1.5	1.67
$1-l$	-10.2%	-10.6%	-11.2%
w^M	7.6%	7.7%	8.0%
ϑ	1.5%	1.7%	1.9%

Table 2.7: Varying η

Compared to the other parameters, the simulation results are relatively more sensitive to the weight parameter μ in the production function. A lower μ translates into a smaller share of spending on low-skill goods. Because of this, there is less incentive to substitute migrant workers with urban residents when the housing price increases and the low-skill good becomes more expensive. Therefore, a lower μ leads to higher wages for migrant workers who produce the low-skill goods whereas a higher μ contributes to lower wages. The changes in the weight parameter in the production function do not alter the sign of the variables of interest. It yields monotonic changes in migrant wages but not the other two variables. The lower the weight parameter μ, the larger impact of housing cost on migrant wages. The size of migration and the shortage ratio as a result of rising housing cost respond stronger when firms put asymmetric weights on the low-skill and the high-skill goods. Results are shown in Table 2.8 below.

μ	0.25	0.5	0.75
$1-l$	-12.0%	-10.6%	-11.0%
w^M	9.5%	7.7%	6.3%
ϑ	2.8%	1.7%	1.8%

Table 2.8: Varying μ

In the baseline model, I normalize the unemployment benefits of migrants and set the unemployment benefits for urban residents by referring to their income ratio from the data. In the alternative case, I suppose that migrants and urban residents enjoy the same level of unemployment benefits. Such change produces no significant changes on the resulting migrant labor market outcomes. (Table 2.9)

(z^M, z^C)	(1,1.5)	(1,1)
$1-l$	-10.6%	-10.6%
w^M	7.7%	7.7%
ϑ	1.7%	1.7%

Table 2.9: Varying z^M and z^C

2.6 Conclusion

This paper investigates quantitatively the impact of the rising housing cost on the size of migration, equilibrium migrant wages and migrant labor shortages in China. I extend the standard Diamond-Mortensen-Pissarides search model to incorporate rural and urban sectors. The main finding of the paper is that the rising housing cost is an important factor in explaining the rapid growth in migrant wages over the last decade in urban China. It accounts for about 34 percent of the increase in migrant wages observed in that period. High housing cost also discourages rural to urban migration. If housing prices increase by 60 percent, i.e., the same amount increase as in the average residential selling price in 2004-2008, migrant labor supply in the urban sector would be reduced by 10.6 percent in 2008 relative to the hypothetical scenario in which the housing price was fixed at the 2004 level. The effect of the rising housing cost on migrant labor shortages is rather small though. This is because the shortages are not only affected by unemployment but also through vacancy posting. Vacancies adjust in the opposite direction which mitigates the impact from a smaller pool of available migrant workers, leading to a very mild increase in the shortage ratio.

Chapter 3

Quantile Regression and Decomposition on Firm Performance Differentials: Evidence from Chinese Manufacturing Firms

3.1 Introduction

The relationship between ownership structure and firm performance is an important issue in the study of corporate governance. This paper examines empirically how ownership type (i.e., state-owned enterprises (SOEs) vs. private firms (non-SOEs)) affects firm financial performance using microdata from China.

China undertook substantial reforms in its industrial state sector starting in the 1980s. A new stage of reforms started in 1999 which aimed to "grasp the large and let go of the small" (1999 reform from hereon), meaning to keep the large state-owned enterprises under state control while privatizing or closing the small ones. By 2007 the reform reached its end and a large number of state-owned firms which used to dominate China's industrial sector was shutdown or privatized. Meanwhile, the large SOEs were transformed through restructuring to be limited liability corporations but with its governance still under the control of the Chinese government.[1] Figure 3.1 shows the diminishing share of state-owned enterprises in the manufacturing industry over the period 1998-2007. The number of SOEs is represented by the unfilled bars with red outlines. In 1998 when the reform just started, more than a third of the manufacturing firms were SOEs, but by 2007 the number had shrunk to be only about

[1]The Chinese government achieves its control over the restructured SOEs through executive appointment and the determination of executive compensation.

5 percent. The number of non-SOEs represented by the blue bars, on the contrary, were steadily growing with its total number tripled by 2007. Figure 3.2 shows that the SOE share in the gross industrial output had also seen a significant decline, from 45 percent in 1998 to less than 25 percent in 2007. Its share was more than halved between 1998 and 2007.

Over the years of the reform, the average return on assets (ROA), an indicator of firm performance, rose for both SOEs and non-SOEs. In particular, the average ROA for SOEs even increased from negative to positive as shown in Figure 3.3.

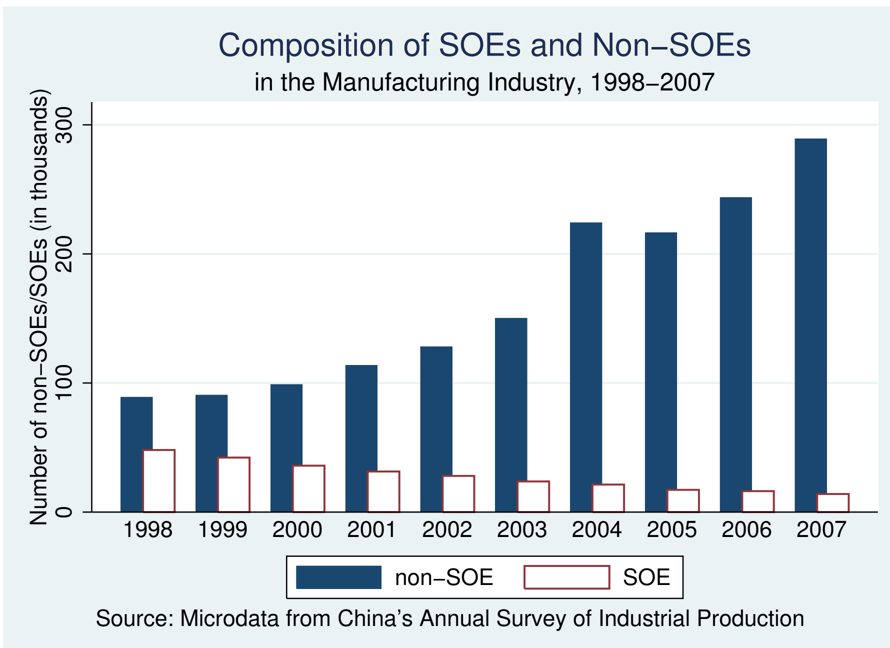

Figure 3.1: Composition of Manufacturing Industry by Ownership

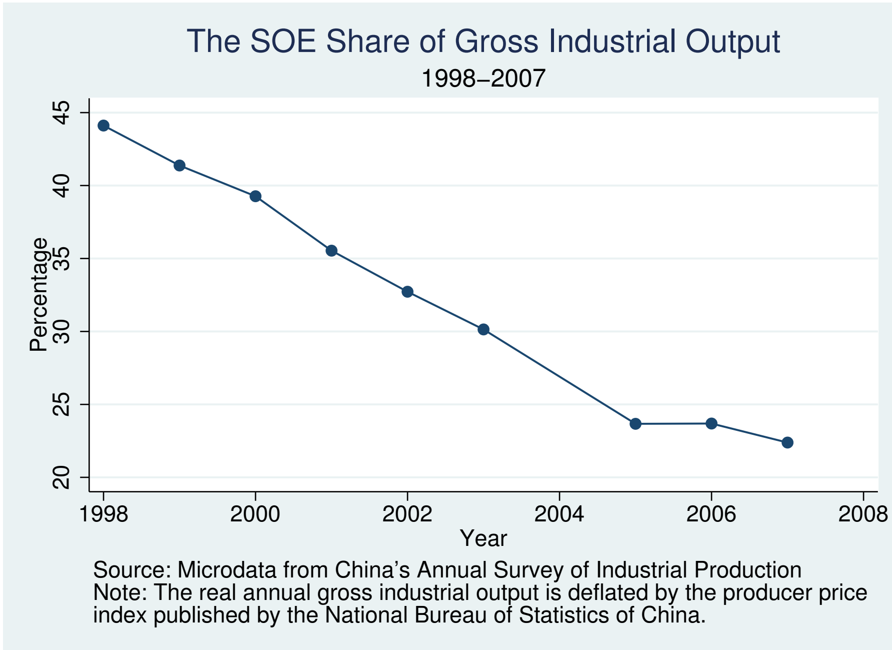

Figure 3.2: The Shrinking SOE Share of Gross Industrial Output

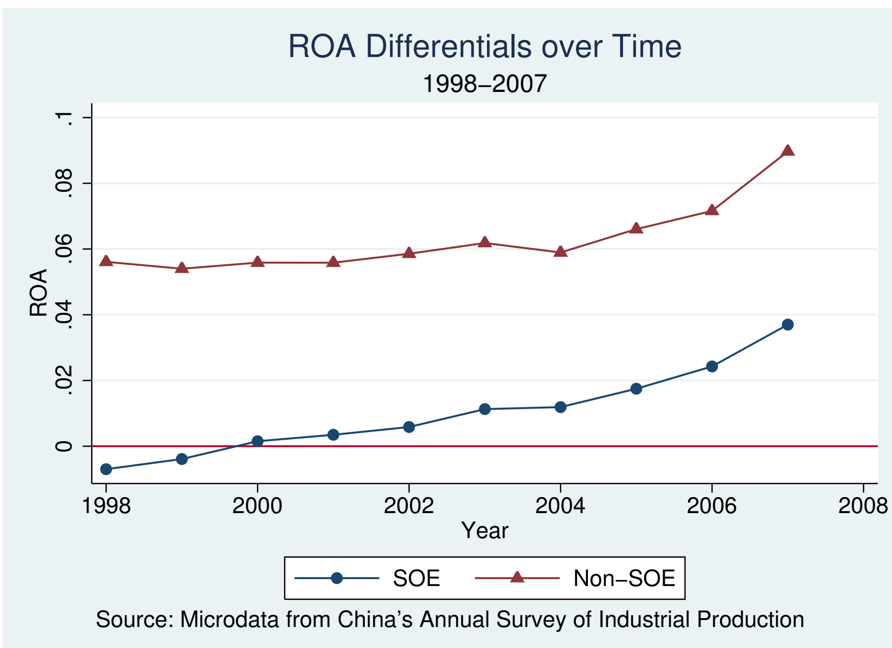

Figure 3.3: Average Return on Assets by Ownership Type, 1998-2007

A second observation from the chart is that in spite of fast growth, there was little convergence in the average ROA between SOEs and non-SOEs in the period of 1998-2007. Many existing studies have been done to investigate the relationship between ownership structure and firm performance, especially for emerging markets. In the context of China, there has been a growing literature studying the performance differentials between SOEs and non-SOEs since the beginning of China's industrial reforms in the mid-1980s. These studies mainly focus on the average performance differentials and have reported ambiguous results. Some scholars find negative correlations between government shareholding and corporate performance (Xu and Wang, 1999; Qi, Wu and Zhang, 2000; Sun and Tong, 2003; Bai, Liu, Lu, Song, Zhang, 2004; Li, Sun, Zou, 2009). Others report positive correlation, for example Chen and Gong (2002). Li, Sun, and Zou (2009) attribute the conflicting results to the inappropriate use of the least squares method. They suggest that it is more reasonable to employ the method of quantile regressions as it covers the full spectrum of the performance distribution and allows a richer characterization of the data. It is conceivable that state ownership may only have an impact on the financial prospects of a certain subset of firms (e.g., bad performing firms, or mediocre firms, or good performing firms).

As a matter of fact, what makes the 1999 reform different from previous rounds of China's industrial reforms is that it had a special focus on the very "big" firms and the very "small" firms, namely firms at the upper and lower quantiles of the performance distribution. Adopting quantile regressions therefore provides more insights on the impact of ownership type throughout the reform periods.

This paper uses detailed firm-level data and utilizes the unconditional quantile regression method (Firpo et al., 2007; Firpo et al., 2009) to achieve two goals: 1) to decompose the performance differentials and to investigate the effects of ownership on firms at different points of the performance distribution, and 2) to compare and document the different roles ownership types play in firm performance before and after the reform.

At the aggregate level, the performance differential is mainly attributed to different firm characteristics for firms in the middle range of the ROA distribution. However, for firms with either very good or very bad performance, it appears that differences in the return to characteristics (i.e. the SOE structure effect) are more important to "explain" the accelerating ROA differentials below the 10th quantile and beyond the 80th quantile. These findings imply that both before and after the 1999 reform, 1) SOE ownership acts as an impediment for firms in the upper quantiles from attaining top tier performers in the manufacturing industry, and 2) it is an important factor in explaining why some SOEs are trapped in the bottom quantiles. The comparison of decomposition results for pre- and post-reform periods shows that although the SOE ownership is a hurdle in achieving better performance in both time periods, it is noteworthy that the SOE structure effect becomes less important in "explaining" ROA differential at the upper quantiles in 2006-07.

3.2 Literature Review

This paper is related to the literature that explores the relation between ownership structure and firm performance. This strand of literature has been growing and developing based on the pioneering works done by Berle and Means and Coase in the 1930s (Berle and Means, 1932; and Coase, 1937). More recently, many scholars argue that state ownership in competitive markets is less efficient than private ownership, mostly because of government's asymmetric information, higher transaction cost, and deviation of its decision-making away from profit maximization (Boycko, Shleifer and Vishny, 1996; and Dewenter and Malatesta, 2001). There has also been a larger number of studies on the association between state ownership and firm performance in the emerging markets, for example China (Xu and Wang, 1999; Qi, Wu and Zhang, 2000; Sun and Tong, 2003; Bai, Liu, Lu, Song, and Zhang, 2004; Li, Sun, and Zou, 2009). The most relevant paper to my paper is Li, Sun and Zou (2009). They use quantile regressions to assess the effect of ownership concentration on firm performance with a sample of 643 public firms in China and find a significantly negative relationship in the upper quantiles of the performance distribution. This paper differs from theirs in three respects. First, instead of ownership concentration[2], this paper focuses on ownership identity, i.e., state owned vs. private owned. Secondly, this paper relies on the method of unconditional quantile regressions (Firpo et al., 2007; Firpo et al., 2009) which allows for a computationally efficient decomposition analysis which is not conducted in their paper. Last but not least, the data used in this paper is from a census which includes almost all the manufacturing firms in China.

In the ownership literature, a common debate is whether the ownership structure is endogenous. The ownership endogeneity argument was first put forward in Demsetz (1983) which suggests that ownership structure is determined such that "various cost advantages and disadvantages are balanced to arrive at an equilibrium". It is later supported by a few empirical studies. For example, Himmelberg et al. (1999) use US panel data and find that insider ownership is endogenous to performance. Gugler and Weigand (2003), however, argue that although managerial ownership is endogenous for US firms, the largest shareholder affects performance exogenously. In the context of China's SOE reform, of which privatization is a core component, one may argue that the government strategically chooses which firms to privatize based on firms' characteristics such as size and therefore ownership type would be endogenous. How-

[2] Ownership concentration refers to the amount of stock owned by individual investors and large-block shareholders.

ever, this paper views the ownership type as exogenous because the two time periods: 1998-1999 (pre-reform) and 2006-2007 (post-reform) saw few privatization attempts and firm ownership type was relatively stable. Moreover, even if some endogeneity is present, this paper, as an accounting exercise, is still useful to identify which portion of the performance difference between SOEs and non-SOEs at different quantiles can be "explained" by ownership type. Lastly, to my knowledge, Frolich and Melly (2007) is the only paper that tackles the endogeneity problem with instrumental variables in unconditional quantile regressions. However, their method requires a qualified instrument which is not easy to find for ownership type in this paper's context. Even with a qualified instrumental variable, it would be computationally challenging to conduct the decomposition analysis.

This paper is also connected to studies on China's industrial reforms. The empirical work done by Yusuf, Nabeshima, and Perkins (2005) uses firm-level survey data and shows that foreign ownership, reformed SOEs, and non-SOE ownership all enhance productivity. Using the same survey data as Xu et al. (2005), Lin and Zhu (2001) focus on the effectiveness of shareholder reform with a special focus on survey responses. The authors report that 34 percent of the respondents to the survey identified improved internal management mechanisms, 23 percent claimed clearer property rights, and 11 percent indicated that restructuring had led to a significant improvement in performance. Jefferson and Su (2005) studies the performance impact of conversion on China's SOEs and their results provide empirical evidence supporting the conventional wisdom that shifting firms toward private control leads to improved firm performance.

The rest of the paper is organized as follows. Section 3 describes the data and introduce the variables. Section 4 illustrates the unconditional quantile regression model and presents regression results. Section 5 carries out the decomposition analysis. Section 6 concludes.

3.3 Data and Variables

3.3.1 Data and descriptive statistics

The data used in this paper is from China's Annual Survey of Industrial Production (1997-2007). The survey is a census conducted by China's National Bureau of Statistics (NBS), and covers all the SOEs and non-SOEs with annual sales over 5 million RMB (roughly 800,000 USD). It is the largest data set after China's general economic census data. The survey extracts detailed information on firm basics including ownership type, years of establishment, legal registration, industry, number of employees, etc. It also collects firm accounting and financial information such as assets, liabilities, sales, profits, etc. It covers three major industries including mining, manufacturing and utilities. To avoid state monopolies, this paper focuses on the sub-sample of manufacturing firms which are classified into 30 two-digit-level categories ranging from "agricultural product and byproduct processing", "food production to craftwork and other manufacturing" and "waste resources and materials recovering" based on "Codes of Industrial Classification for National Economic Activities" (GB/T4754-2002). I use the unique registration ID to match firms over time. To eliminate potential selection bias, I restrict my attention to only active firms.

Before introducing the variables, it is crucial to first clarify the definition of state ownership. There are six types of legal registration for firms in China, i.e., state-owned, collectively owned, privately owned, limited-liability corporations, shareholding firms, and foreign firms. The traditional approach counts only firms as state-owned enterprises when they are legally registered as state-owned. However, this may understate the size of the state sector (Hsieh and Song 2015) as many ultimately state-owned firms are registered as foreign firms as long as a third of their ownership is foreign-held. Instead of using the traditional approach, I follow the classification proposed by Hsieh and Song (2015) and define a firm as a state-owned enterprise when the state share of its registered capital is 50 percent or more when the state is reported as the controlling shareholder. Based on these criteria, there were 51,715 SOEs which accounted for more than a third of manufacturing firms. By 2007, the number of SOEs went down to 15,005 which only constitute a small portion (less than 5 percent) in the manufacturing industry. Table 3.1 summarizes the changes in the composition of the manufacturing industry by ownership type in 1998 and 2007, respectively.

	1998		2007	
	Freq.	Percent	Freq.	Percent
Non-SOE	91,138	63.80	295,086	95.16
SOE	51,715	36.20	15,005	4.84
Total	142,853	100	310,091	100

Table 3.1: Constitution of the Manufacturing Industry by Ownership Type

3.3.2 Variables

In the literature, a common indicator of firm performance is the accounting return on assets (*ROA*). This paper is no exception and is consistent with the literature. Return on assets is calculated as net income over total assets which shows in general how efficient the firm is generating profits using its assets. Table 3.2 summarizes the descriptive statistics of ROA by firm ownership type. Due to concerns with measurement error, I trim the top and bottom 1 percent of the ROA distribution. The top panel of Table 3.2 shows that on average non-SOEs outperform SOEs. The average

ROA for non-SOEs is 0.0986, more than triple of that of the SOEs (0.0306). Similar patterns are manifested at the 5th, 25th, 50th, 75th and 95th quantiles as illustrated in the bottom panel of Table 3.2.

	N	Mean	S.D.	Min	Max
non-SOE	228,487	0.0986	0.2782	-8.6858	18.4831
SOE	12,360	0.0306	0.1706	-2.1887	10.1057
Total	240,847	0.0951	0.2741	-8.6858	18.4831
Quantile	**0.05**	**0.25**	**0.50**	**0.75**	**0.95**
non-SOE	-0.0443	0.0072	0.0368	0.1112	0.4262
SOE	-0.1073	0.0000	0.0149	0.0568	0.1915
Total	-0.0480	0.0066	0.0354	0.1076	0.4147

Table 3.2: Firm Performance (ROA) by Ownership Type (2007)

This is also reflected in Figure 3.4 and Figure 3.5 which plot the pre- and post-reform ROA by quantile for SOEs and non-SOEs, respectively. In both time periods, non-SOEs have higher return to assets than SOEs at any quantile along the performance distributions. For example, the performance differentials are larger below the 20th quantile and above the 40th quantile in the post-reform period. In particular, the ROA of non-SOEs at the 90th quantile of its ROA distribution is about twice the ROA of SOEs at the 90th quantile of the SOE's ROA distribution. Compared to the pre-reform ROA distributions, the gaps for firms in the middle range of the ROA distribution, (i.e., between the 20th quantile and the 60th quantile) were narrowed.

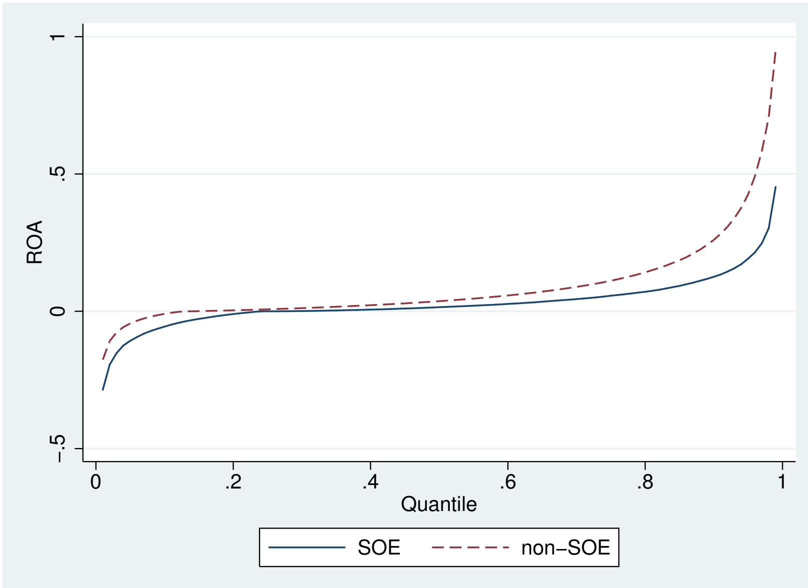

Figure 3.4: ROA by Ownership Type by Quantile, 2006-07

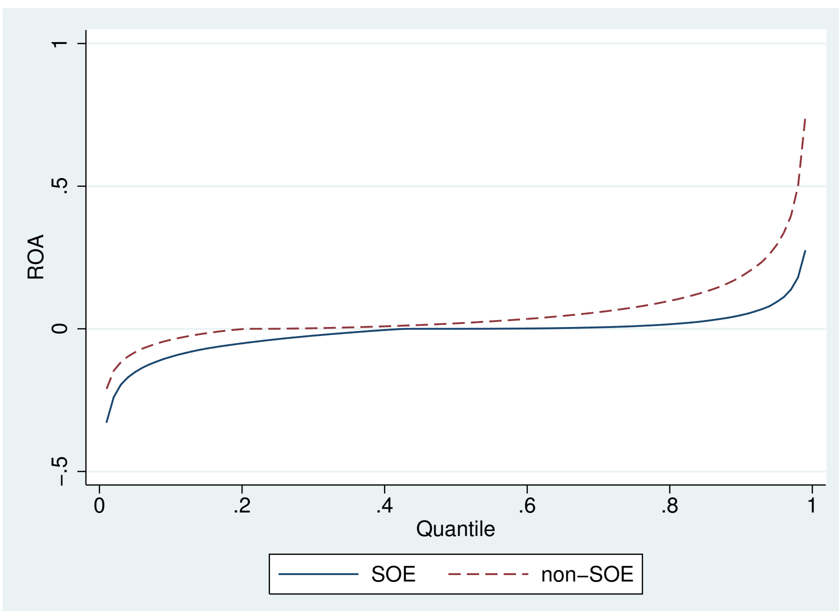

Figure 3.5: ROA by Ownership Type by Quantile, 1998-99

To get a more direct view of the performance differential, I calculate the differences between the ROA distributions, i.e., the gaps between the red curve and the blue curve in Figure 3.4 and Figure 3.5, and plot them along quantiles in Figure 3.6 and

Figure 3.7. The performance differentials follow a J-shaped pattern along the ROA distribution in both pre- and post-reform periods. The red solid line represents the average ROA differential. The two dashed red lines refer to the one standard deviation away from the average ROA differential. The quantile differentials are shown as the black curve. They tend to be larger at both the upper and bottom quantiles but more so at the top of the distribution. In particular, although most of the quantile differentials lie within one standard deviation of the average ROA differential, it increases with an acceleration after the 75th quantile and ultimately falls out of the one standard deviation range when it gets to the 90th quantile.

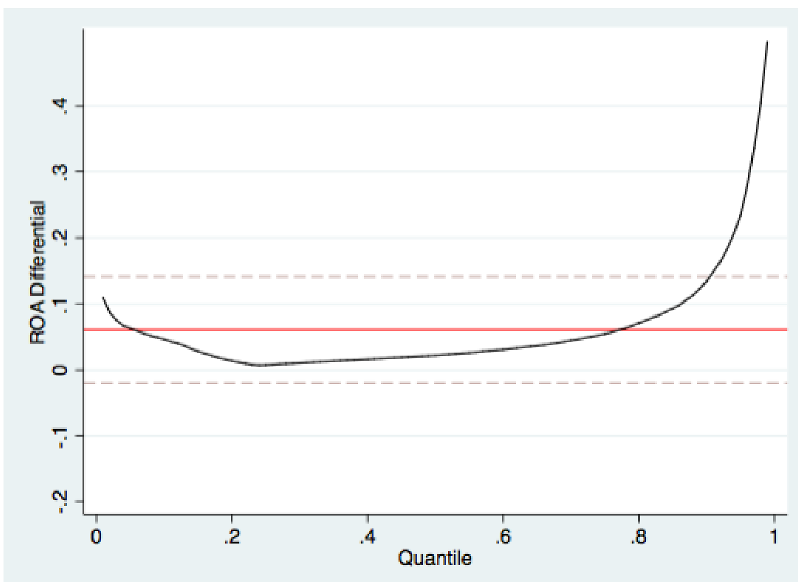

Figure 3.6: ROA Differentials by Quantile, 2006-07

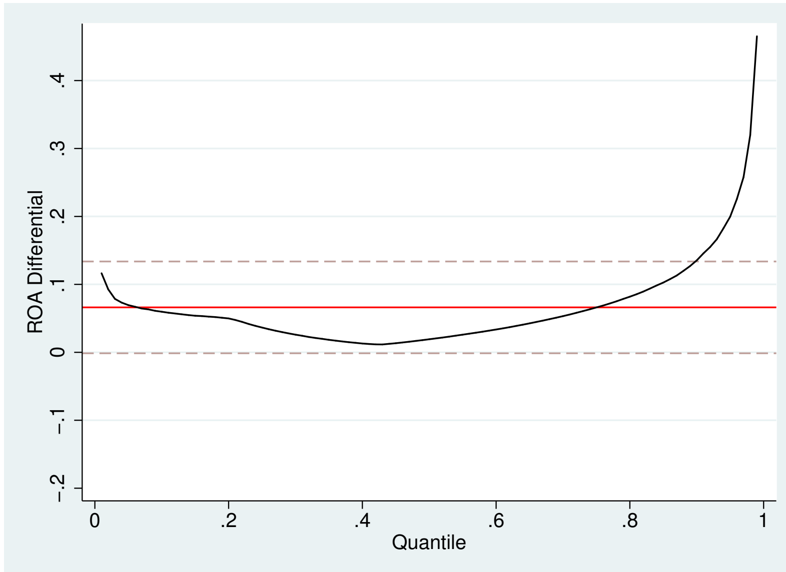

Figure 3.7: ROA Differentials by Quantile, 1998-99

The same J-shaped pattern in the ROA differential is also observed in most of the two-digit level sub-industries coded with “13-30” (without 38) in “Codes of Industrial Classification for National Economic Activities”.

Figure 3.8 and Figure 3.9 plot the kernel density estimates of ROA for both SOEs and non-SOEs both before and after the reform. The graphs clearly show that the distributions for SOEs and non-SOEs are different. This is also confirmed by a two-sample Kolmogorov-Smirnov test which rejects the hypothesis that the ROA for the SOEs and non-SOEs are from the same distribution.

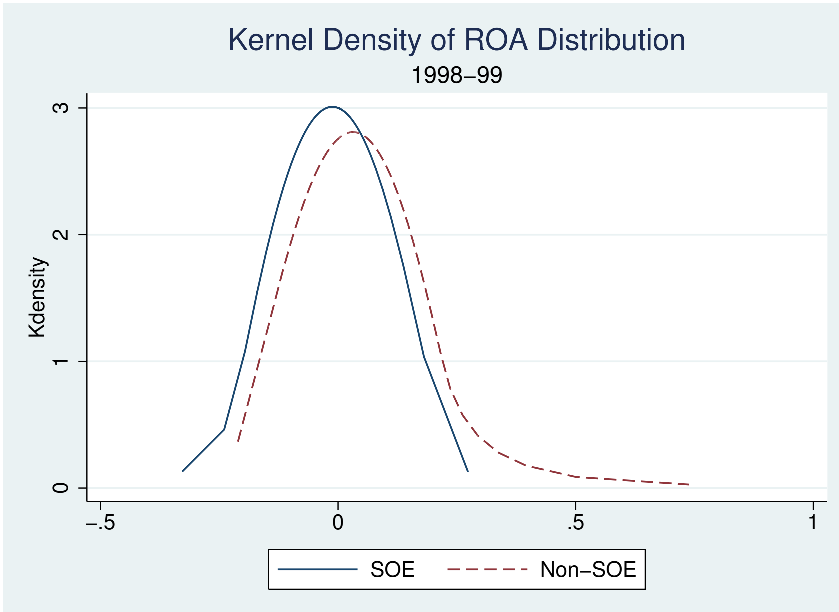

Figure 3.8: Kernel Density Estimates of ROA Distributions by Ownership Type, 1998-99

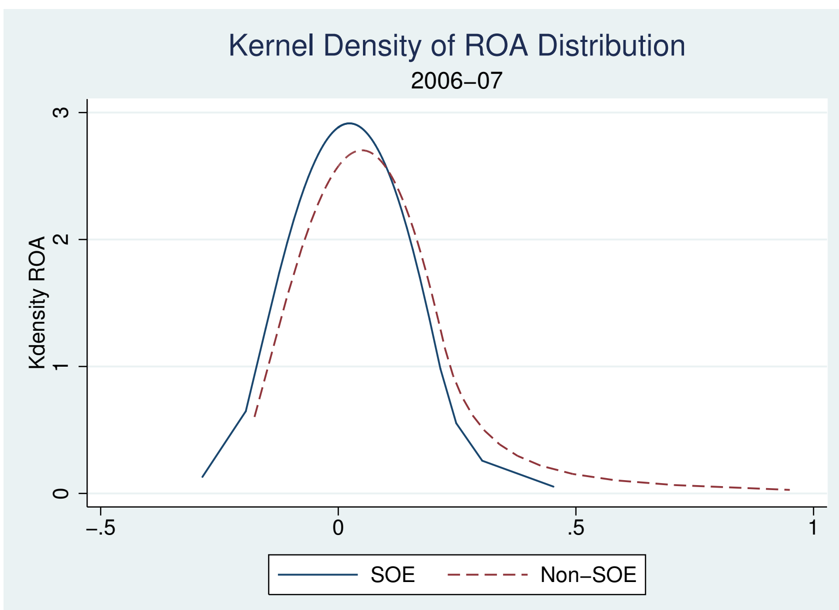

Figure 3.9: Kernel Density Estimates of ROA Distributions by Ownership Type, 2006-07

The main explanatory variable of interest in this paper is ownership type (SOE), a binary variable that equals 1 if the firm is an SOE and 0 otherwise. Besides ownership type, other factors can affect firm performance. Based on Li, Sun and Zou (2009), I include a set of variables to control for industry type, region, firm age, size of firm, leverage, input structure, and earning capability.

Performance of manufacturing firms in China exhibits large disparity across regions. Although the Chinese government has been trying to bridge the East-West gap, firms on the east coast have in general enjoyed greater benefits since the Open Door policy. For such regional differences, I categorize firms into six groups based on their geographical locations associated with regional dummies.

Firm size also matters. Large firms are more likely to gain from economics of scale, but they may become less efficient in overall corporate governance. To control for the impact of firm size on performance, I use the natural logarithm of total sales (*lsales*) and the natural logarithm of total number of employees (*lemployee*s).

Empirical studies done by Dewenter and Malatesta (2001) show that state-owned enterprises tend to use more leverage than private firms do. In the case of China, SOEs usually have access to various preferential treatments from the government[3]; for example, Boyreau-Debray and Wei (2005) document that Chinese banks, which are mostly state-owned, offer easier credit to SOEs. To capture the differences in the financing capability, I employ the long-term debt to total equity ratio (D/E) and the total debt to total assets ratio (D/A) to control for the firm's leverage risk.

In the corporate literature, input structure and earning ability are also considered as important factors that can affect firm performance and hence the capital over sales ratio (K/S) and the operating profit margin, i.e., operating profit over sales ratio

[3]This is usually referred to as "state favoritism" which summarizes the various preferential treatments SOEs get from the government such as access to credit, etc.

(PM) are included in the regression model as controls. Finally, I also control for firm age (Age), i.e., years of establishment in the manufacturing industry.

Just as with other empirical studies using quantile regressions, some variables in this paper are "arguably" endogenous. In addition, sales are incorporated in independent variables (size, K/S and PM). Nonetheless, it is still useful to look at the extent to which the performance differences between SOEs and non-SOEs at different quantiles can be "explained" by the above variables as an accounting exercise, and to provide an evaluation of the performance of SOEs and non-SOEs after the 1999 reform. I also utilize lagged independent variables in the regressions to minimize the impact of endogeneity. Table 3.3 provides the summary statistics of the key variables.

	2006-07		1998-99		Difference in
	Mean	S.D.	Mean	S.D.	Mean
Age	8.9777	8.7473	14.41245	14.38425	-5.43475
lemployees	4.7223	1.0904	4.967486	1.202613	-0.245186
lsales_lag	10.2243	1.2491	9.429373	1.491684	0.794927
D/E_lag	0.7181	136.0556	.861671	54.14565	-0.143571
D/A_lag	0.5639	0.2958	.6682754	.3460661	-0.1043754
K/S_lag	0.3664	4.5743	1.518278	30.67096	-1.151878
PM_lag	0.0296	1.1513	-.1329254	4.489107	0.1625254

Table 3.3: Summary Statistics of Control Variables

Note: The variable "Age" is in years, the number of employees is in heads, and sales are in Yuan.

Table 3.4 checks the pairwise correlation between the variables and find no multicollinearity.

2006-07	SOE	Age	lemployees	lsales_lag	D/E_lag	D/A_lag	K/S_lag	PM_lag
SOE	1.00							
Age	0.30	1.00						
lemployees	0.15	0.20	1.00					
lsales_lag	0.13	0.13	0.65	1.00				
D/E_lag	0.00	0.00	0.00	0.00	1.00			
D/A_lag	0.06	0.09	0.04	0.01	0.01	1.00		
K/S_lag	0.04	0.02	0.01	-0.06	0.00	0.01	1.00	
PM_lag	-0.03	-0.01	-0.00	0.04	0.00	-0.03	-0.40	1.00

1998-99	SOE	Age	lemployees	lsales_lag	D/E_lag	D/A_lag	K/S_lag	PM_lag
SOE	1.00							
Age	0.42	1.00						
lemployees	0.07	0.28	1.00					
lsales_lag	-0.25	-0.04	0.62	1.00				
D/E_lag	0.01	-0.00	0.00	-0.01	1.00			
D/A_lag	0.18	0.16	0.06	-0.11	0.01	1.00		
K/S_lag	0.04	0.03	-0.01	-0.11	0.00	0.02	1.00	
PM_lag	-0.04	-0.03	0.00	0.10	-0.00	-0.06	-0.72	1.00

Table 3.4: Pairwise Correlation between Key Variables
Note: The variable “Age” is in years, the number of employees is in heads, and sales are in Yuan.

3.4 RIF-Regressions

In this section, I first introduce the quantile regression technique that I use for the decomposition analysis. Quantile regression is a tool to model the relationship between the explanatory variables and the distribution of the outcome variable. In particular, it is very useful when the extremes are important. In the context of this paper, the quantile regression model is appropriate for two reasons: 1) the 1999 reform has a special focus on firms that are either very big or very small, and 2) the large ROA differentials mainly appear at the upper and bottom quantiles. In the literature, there are two approaches to model quantile regressions: the conditional quantile regressions developed by Koenker and Bassett (1978) and Koenker (2005), and the more recent unconditional quantile regression technique proposed by Firpo, Fortin and Lemieux

(2009)[4]. The key difference between conditional and unconditional quantile regressions is that the parameters from a conditional quantile regression indicate the effect of a covariate on a conditional quantile, but the effect on the unconditional quantile (i.e., the unconditional quantile partial effect) may be of more policy or economic interest in many cases. However, unlike the mean regressions, the conditional quantile coefficient does not directly imply the unconditional quantile partial effect because the law of iterated expectations does not extend to the quantiles. I therefore use the method proposed by FFL (2009), which provides a direct estimate of the unconditional quantile partial effect, to compute marginal effects of changes in ownership type on quantiles of the unconditional distribution of return on assets. The key to the unconditional quantile regression approach is the recentered influence function (RIF). The influence function is a directional derivative of the mixing distribution $F_{Y,\varepsilon \cdot G_Y}$, i.e.,

$$IF\left(y; v, F_Y\right) = \frac{\partial v\left(F_{Y,\varepsilon \cdot G_Y}\right)}{\partial \varepsilon} \mid_{\varepsilon=0},$$

where $v\left(\cdot\right)$ is the real-value functional such that $v: \ \mathcal{F}_v \rightarrow \mathbb{R}$, and $\mathcal{F}_v$ is a class of distribution functions such that $F_Y, \ G_Y \in \mathcal{F}_v$, and the mixing distribution $v\left(F_{Y,\varepsilon \cdot G_Y}\right)$ is defined as

$$F_{Y,\varepsilon \cdot G_Y} \equiv \left(1-\varepsilon\right) F_Y + \varepsilon \cdot G_Y \ 0 \leq \varepsilon \leq 1,$$

which is ε away from F_Y in the direction of the probability distribution G_Y.

The recentered influence function is

$$\begin{aligned} RIF\left(y; v, F_Y\right) &= v\left(F_Y\right) + \int IF\left(s; v, F_Y\right) \cdot d\Delta_y \\ &= v\left(F_Y\right) + IF\left(y; v, F_Y\right), \end{aligned}$$

[4]FFL (2009) from hereon.

which is defined as the first order von Mises linear approximation (VOM) of the real-value functional $v(F_Y)$ where $G_Y = \Delta_Y$ and $\varepsilon = 1$ with Δ_Y denoting the probability measure that puts mass 1 at the value y. The last equality holds because $\int IF(s; v, F_Y) \cdot dF_Y(s) = 0$ by definition.

One important feature of the RIF is that in the presence of covariates X the following equalities hold, i.e.,

$$\begin{aligned} v(F_Y) &= \int RIF(y; v, F_Y) \cdot dF_Y(y) \\ &= \int \int RIF(y; v, F_Y) \cdot dF_{Y|X}(y|X = x)\, dF_X(x) \\ &= \int E[RIF(Y; v) \mid X = x] \cdot dF_X(x). \end{aligned}$$

The first equality follows because the influence function integrates to zero over the distribution of F_Y by definition; the second equality follows by the definition of the unconditional distribution function of Y; and the third equality holds because of law of iterated expectations. In FFL (2009), the authors prove that the unconditional average partial effect (UAPE) can be derived from the vector of average derivatives, i.e.,

$$UAPE = \int \frac{E[RIF(Y; v) \mid X]}{dx} \cdot dF_X(x).$$

In the case of quantiles, let q_θ be the θth quantile of a random variable y, i.e.,

$$q_\theta = v_\theta(F_Y) = \inf_q \{q : F_Y(q) \geq \theta\}.$$

The influence function for the q_θ is

$$IF(Y; q_\theta) = \frac{\theta - I(Y \leq q_\theta)}{f_Y(q_\theta)} \tag{3.1}$$

and the recentered influence function for the θth quantile is obtained by adding back q_θ to equation (3.1), i.e.,

$$RIF(Y; q_\theta) = q_\theta + \frac{\theta - I(Y \leq q_\theta)}{f_Y(q_\theta)}. \tag{3.2}$$

Assume a linear model $E\left[RIF(Y;q_\theta)\right] = X'\beta_\theta$, where Y is the return on assets and X includes all the covariates including ownership type. The main advantage of using the recentered influence function is that the coefficient β_θ from the above linear model is equal to the average (over the distribution of X) partial derivatives of the linear model represented by $E\left[\frac{dE[RIF(Y;q_\theta)|X]}{dx}\right]$ for any quantile. FFL (2009) defines this average marginal effect $E\left[\frac{dE[RIF(Y;q_\theta)|X]}{dx}\right]$ as the unconditional quantile partial effect (UQPE) which can then easily be computed from RIF regressions using the recentered influence function as the dependent variable.

In doing this, I first estimate the kernel density of q_θ using the Epanechnikov kernel function with a Sheather-Jones bandwidth of 0.1. Following equation ($[eq:IF]$), I compute the influence function for each observation using the sample estimate of q_θ, add back q_θ and get the recentered influence function. Under the linearity assumption, the general model is specified as

$$RIF\left(ROA;q_\theta\right) = \beta^0_\theta + \beta^1_\theta SOE + X^T\gamma_\theta + \varepsilon.$$

The coefficients (β's and γ's) from the above RIF regressions indicate the marginal effects of covariates on different ROA quantiles. The coefficient estimates ($\hat{\beta}$'s and $\hat{\gamma}$'s) are interpreted as the estimated return to the regressors at the θth quantile of the ROA distribution.

Due to the concern of endogeneity, I carry out a series of quantile regressions to examine the effects of ownership type on the ROA differential at different quantiles. I start with a regression on ownership type without any other control variables, which essentially captures the raw SOE disadvantage. The RIF regression coefficients for the 10th, 50th, and 90th quantiles along with their standard errors are reported in Table 3.5. On the whole, the SOE effects are negative and non-monotonic across the entire distribution of return on assets. The SOE disadvantage tends to shrink at the lower quantiles and expand in the upper quantiles. For example, at the 10th quantile, the SOE disadvantage is of 0.063 for 1998-99 and of 0.038 for 2006-07. At the 90th quantile, the SOE disadvantage rises from 0.125 to 0.146 after the reform. I also present the corresponding OLS coefficients in the last column of each panel for comparison.

Quantile	10th	50th	90th	OLS
			1999	
SOE	-0.063	-0.025	-0.125	-0.073
	(0.001)**	(0.000)**	(0.002)**	(0.002)**
_cons	-0.042	0.016	0.184	0.059
	(0.001)**	(0.000)**	(0.001)**	(0.001)**
Industry	No	No	No	No
Region	No	No	No	No
N	135,113	135,113	135,113	135,113
			2007	
SOE	-0.038	-0.027	-0.146	-0.068
	(0.001)**	(0.001)**	(0.003)**	(0.003)**
_cons	-0.010	0.037	0.260	0.099
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
Industry	No	No	No	No
Region	No	No	No	No
N	240,845	240,845	240,845	240,845

* p<0.05; ** p<0.01

Table 3.5: Unconditional Quantile Regression Coefficients (No Controls)

Figure 3.10 gives a visual summary of the the RIF regression results for firms in 2007. The curve depicts the coefficients corresponding to each of the 91 different ROA quantiles equally spread over the 5th to 95th quantile of the ROA distribution. The green line represents the estimated coefficients for firms with an ROA between the 5th and 95th quantiles. The shaded grey area depicts the 95 percent confidence interval around the quantile regression coefficients. The solid red line represents the OLS estimate which does not vary across quantiles. The two dashed red lines reflect one standard deviation change around the OLS estimate of the mean effect. The SOE effect varies over quantiles of the ROA distribution and most of the quantile regression estimates lie outside the 95 percent confidence interval of the OLS estimate, which indicates that it is inadequate just looking at the average performance differential.

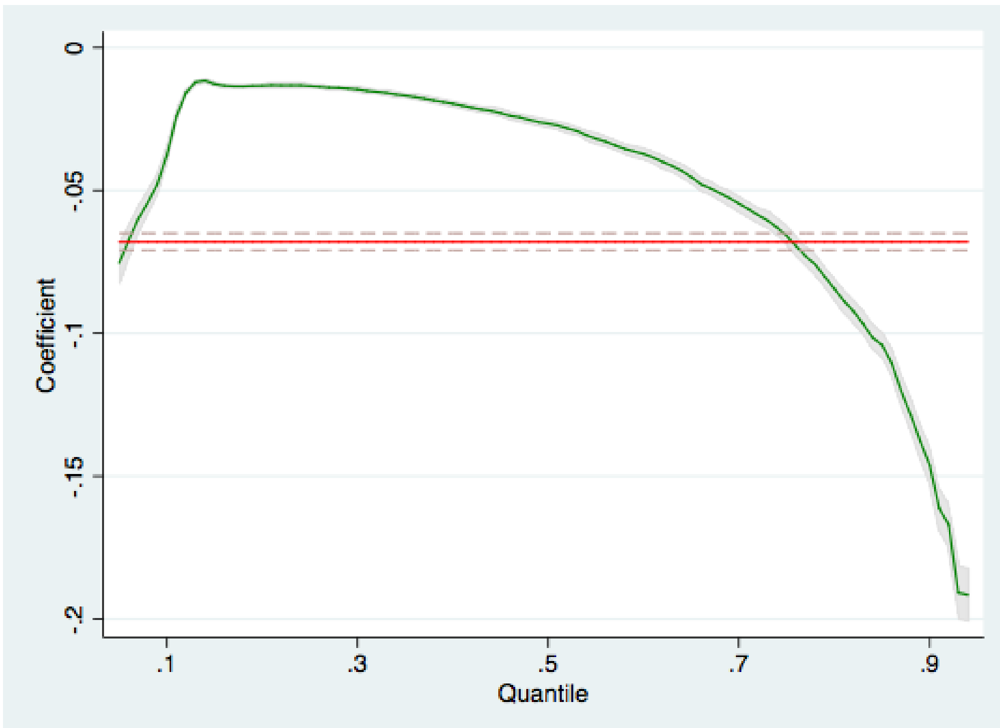

Figure 3.10: Quantile estimates vs OLS estimates

I now begin to add other control variables starting with firm age, number of employees, industry and region dummies. These variables are commonly used in the literature but are arguably endogenous. Nonetheless it is useful to know the extent to which the ROA differential at different quantiles are "explained" by these additional regressors. The estimated coefficients are reported in Table 3.6. Controlling for these variables decreases the SOE effects for all quantiles in both time periods. For example, at the 90th quantile, the raw ROA disadvantage is 0.146 whereas after adding the controls, the SOE effect decreases to 0.078. The magnitude of the OLS coefficients also declines.

Quantile	10^{th}	50^{th}	90^{th}	OLS
		1999		
SOE	-0.051	-0.021	-0.113	-0.067
	(0.001)**	(0.000)**	(0.002)**	(0.002)**
Age	-0.001	-0.000	-0.001	-0.001
	(0.000)**	(0.000)**	(0.000)**	(0.000)**
lemployees	0.003	-0.001	-0.028	-0.009
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
_cons	-0.068	0.027	0.432	0.155
	(0.003)**	(0.001)**	(0.006)**	(0.004)**
Industry	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	134,735	134,735	134,735	134,735
		2007		
SOE	-0.028	-0.018	-0.078	-0.043
	(0.001)**	(0.001)**	(0.004)**	(0.003)**
Age	-0.000	-0.001	-0.004	-0.002
	(0.000)**	(0.000)**	(0.000)**	(0.000)**
lemployees	0.001	0.005	-0.011	-0.004
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
_cons	-0.017	0.030	0.481	0.190
	(0.001)**	(0.001)**	(0.008)**	(0.004)**
Industry	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	240,819	240,819	240,819	240,819

* $p<0.05$; ** $p<0.01$

Table 3.6: Unconditional Quantile Regression Coefficients (With Controls)

Table 3.7 presents the estimates when all the other variables are included to control for sales, financing capability, input structure, and earning management. These variables can affect firm performance and are frequently used in the literature but induce an even stronger argument for endogeneity. After adding all these control variables, the SOE disadvantage is reduced but the inverse U-shape pattern remains (i.e. smaller in the middle and larger at the upper and lower quantiles). It is worth noting that controlling for all these factors has the greatest effect at the top of the distribution. For example, at the 90th quantile, with controls the average ROA for SOEs in 2007 is less than that of non-SOEs by 0.072 (the 8th column in Table 3.7) whereas the different is 0.146 without (the 8th column in Table3.5). At the 10th quantile, the same coefficient only falls from 0.038 (the 6th column in Table 3.7) to

0.025 (the 6th column in Table 3.5).

Quantile	10th	50th	90th	OLS
			1999	
SOE	-0.027	-0.012	-0.073	-0.041
	(0.001)**	(0.000)**	(0.002)**	(0.002)**
Age	-0.000	-0.000	-0.001	-0.000
	(0.000)	(0.000)**	(0.000)**	(0.000)**
lemployees	-0.010	-0.005	-0.057	-0.028
	(0.001)**	(0.000)**	(0.001)**	(0.001)**
lsales_lag	0.017	0.006	0.036	0.023
	(0.001)**	(0.000)**	(0.001)**	(0.001)**
D/E_lag	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
D/A_lag	-0.090	-0.017	-0.133	-0.085
	(0.004)**	(0.001)**	(0.006)**	(0.002)**
K/S_lag	0.000	0.000	0.000	0.000
	(0.000)**	(0.000)	(0.000)	(0.000)*
PM_lag	0.002	-0.000	-0.001	0.000
	(0.001)**	(0.000)	(0.000)*	(0.000)
_cons	-0.128	-0.002	0.286	0.061
	(0.006)**	(0.001)*	(0.009)**	(0.007)**
Industry	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	109,130	109,130	109,130	109,130
			2007	
SOE	-0.025	-0.016	-0.072	-0.039
	(0.001)**	(0.001)**	(0.004)**	(0.003)**
Age	-0.000	-0.001	-0.003	-0.001
	(0.000)**	(0.000)**	(0.000)**	(0.000)**
lemployees	-0.008	-0.006	-0.071	-0.032
	(0.000)**	(0.000)**	(0.002)**	(0.001)**
lsales_lag	0.011	0.013	0.073	0.035
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
D/E_lag	0.012	0.013	0.078	0.037
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
D/A_lag	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)**
K/S_lag	-0.048	-0.059	-0.260	-0.134
	(0.001)**	(0.001)**	(0.006)**	(0.002)**
PM_lag	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
_cons	-0.076	-0.038	0.056	-0.006
	(0.002)**	(0.002)**	(0.013)**	(0.006)
Industry	Yes	Yes	Yes	Yes
Region	Yes	Yes	Yes	Yes
N	240,276	240,276	240,276	240,276

* $p<0.05$; ** $p<0.01$

Table 3.7: Unconditional Quantile Regression Coefficients (All Controls)

Similar to Figure 3.10, Figure 3.11 plots the coefficients on the binary variable *SOE* from the RIF regression and its corresponding OLS coefficient. After adding all the control variables, most of the quantile regression estimates still lie outside the 95 percent confidence interval of the OLS estimate.

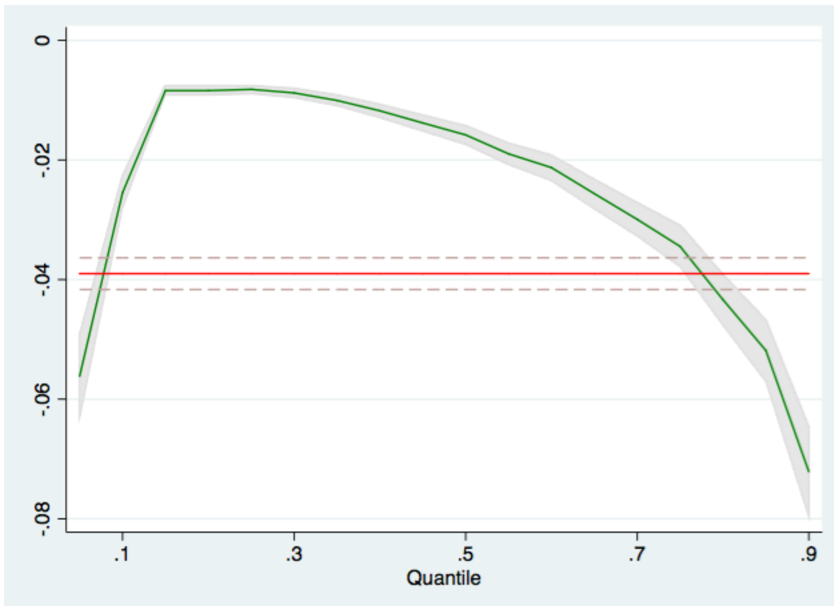

Figure 3.11: Quantile Estimates vs OLS Estimates (2007)

Figure 3.12 shows the combined patterns of estimated coefficients from RIF regressions for 1998-99 and 2006-07, respectively. The solid red line with triangle markers represent the regression coefficients using 2006-07 data and the dashed blue line with cross markers represents the coefficients using the 1998-99 data. The patterns look similar in both periods: first rising and then declining, and the magnitude of the SOE effect is negative across all quantiles. The SOE effect tends to be larger for either very bad performers or very good performers but more so for firms at the higher end of the distribution. The main difference occurs at the bottom quantiles. The magnitude of the SOE effect in the post-reform period is larger for firms in the lowest 10th quantile. For firms located between the 10th and the 30th quantiles, the post-reform SOE effect turns to be smaller relative to the pre-reform period; for example, at the 20th quantile, the post-reform effect of ownership type is only a third of that in the pre-reform period. The SOE effect becomes stronger once it get beyond the 30th quantile. More importantly, the results also indicate that the ROA differential between SOEs and non-SOEs is mainly driven by the gap among firms in the upper quantiles of the ROA distribution for both pre- and post-reform periods.

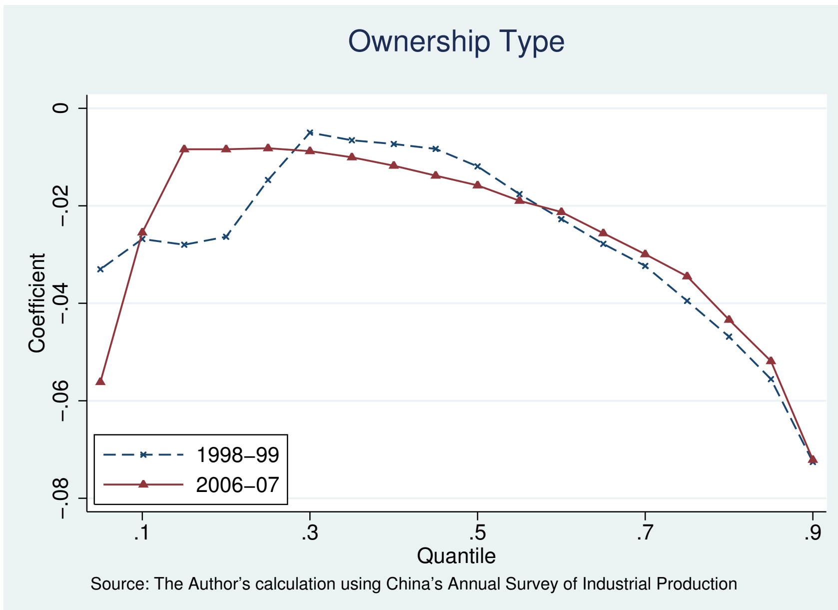

Figure 3.12: RIF Regression Coefficients on Ownership Type: 1998-99 and 2006-07

3.5 Decomposition

The decomposition exercise contains both the aggregate and detailed decompositions. The aggregate decomposition examines what portion of the ROA differential is attributable to differences in firm characteristics (i.e. the "composition effect") and which is due to ownership type differences in returns to these characteristics (i.e. the "SOE structure effect"). The detailed decomposition identifies which characteristics and the return to these characteristics are relatively more important in "explaining" the ROA differential. The decomposition analysis is based on the RIF regressions introduced in Section 3.4. The advantage of RIF-regression-based decomposition (as opposed to other quantile decomposition methods) is that under a linearity assumption the RIF regression coefficients can be used to perform the detailed decomposition in an analogous way to the standard Oaxaca-Blinder methodology.

Denote q_g^θ the θth quantile of ROA, where $g \in (SOE,\ non\text{-}SOE)$. Let Δ_O^θ be the overall ROA differential between non-SOE and SOE at the θth quantile. By definition,

$$\Delta_O^\theta \equiv q_{SOE}^\theta - q_{non\text{-}SOE}^\theta.$$

Under the assumptions of ignorability and common support, Δ_O^θ can be decomposed in two parts, i.e.

$$\Delta_O^\theta = \left[q_{SOE}^\theta - q_C^\theta\right] - \left[q_{non\text{-}SOE}^\theta - q_C^\theta\right] = \Delta_S^\theta + \Delta_X^\theta,$$

where q_C^θ is a quantile statistic of a counterfactual distribution which would have prevailed with the non-SOE ownership type but with observed and unobserved firm characteristics as of an SOE. The first term captures the SOE structure effect and the second term represents the composition effect.

One nice feature of the recentered influence function shown in equation (3.2) is that the following equality always hold for any quantile, i.e.,

$$q_g^\theta = E_X\left[E\left[RIF\left(y_g; q_g^\theta\right)|X\right]\right] \text{ and } g \in (SOE,\ non\text{-}SOE).$$

Then $\left\{\Delta_O^\theta,\ \Delta_S^\theta,\ \Delta_X^\theta\right\}$ can be written as

$$\Delta_O^\theta = E_X\left[E\left[RIF\left(y_{SOE}; q_{SOE}^\theta\right)|X\right]\right] - E_X\left[E\left[RIF\left(y_{non\text{-}SOE}; q_{non-SOE}^\theta\right)|X\right]\right], \tag{3.3}$$

$$\Delta_S^\theta = E_X\left[E\left[RIF\left(y_{SOE}; q_{SOE}^\theta\right)|X\right]\right] - E_X\left[E\left[RIF\left(y_{non-SOE}; q_C^\theta\right)|X\right]\right], \tag{3.4}$$

$$\Delta_X^\theta = E_X\left[E\left[RIF\left(y_{non\text{-}SOE}; q_C^\theta\right)|X\right]\right] - E_X\left[E\left[RIF\left(q_{non-SOE}^\theta; q_{SOE}^\theta\right)|X\right]\right]. \quad (3.5)$$

Following the literature, assume $E\left[RIF\left(q_g^\theta;\theta\right)|X\right]$ takes a linear functional form, i.e.,

$$E\left[RIF\left(y;q_\theta\right)|X\right] = X' \cdot \gamma_\theta.$$

With the linear specification, the decomposition based on RIF regressions looks very much like standard Oaxaca-Blinder decomposition. It is shown in FFL (2006) that we can rewrite equation (3.4) and equation (3.5) as

$$\Delta_S^\theta = E_X\left[X|g = non\text{-}SOE\right]' \cdot \left(\gamma_{non\text{-}SOE,\theta} - \gamma_{C,\theta}\right)$$

and

$$\Delta_X^\theta = E_X\left[X|g = non\text{-}SOE\right]' \cdot \gamma_{C,\theta} - E_X\left[X|g = SOE\right]' \cdot \gamma_{SOE,\theta}.$$

The results of the decomposition are presented in Figures 3.13-3.16 and Tables 3.8-3.9. The base group used in the RIF regression is the non-SOEs. The covariates used in the RIF regressions contain the full set of control variables. Figure 3.13 shows the overall ROA differentials (i.e. Δ_O^θ) at 19 points equally distributed between the 5th and 95th quantiles, and the decomposed composition effect (i.e. Δ_X^θ) and SOE structure effect (i.e. Δ_S^θ). At the aggregate level, the overall ROA differential has a J-shaped pattern both pre- and post-reform, i.e., the ROA differential is larger at both the bottom and upper end of the distribution. For firms in the middle range of the ROA distribution, the performance differential is mainly due to different firm characteristics. However, differences in the return to characteristics (i.e. the SOE structure effect) are more important to "explain" the widening ROA differentials below the 10th quantile and beyond the 80th quantile. This is consistent with the canonical corporate theory that state ownership is usually less efficient than private ownership. Due to inefficiencies in corporate governance, the SOE ownership tends to be a bottleneck that blocks firms in the upper quantiles from becoming top-tier competitors in the manufacturing industry. For firms at the bottom of the ROA distribution, state ownership keeps these firms in a low performance trap. The comparison of Figure 3.13 and Figure 3.14 reflects how the composition and SOE structure effects change in the pre- and post- reform periods. In 1998-99, the ROA differential at the upper quantiles is primarily attributable to the SOE structure effect. It is interesting to note that this effect becomes less important in 2006-07, indicating that although SOE ownership remains a bottleneck for good performing firms, it is less of a hurdle after the reform. This may be possibly due to the improvement of corporate governance or the exit of inefficient SOEs.

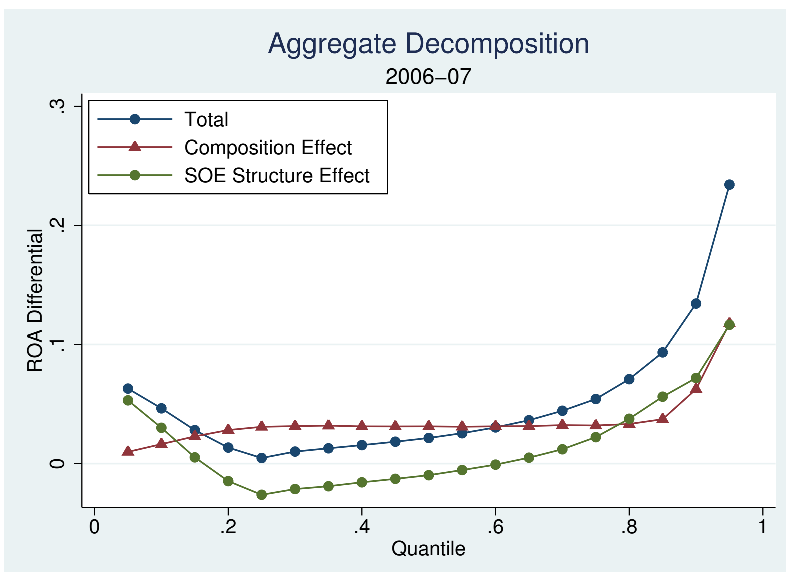

Figure 3.13: Decomposition of Total ROA Differential into Composition and SOE Structure Effects

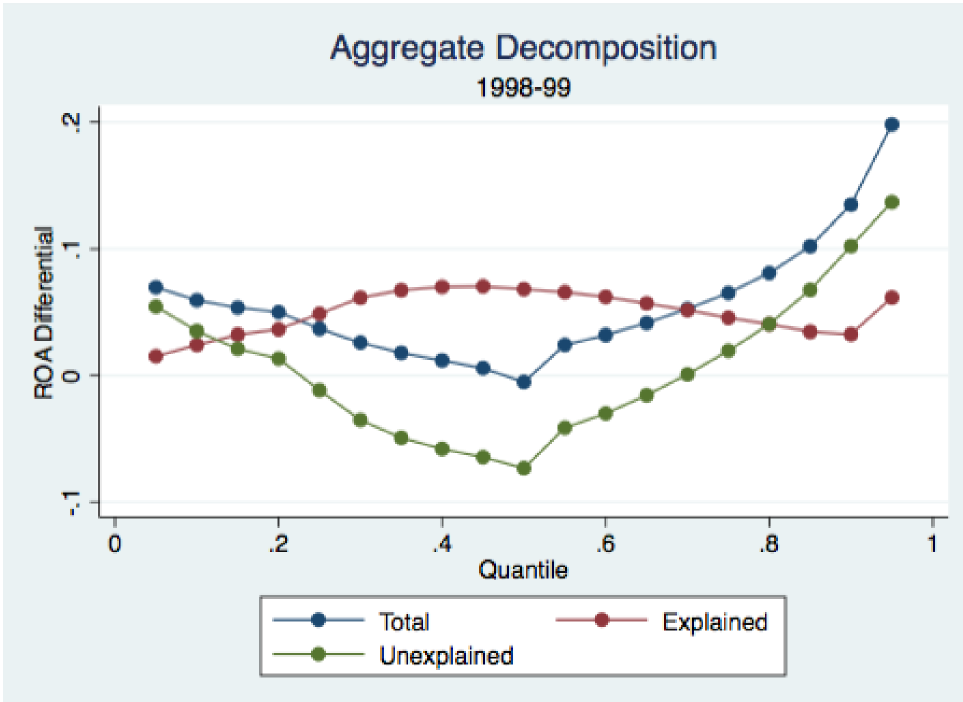

Figure 3.14: Decomposition of Total ROA Differential into Composition and SOE Structure Effects

Figure 3.15 and Figure 3.16 illustrate the detailed decomposition which looks at the contribution of six main sets of factors. In 1998-99, size appears to be the most important firm characteristic to the composition effect but it only "explains" a very small portion of the upper quantile ROA differential in the post-reform period. Despite its decreasing contribution to the composition effect, size is still recognized as the most decisive factor to the post-reform SOE structure effect.

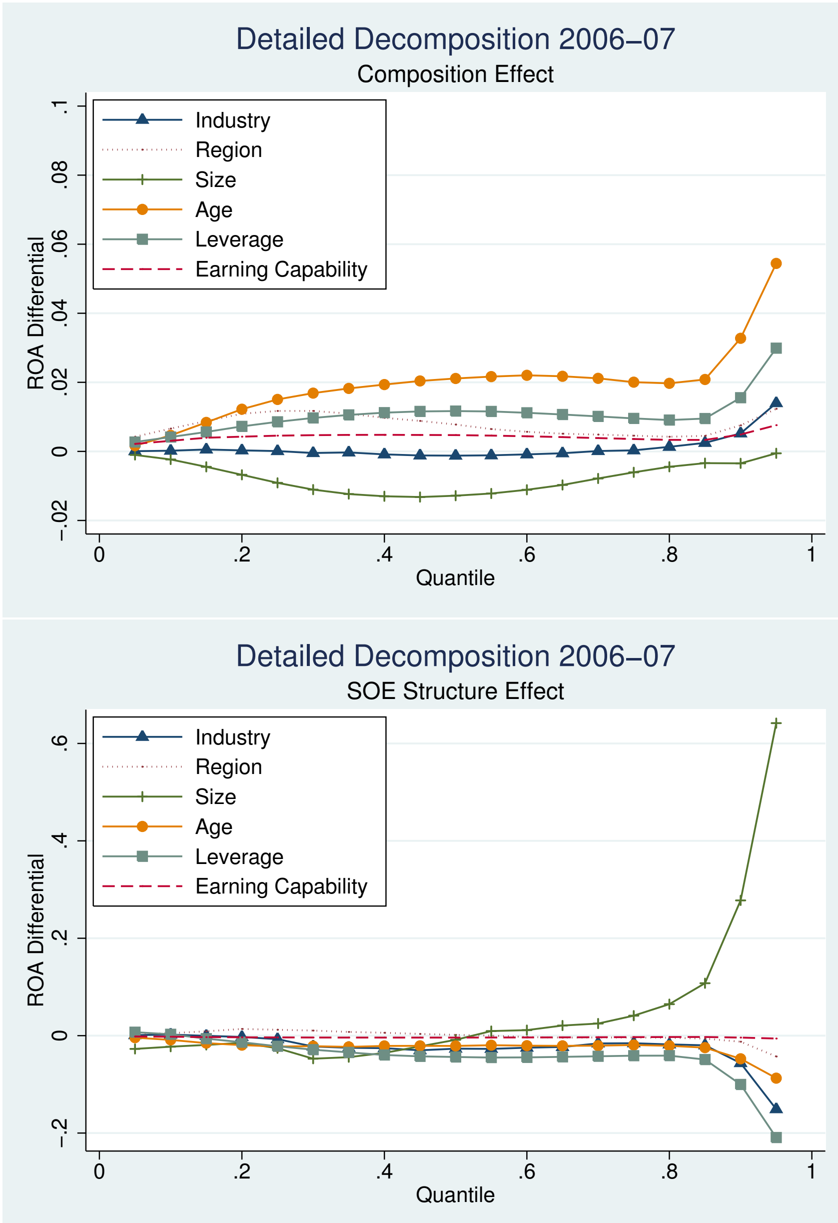

Figure 3.15: Detailed Decomposition 2006-07

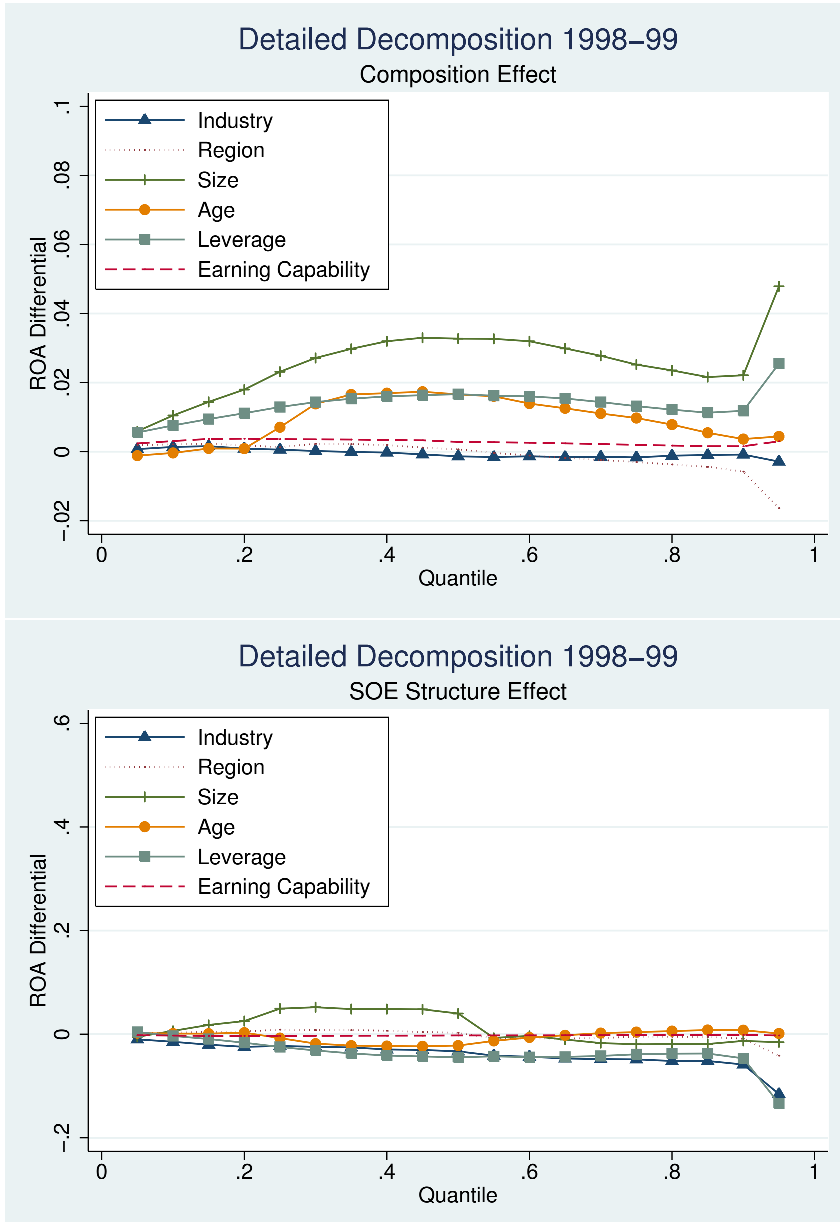

Figure 3.16: Detailed Decomposition 1998-99

Table 3.8 and Table 3.9 summarize results from the detailed decomposition for 1998-99 and 2006-07 at the 10th, 50th and 90th quantiles. The SOE structure effect accounts for about 76 percent[5] of the estimated ROA differential at the 90th quantile

[5]The number is obtained by 0.1022007/0.134743=0.76

in the pre-reform period but declines to 53 percent after the reform. One explanation for the reduction in the SOE structure effect is the post-reform improved internal management mechanisms and the enhanced productivity suggested by Yusuf, Nabeshima, and Perkins (2005), Xu et al. (2005) and Lin and Zhu (2001). Firms below the 10th quantile, however, still endure a rising effect of SOE ownership from 59 percent in 1998-99 to 65 percent in 2006-07.

	(1)		(2)		(3)	
Quantiles	10th		50th		90th	
Unadjusted ROA differential	0.047		0.022		0.134	
Estimated difference	0.046	(.001)	0.02	(.002)	0.134	(.001)
Composition effects attributable to						
Age	0.005	(.000)	0.021	(.001)	0.033	(.001)
Size (employment, sales)	-0.002	(.000)	-0.013	(.000)	-0.006	(.001)
Leverage	0.004	(.000)	0.012	(.000)	0.016	(.001)
Input structure, earning ability	0.003	(.001)	0.005	(.001)	0.005	(.001)
Industrial sectors	0.000	(.000)	-0.001	(.000)	0.005	(.001)
Region	0.007	(.000)	0.008	(.000)	0.008	(.001)
Total explained by model	0.016	(.001)	0.031	(.002)	0.062	(.003)
Structure effects attributable to						
Age	-0.009	(.001)	-0.021	(.002)	-0.047	(.003)
Size (employment, sales)	-0.023	(.007)	-0.009	(.010)	0.278	(.011)
Leverage	0.003	(.002)	-0.044	(.003)	-0.100	(.003)
Input structure, earning ability	-0.002	(.001)	-0.004	(.001)	-0.004	(.001)
Industrial sectors	0.002	(.004)	-0.027	(.006)	-0.056	(.006)
Region	0.005	(.002)	0.002	(.002)	-0.012	(.003)
Constant	0.053	(.009)	0.093	(.013)	0.014	(.013)
Unexplained ROA differential	0.030	(.001)	-0.010	(.002)	0.072	(.003)

Table 3.8: ROA Differential by Ownership Type: RIF Decomposition Results (2006-07)

Quantiles	(1) 10th		(2) 50th		(3) 90th	
Unadjusted ROA differential	0.060		0.019		0.135	
Estimated difference	0.059	(.001)	-0.005	(.001)	0.135	(.001)
Composition effects attributable to						
Age	-0.000	(.001)	0.017	(.001)	0.004	(.001)
Size (employment, sales)	0.010	(.000)	0.033	(.001)	-0.003	(.001)
Leverage	0.008	(.000)	0.017	(.000)	0.012	(.000)
Input structure, earning ability	0.003	(.000)	0.003	(.000)	0.002	(.000)
Industrial sectors	0.001	(.000)	-0.001	(.001)	-0.001	(.001)
Region	0.002	(.000)	0.001	(.000)	-0.006	(.000)
Total explained by model	0.024	(.001)	0.068	(.001)	0.062	(.003)
Structure effects attributable to						
Age	0.001	(.001)	-0.022	(.002)	0.008	(.002)
Size (employment, sales)	0.006	(.005)	0.040	(.007)	-0.013	(.006)
Leverage	-0.003	(.002)	-0.045	(.002)	-0.047	(.002)
Input structure, earning ability	-0.003	(.000)	-0.002	(.000)	-0.001	(.000)
Industrial sectors	-0.015	(.002)	-0.033	(.003)	-0.059	(.003)
Region	0.003	(.001)	0.003	(.002)	-0.009	(.002)
Constant	0.045	(.006)	-0.013	(.009)	0.223	(.008)
Unexplained ROA differential	0.035	(.001)	-0.073	(.002)	0.102	(.002)

Table 3.9: ROA Differential by Ownership Type: RIF Decomposition Results (1998-99)

3.6 Conclusion

This paper examines the effect of SOE ownership type on changes in the pre- and post-reform distribution of return on assets. I adopt the RIF quantile regression methodology proposed by Firpo, Fortin and Lemieux (2009). The state ownership has a negative impact on the ROA measure across the entire distribution in both 1998-99 and 2006-07. The effect is especially larger at the upper and bottom quantiles. Based on the RIF-regression results, I carry out a decomposition analysis with a special focus on the higher and lower ends of the ROA distribution. The overall ROA differential is then divided into a composition effect which reflects the extent to which the ROA differential is "explained" by the differences in firm characteristics and a SOE structure effect which captures the portion of the ROA differential contributed by the differences in the return to the characteristics. At the aggregate level, the performance differential is mainly attributed to different firm characteristics for firms in the middle range of the ROA distribution. However, for firms with either very good or very bad performance, it appears that differences in the return to characteristics (i.e. the SOE structure effect) are more important to "explain" the accelerating ROA differentials below the 10th quantile and beyond the 80th quantile. These findings imply that no matter before or after the reform 1) the SOE ownership keeps firms in the upper quantiles from attaining top tier in the manufacturing industry, and 2) traps firms at the bottom quantiles from moving up. The comparison of decomposition results for pre- and post-reform periods shows that although the SOE ownership is of a hurdle in both time periods, it is worth noting that the SOE structure effect becomes less important in "explaining" ROA differential at the upper quantiles in 2006-07.

Bibliography

[1] Acemoglu, D. (2001). Good jobs versus bad jobs. *Journal of labor Economics 19*(1), 1–21.

[2] Albrecht, J., L. Navarro, and S. Vroman (2009). The effects of labour market policies in an economy with an informal sector. *The Economic Journal 119*(539), 1105–1129.

[3] Cao, K. H. and J. A. Birchenall (2013). Agricultural productivity, structural change, and economic growth in post-reform China. *Journal of Development Economics 104*, 165–180.

[4] Chan, K. W. and Y. Hu (2003). Urbanization in China in the 1990s: New definition, different series, and revised trends. *China Review*, 49–71.

[5] David, H., L. F. Katz, and A. B. Krueger (1997). Computing inequality: have computers changed the labor market? Technical report, National Bureau of Economic Research.

[6] Davidoff, T. and J. Yoshida (2008). Reconciling micro and macro estimates of substitution between housing and non-housing consumption by relaxing homotheticity restriction.

[7] Davis, M. A. and J. Heathcote (2005). Housing and the business cycle. *International Economic Review 46*(3), 751–784.

[8] Démurger, S., M. Gurgand, S. Li, and X. Yue (2009). Migrants as second-class workers in urban China? a decomposition analysis. *Journal of Comparative Economics 37*(4), 610–628.

[9] Dollar, D. and B. F. Jones (2013). China: An institutional view of an unusual macroeconomy. Technical report, National Bureau of Economic Research.

[10] Fang, C. and W. Dewen (2008). Impacts of internal migration on economic growth and urban development in China. pp. 245.

[11] Fang, C. and Z. Zhao (2009). Employment and inequality outcomes in China. *Institute of Population and Labour Economics, Chinese Academy of Social Sciences*.

[12] Flavin, M. and S. Nakagawa (2008). A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence. *The American economic review*, 474–495.

[13] Gete, P. (2010). Housing markets and current account dynamics. *Available at SSRN 1558512*.

[14] Golley, J. and X. Meng (2011). Has China run out of surplus labour? *China Economic Review 22*(4), 555–572.

[15] Guerrieri, V. (2007). Heterogeneity, job creation and unemployment volatility. *The Scandinavian Journal of Economics 109*(4), 667–693.

[16] Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies 57*(2), 279–298.

[17] Johnson, G. E. (1997). Changes in earnings inequality: the role of demand shifts. *Journal of Economic Perspectives 11*, 41–54.

[18] Kahn, J. A. (2008). What drives housing prices? Technical report, Staff Report, Federal Reserve Bank of New York.

[19] Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics 107*(1), 35–78.

[20] Knight, J., Q. Deng, and S. Li (2011). The puzzle of migrant labour shortage and rural labour surplus in China. *China Economic Review 22*(4), 585–600.

[21] Knight, J. and L. Yueh (2009). Segmentation or competition in China's urban labour market? *Cambridge journal of economics 33*(1), 79–94.

[22] Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

[23] Laing, D., C. Park, and P. Wang (2005). A modified harris–todaro model of rural–urban migration for china. *Critical Issues in China's Growth and Development*, 245–264.

[24] Meng, X. and J. Zhang (2001). The two-tier labor market in urban China: occupational segregation and wage differentials between urban residents and rural migrants in shanghai. *Journal of comparative Economics 29*(3), 485–504.

[25] Mortensen, D. T. and E. Nagypal (2007). More on unemployment and vacancy fluctuations. *Review of Economic dynamics 10*(3), 327–347.

[26] Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature*, 390–431.

[27] Piazzesi, M., M. Schneider, and S. Tuzel (2007). Housing, consumption and asset pricing. *Journal of Financial Economics 83*(3), 531–569.

[28] Pissarides, C. (2000). Equilibrium unemployment theory, 1990.

[29] Shi, L. et al. (2008). *Rural migrant workers in China: Scenario, challenges and public policy*. ILO.

[30] Taylor, G. (2011). China's floating migrants: Updates from the 2005 1survey. *Working Papers*.

[31] Wu, J., J. Gyourko, and Y. Deng (2012). Evaluating conditions in major Chinese housing markets. *Regional Science and Urban Economics 42*(3), 531–543.

[32] Yan, S. and Q. Wu (2014). Housing supply elasticiy in twenty major Chinese cities before and after the land supply reform: An empirical research based on the panel data of twenty cities. *Journal of Shanghai University of Finance and Economics 16*.

[33] Zhao, Y. (2003). The role of migrant networks in labor migration: The case of China. *Contemporary Economic Policy 21*(4), 500–511.

Appendix

Chapter 2 and Chapter 3

A.1 A Simple Model (Partial Equilibrium)

Time is continuous. There is measure one of rural workers. There are no urban residents in the simple economy. Rural workers are homogeneous and endowed with one unit of labor service. They can choose either to work in the rural sector or become a rural-urban migrant and search for jobs in the urban sector. In either case, they supply labor inelastically. There is a one-time cost of transportation $t > 0$ if migration is incurred. The measure of rural workers remaining in the rural sector is l and the measure of rural-urban migrant workers in the urban sector is $1 - l$.

In the simple model, all migrant workers are identical with productivity y. They exhibit the same preferences regardless of their current residence. Individuals consume two types of goods, the non-housing good c (numeraire) and the housing good h. Let p be the relative price of housing.

Job search in the urban sector is governed by a random matching process $M(\cdot)$. It has three standard properties: increasing in both arguments, concave, and homogeneous of degree 1. Let θ be the market tightness which is defined as the vacancy to unemployment ratio $\frac{v}{u}$. The job finding rate can be written as $m(\theta) = M\left(1, \frac{v}{u}\right)$. Each firm posts a single vacancy at a flow cost of $k > 0$ and it can only be filled by a single worker. All vacancies are identical in every respect and they are filled at rate $\frac{m(\theta)}{\theta}$. Let V and J be the present discounted value of an unfilled vacancy and a filled vacancy, respectively. Workers in the urban sector are either employed or unemployed. Jobs arrive at rate $m(\theta)$. Upon a successful match, wage w is negotiated and determined by Nash bargaining. Job separation is assumed to be exogenous at rate λ. Unemployed workers receive unemployment benefit z. There is no on-the-job search. Let W and U be the value of being employed and unemployed, respectively. All firms and workers in the economy discount the future at the rate $r > 0$.

A.1.1 Workers

The steady-state value functions of a migrant worker are:

$$\begin{aligned} rW &= \max u(c,h) + \lambda [U - W] \\ &s.t. \quad c + ph = w; \end{aligned} \tag{1}$$

$$\begin{aligned} rU &= \max u(c,h) + m(\theta)[W - U] \\ &s.t. \quad c + ph = z. \end{aligned} \tag{2}$$

Equation (1) states that the flow value of an employed migrant worker rW is equal to the instantaneous utility plus the expected value of the change of state from employment to unemployment. Equation (2) shows that the flow value of an unemployed migrant worker rU is equal to the instantaneous utility plus the expected value of the change of state from unemployment to employment.

ln the rural sector, rural production of the consumption good exhibits diminishing returns to scale. The rural production function is

$$Y = Al^{\gamma}, \qquad \gamma \in (0,1),$$

where Y denotes the rural aggregate output and A is the rural production technology parameter. Assume the rural labor market is perfectly competitive and rent/profit is equally distributed among rural workers. This implies that each rural worker gets the same share of income $\frac{Y}{l}$ to spend on the general consumption good. For simplicity, I abstract away from a meaningful rural housing market by imposing unit housing endowment to each rural worker staying in the rural sector. The steady-state present discounted value of employment in rural sector W^R (superscript R represents rural workers) is equal to the present discounted value of lifetime utility

$$rW^R = u\left[\frac{Y}{l}, 1\right]. \tag{3}$$

A.1.2 Firms

The setup of firms follow the standard Diamond-Mortensen-Pissarides search model. The following two equations describe the asset values of firms. Equation (4) states that the flow value of a filled job with a rural migrant worker rJ is equal to the net return of the job $y - w$ plus the expected capital loss from a possible job break-up with probability λ. Equation (5) shows that the flow value of an unfilled job with a rural migrant worker rV is equal to the expected capital gain from a successful match $J - V$ less the cost of job posting k

$$rJ = y - w + \lambda[V - J]; \tag{4}$$

$$rV = -k + \frac{m(\theta)}{\theta}\max[J - V, 0]. \tag{5}$$

A.1.3 Wage determination

Workers and firms engage in Nash bargaining with bargaining power β and $1 - \beta$ respectively. The wage rate w is determined by maximizing their joint surplus, i.e.,

$$w = \arg\max[W - U]^{\beta}[J - V]^{1-\beta}. \tag{6}$$

The maximization implies that the worker's share of match surplus is the constant β, i.e.,

$$W - U = \beta \left(W - U + J - V \right). \tag{7}$$

In the steady state equilibrium, free entry of firms implies that $V = 0$. From the firm's value functions, the equilibrium wage can be written as a function of θ

$$w = y - (r + \lambda) \frac{k\theta}{m(\theta)}. \tag{8}$$

Assume the utility function is Cobb-Douglas, i.e., $u(c, h) = c^{\alpha} h^{1-\alpha}$, with some simple algebra using equations (1), (2), (4), (5), and (7), the wage equation (8) can be rewritten as

$$[r + \lambda + m(\theta)] \frac{\beta}{1-\beta} \frac{k\theta}{m(\theta)} = \left[\frac{\alpha^{\alpha} (1-\alpha)^{1-\alpha}}{p} \right] \left[y - (r + \lambda) \frac{k\theta}{m(\theta)} - z \right]. \tag{9}$$

A.1.4 Steady-State Equilibrium

In steady-state equilibrium, there are no net migration flows between the urban and the rural sector, which implies that the present discounted value of staying in the rural sector is equal to the difference between the present discounted value of being unemployed in the urban sector and the transportation cost t (a.k.a. the no arbitrage condition), i.e.,

$$W^{R} = U - t.$$

Moreover, there should be no net flows between unemployment and employment in the urban sector in the steady state, i.e.,

$$u = \frac{\lambda}{m(\theta) + \lambda}. \tag{10}$$

A.1.5 Analytical Results

Proposition: In the simple economy, migrant wages increase in the housing price, i.e.,$\frac{\partial w}{\partial p} \geq 0$.

Proof. Rearrange equation (9) and get

$$[r+\lambda]\left[\frac{\beta}{1-\beta}+\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}\right]\frac{k\theta}{m(\theta)}+\frac{\beta}{1-\beta}k\theta=\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}[y-z] \tag{11}$$

Using the firm's value functions (4) and (5), some simple algebra yields

$$\frac{y-w}{r+\lambda}=k\frac{\theta}{m(\theta)}.$$

Assume the matching function takes the form

$$m(\theta)=\omega\theta^{\phi},\qquad \phi\in(0,1) \tag{12}$$

where ω is the matching efficiency and ϕ is the matching elasticity.

Substitution from equations (4), (5), and (12) into (11) yields

$$[r+\lambda]\left[\frac{\beta}{1-\beta}+\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}\right]\frac{y-w}{r+\lambda}+\frac{\beta}{1-\beta}k\left[\frac{(r+\lambda)k}{(y-w)\omega}\right]^{\frac{1}{\phi-1}}=\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}[y-z]. \tag{13}$$

Apply the Implicit Function Theorem to equation (13) and derive

$$\frac{\partial w}{\partial p}=-\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}p^{-2}(w-z)}{\frac{\beta}{1-\beta}\left(\frac{r+\lambda}{\omega}\right)^{\frac{1}{\phi-1}}k^{\frac{\phi}{\phi-1}}\frac{1}{\phi-1}(y-w)^{-\frac{\phi}{\phi-1}}-\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}-\frac{\beta}{1-\beta}}.$$

The numerator is greater than or equal to zero because $w\geq z$ is the necessary and sufficient condition for $W\geq U$, and the denominator is negative. Therefore,

$$\frac{\partial w}{\partial p}\geq 0.$$

Proposition: In the simple economy, the shortage ratio decreases in the housing price, i.e.,$\frac{\partial \vartheta}{\partial p} \leq 0$.

Proof. Apply the Implicit Function Theorem to equation (9) and get

$$\begin{aligned}\frac{\partial \theta}{\partial p} &= -\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha} p^{-2}\left[(y-z)-(r+\lambda)\frac{k}{\omega\theta^{\phi-1}}\right]}{(r+\lambda)\left[\frac{\beta}{1-\beta}+\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}\right]k\frac{1-\phi}{\omega\theta^{\phi}}+\frac{\beta}{1-\beta}k} \\ &= -\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha} p^{-2}[w-z]}{(r+\lambda)\left[\frac{\beta}{1-\beta}+\frac{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}{p}\right]k\frac{1-\phi}{\omega\theta^{\phi}}+\frac{\beta}{1-\beta}k}.\end{aligned}$$

The numerator is greater than or equal to zero, i.e., $\alpha^{\alpha}(1-\alpha)^{1-\alpha} p^{-2}[w-z] \geq 0$ for the same reason I state in the proof of Proposition 1. The denominator is strictly greater than zero given $\phi \in (0,1)$. Therefore,

$$\frac{\partial \theta}{\partial p} \leq 0,$$

i.e., θ decreases in the housing price p.

Followed by definition, the shortage ratio can be written as

$$\begin{aligned}\nu &\equiv \frac{v}{v+1-u}. \\ &= \frac{u\theta}{u\theta+1-u} \\ &= \frac{\theta}{\theta+\frac{1}{u}-1}. \end{aligned} \tag{14}$$

Apply equation (10) to equation (14) and get

$$\vartheta = \frac{\lambda}{\lambda+\omega\theta^{\phi-1}}. \tag{15}$$

Because $\phi < 1$, ϑ increases in θ and hence decrease in p.

The general model differs from the simple model in that migrant workers produce only the low-skill intermediate good, the price of which increases as the housing price increases. Therefore the reduction in the value of the vacancy is not so much as that in the simple model where the rising housing price has a direct negative impact on the value of the vacancy. In other words, the market is less tight in the full model because of the substitutability between urban workers and migrants.

A.2 Derivation of the Indirect Utility Functions

- Migrants solve their maximization problem:

$$\begin{aligned}\max_{q,h} \Gamma(q,h) &= \left[\left(\delta^M\right)^{\frac{1}{\varepsilon}} q^{\frac{\varepsilon-1}{\varepsilon}} + \left(1-\delta^M\right)^{\frac{1}{\varepsilon}} h^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}} \\ & \quad s.t. \qquad q + ph = m^M.\end{aligned}$$

The demand functions for employed migrant workers are

$$\begin{aligned} c_D^M\left(p, y, w^M(y)\right) &= \frac{\delta^M p^\varepsilon}{\delta^M p^\varepsilon + p\left(1-\delta^M\right)} w^M \ (y); \\ h\left(p, y, w^M(y)\right) &= \frac{\left(1-\delta^M\right)}{\delta^M p^\varepsilon + p\left(1-\delta^M\right)} w^M(y).\end{aligned}$$

The demand functions for unemployed migrant workers are

$$\begin{aligned} c_D^M\left(p, z^M\right) &= \frac{\delta^M p^\varepsilon}{\delta^M p^\varepsilon + p\left(1-\delta^M\right)} z^M; \\ h_D^M\left(p, z^M\right) &= \frac{\left(1-\delta^M\right)}{\delta^M p^\varepsilon + p\left(1-\delta^M\right)} z^M.\end{aligned}$$

- Likewise, urban residents solve their maximization problem:

$$\begin{aligned}\max_{q,h} \Gamma(q,h) &= \left[\left(\delta^C\right)^{\frac{1}{\varepsilon}} q^{\frac{\varepsilon-1}{\varepsilon}} + \left(1-\delta^C\right)^{\frac{1}{\varepsilon}} h^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}} \\ & \quad s.t. \qquad q + ph = m^C.\end{aligned}$$

The demand functions for employed urban workers are

$$c_D^C\left(p, w^C\right) = \frac{\delta^C p^\varepsilon}{\delta^C p^\varepsilon + p\left(1-\delta^C\right)} w^C \text{ and } h_D^C\left(p, w^C\right) = \frac{\left(1-\delta^C\right)}{\delta^C p^\varepsilon + p\left(1-\delta^C\right)} w^C.$$

The demand functions for unemployment urban workers are

$$c_D^C\left(p, z^C\right) = \frac{\delta^C p^\varepsilon}{\delta^C p^\varepsilon + p\left(1-\delta^C\right)} z^C \text{ and } h_D^C\left(p, z^C\right) = \frac{\left(1-\delta^C\right)}{\delta^C p^\varepsilon + p\left(1-\delta^C\right)} z^C.$$

A.3 Summary of Equations in the Model

1. Value functions of workers in the urban sector:

$$\begin{aligned}
& \left[r+\lambda^{M}+m\left(\theta^{M}\right)\right] \quad \left[W^{M}(y)-U^{M}(y)\right] \\
= & \left[\delta+(1-\delta)p^{1-\varepsilon}\right]^{\frac{1}{\varepsilon-1}} \quad \left[w^{M}(y)-z^{M}\right], \qquad (16)\\
& \left[r+\lambda^{C}+m\left(\theta^{C}\right)\right] \quad \left[W^{C}-U^{C}\right] \\
= & \left[\delta+(1-\delta)p^{1-\varepsilon}\right]^{\frac{1}{\varepsilon-1}} \quad \left[w^{C}-z^{C}\right]. \qquad (17)
\end{aligned}$$

2. Value functions of firms in the urban sector:

$$k_L = \frac{m\left(\theta^{M}\right)}{\theta^{M}} \int_{\underline{y}}^{\infty} \frac{p_L y - w^{M}(y)}{r+\lambda^{M}} \frac{g(y)}{1-G\left(\underline{y}\right)} dy, \qquad (18)$$

$$k_H = \frac{m\left(\theta^{C}\right)}{\theta^{C}} J_H = \frac{m\left(\theta^{C}\right)}{\theta^{C}} \frac{p_H y_H - w^{C}}{r+\lambda^{C}}. \qquad (19)$$

3. Nash bargaining equations:

$$W^{M}(y)-U^{M}(y) = \frac{\beta}{1-\beta} J_L(y), \qquad (20)$$

$$W^{C}-U^{C} = \frac{\beta}{1-\beta} J_H. \qquad (21)$$

4. Value function of a rural worker:

$$rW^{R} = \left[\delta^{M\frac{1}{\varepsilon}}\left(A^{R} l^{\gamma-1}\right)^{\frac{\varepsilon-1}{\varepsilon}} + \left(1-\delta^{M}\right)^{\frac{1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}. \qquad (22)$$

5. Steady-state equilibrium conditions:

   (a) In the urban sector, there is no net flows of workers between employment and unemployment, i.e.,

$$u^{M} = \frac{\lambda^{M}}{m\left(\theta^{M}\right)+\lambda^{M}}, \qquad (23)$$

$$u^{C} = \frac{\lambda^{C}}{m\left(\theta^{C}\right)+\lambda^{C}}. \qquad (24)$$

(b) No net migration flows between the rural and urban sector, i.e.,

$$\left[\delta^{M\frac{1}{\varepsilon}}\left(A^R l^{\gamma-1}\right)^{\frac{\varepsilon-1}{\varepsilon}}+\left(1-\delta^M\right)^{\frac{1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}+rt=rU^M\left(\underline{y}\right). \tag{25}$$

(c) Market clearing conditions:

i. the labor market clears, i.e.,

$$G(\underline{y})=l; \tag{26}$$

ii. the housing market clears:

$$\begin{aligned}&\frac{1-\delta^M}{\delta^M p^{\varepsilon}+p\left(1-\delta^M\right)}\left[\int_{\underline{y}}^{\infty} w^M(y)(1-l)\left(1-u^M\right)\frac{g(y)}{1-G\left(\underline{y}\right)}dy+z^M(1-l)u^M\right]\\&+\frac{1-\delta^C}{\delta^C p^{\varepsilon}+p\left(1-\delta^C\right)}\left[w^C\kappa\left(1-u^C\right)+z^C\kappa u^C\right]=\left[\frac{p}{c_0\xi}\right]^{\frac{1}{\xi-1}};\end{aligned} \tag{27}$$

goods market clear:

$$\begin{Bmatrix}\frac{\delta^M p^{\varepsilon}}{\delta^M p^{\varepsilon}+p(1-\delta^M)}\left[\int_{\underline{y}}^{\infty} w^M(y)(1-l)\left(1-u^M\right)\frac{g(y)}{1-G\left(\underline{y}\right)}dy+z^M(1-l)u^M\right]\\+\frac{\delta^C p^{\varepsilon}}{\delta^C p^{\varepsilon}+p(1-\delta^C)}\left[w^C\kappa\left(1-u^C\right)+z^C\kappa u^C\right]+\rho\left[\frac{p}{\rho\xi}\right]^{\frac{\xi}{\xi-1}}\end{Bmatrix}$$
$$=\mu^{-1}p_L^{\eta-1}\int_{\underline{y}}^{\infty} p_L y(1-l)\left(1-u^M\right)\frac{g(y)}{1-G\left(\underline{y}\right)}dy \tag{28}$$

$$\begin{Bmatrix}\frac{\delta^M p^{\varepsilon}}{\delta^M p^{\varepsilon}+p(1-\delta^M)}\left[\int_{\underline{y}}^{\infty} w^M(y)(1-l)\left(1-u^M\right)\frac{g(y)}{1-G\left(\underline{y}\right)}dy+z^M(1-l)u^M\right]\\+\frac{\delta^C p^{\varepsilon}}{\delta^C p^{\varepsilon}+p(1-\delta^C)}\left[w^C\kappa\left(1-u^C\right)+z^C\kappa u^C\right]+\rho\left[\frac{p}{\rho\xi}\right]^{\frac{\xi}{\xi-1}}\end{Bmatrix}$$
$$=(1-\mu)^{-1}p_H^{\eta}y_H\kappa\left(1-u^C\right) \tag{29}$$

$$q_L=\mu p_L^{-\eta}q \tag{30}$$

$$\mu p_L^{1-\eta}+(1-\mu)p_H^{1-\eta}=1 \tag{31}$$

B.1 Unconditional Quantile Regression Results

	0.1	0.5	0.9	OLS
soe	-0.025	-0.016	-0.072	-0.039
	(0.001)**	(0.001)**	(0.004)**	(0.003)**
industry2	-0.006	-0.003	-0.061	-0.027
	(0.002)**	(0.001)*	(0.011)**	(0.004)**
industry3	-0.005	-0.001	-0.066	-0.019
	(0.002)*	(0.002)	(0.012)**	(0.005)**
industry4	0.033	0.030	-0.239	-0.051
	(0.005)**	(0.006)**	(0.030)**	(0.025)*
industry5	-0.003	-0.016	-0.119	-0.046
	(0.001)*	(0.001)**	(0.007)**	(0.003)**
industry6	-0.003	-0.008	-0.058	-0.023
	(0.001)*	(0.001)**	(0.008)**	(0.003)**
industry7	0.003	-0.005	-0.007	-0.006
	(0.002)	(0.001)**	(0.011)	(0.004)
industry8	0.002	0.008	0.066	0.023
	(0.002)	(0.001)**	(0.012)**	(0.004)**
industry9	-0.002	-0.005	-0.043	-0.024
	(0.002)	(0.002)**	(0.013)**	(0.005)**
industry10	-0.002	-0.010	-0.102	-0.039
	(0.002)	(0.001)**	(0.009)**	(0.004)**
industry11	0.002	-0.019	-0.150	-0.056
	(0.002)	(0.001)**	(0.009)**	(0.005)**
industry12	-0.008	-0.013	-0.091	-0.039
	(0.002)**	(0.002)**	(0.011)**	(0.005)**
industry13	-0.017	-0.016	-0.137	-0.044
	(0.003)**	(0.002)**	(0.016)**	(0.007)**
industry14	-0.001	0.001	-0.091	-0.034
	(0.001)	(0.001)	(0.008)**	(0.003)**
industry15	-0.017	-0.010	-0.128	-0.053
	(0.002)**	(0.001)**	(0.010)**	(0.005)**
industry16	-0.016	-0.025	-0.207	-0.083
	(0.003)**	(0.002)**	(0.015)**	(0.008)**
industry17	0.006	0.003	-0.053	-0.018
	(0.002)**	(0.002)	(0.013)**	(0.005)**
industry18	-0.000	-0.010	-0.127	-0.045
	(0.001)	(0.001)**	(0.008)**	(0.003)**
industry19	-0.001	0.001	-0.030	-0.011
	(0.001)	(0.001)	(0.008)**	(0.003)**
industry20	-0.006	-0.013	-0.100	-0.044
	(0.002)**	(0.001)**	(0.010)**	(0.004)**
industry21	-0.010	-0.013	-0.138	-0.052
	(0.002)**	(0.001)**	(0.010)**	(0.005)**
industry22	0.009	-0.006	-0.128	-0.041
	(0.001)**	(0.001)**	(0.008)**	(0.003)**
industry23	0.014	0.001	-0.085	-0.025
	(0.001)**	(0.001)	(0.007)**	(0.003)**
industry24	0.009	-0.002	-0.106	-0.033
	(0.001)**	(0.001)*	(0.008)**	(0.004)**
industry25	0.004	-0.007	-0.129	-0.043
	(0.001)**	(0.001)**	(0.008)**	(0.003)**
industry26	0.004	-0.009	-0.154	-0.054
	(0.001)**	(0.001)**	(0.007)**	(0.003)**
industry27	-0.018	-0.015	-0.168	-0.068
	(0.002)**	(0.001)**	(0.008)**	(0.004)**
industry28	-0.000	-0.006	-0.158	-0.053
	(0.002)	(0.001)**	(0.010)**	(0.005)**
industry29	-0.001	-0.004	-0.040	-0.010
	(0.002)	(0.001)**	(0.011)**	(0.004)*
industry30	0.010	-0.009	-0.054	-0.006
	(0.004)*	(0.004)*	(0.033)	(0.013)
Region2	0.008	0.001	-0.018	-0.013
	(0.001)**	(0.001)	(0.005)**	(0.003)**
Region3	0.022	0.014	0.031	0.007
	(0.001)**	(0.000)**	(0.004)**	(0.002)**
Region4	0.020	0.018	0.040	0.012
	(0.001)**	(0.001)**	(0.006)**	(0.003)**
Region5	0.005	0.002	-0.044	-0.028
	(0.001)**	(0.001)**	(0.004)**	(0.002)**
Region6	0.003	0.002	-0.048	-0.027
	(0.001)**	(0.001)**	(0.005)**	(0.003)**
lemployees	-0.008	-0.006	-0.071	-0.032
	(0.000)**	(0.000)**	(0.002)**	(0.001)**
age	-0.000	-0.001	-0.003	-0.001
	(0.000)**	(0.000)**	(0.000)**	(0.000)**
lsales_lag	0.012	0.013	0.078	0.037
	(0.000)**	(0.000)**	(0.001)**	(0.001)**
D_E_lag	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)**
D_A_lag	-0.048	-0.059	-0.260	-0.134
	(0.001)**	(0.001)**	(0.006)**	(0.002)**
K_S_lag	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
PM_lag	0.002	0.001	0.003	0.003
	(0.002)	(0.001)	(0.003)	(0.001)**
_cons	-0.076	-0.038	0.056	-0.006
	(0.002)**	(0.002)**	(0.013)**	(0.006)
N	240,276	240,276	240,276	240,276

* p<0.05; ** p<0.01

Table 1: Complete Results from Unconditional Quantile Regression with control variables (2006-2007)

	0.1	0.5	0.9	OLS
soe	-0.027	-0.012	-0.073	-0.041
	(0.001)**	(0.000)**	(0.002)**	(0.002)**
industry2	0.030	0.002	-0.017	-0.023
	(0.004)**	(0.001)**	(0.007)*	(0.006)**
industry3	0.029	0.002	-0.026	-0.006
	(0.005)**	(0.001)**	(0.007)**	(0.006)
industry4	0.056	0.010	-0.032	0.126
	(0.010)**	(0.002)**	(0.017)	(0.017)**
industry5	0.031	0.000	-0.034	-0.008
	(0.003)**	(0.000)	(0.005)**	(0.004)
industry6	0.037	0.005	-0.015	-0.002
	(0.003)**	(0.001)**	(0.007)*	(0.005)
industry7	0.033	0.004	-0.002	-0.005
	(0.004)**	(0.001)**	(0.009)	(0.006)
industry8	0.030	0.006	0.031	0.017
	(0.005)**	(0.001)**	(0.011)**	(0.008)*
industry9	0.047	0.007	0.035	0.025
	(0.005)**	(0.001)**	(0.014)*	(0.009)**
industry10	0.037	0.004	-0.016	-0.001
	(0.004)**	(0.001)**	(0.007)*	(0.006)
industry11	0.037	0.007	-0.025	-0.007
	(0.004)**	(0.001)**	(0.007)**	(0.006)
industry12	0.033	0.006	-0.016	-0.009
	(0.005)**	(0.001)**	(0.010)	(0.008)
industry13	0.034	-0.001	-0.082	-0.007
	(0.006)**	(0.001)	(0.013)**	(0.011)
industry14	0.043	0.003	-0.033	-0.007
	(0.003)**	(0.000)**	(0.005)**	(0.004)
industry15	0.056	0.007	-0.031	-0.005
	(0.004)**	(0.001)**	(0.007)**	(0.006)
industry16	0.011	-0.004	-0.059	-0.028
	(0.008)	(0.001)**	(0.012)**	(0.011)*
industry17	0.044	0.006	-0.017	-0.001
	(0.005)**	(0.001)**	(0.010)	(0.008)
industry18	0.034	0.003	-0.039	-0.015
	(0.003)**	(0.001)**	(0.007)**	(0.005)**
industry19	0.049	0.005	-0.008	0.007
	(0.003)**	(0.000)**	(0.005)	(0.004)
industry20	0.025	-0.002	-0.041	-0.015
	(0.005)**	(0.001)**	(0.008)**	(0.007)*
industry21	0.041	-0.001	-0.050	-0.003
	(0.004)**	(0.001)	(0.009)**	(0.007)
industry22	0.036	0.004	-0.030	-0.006
	(0.003)**	(0.001)**	(0.006)**	(0.005)
industry23	0.048	0.005	-0.026	-0.003
	(0.003)**	(0.000)**	(0.006)**	(0.005)
industry24	0.040	0.005	-0.020	-0.005
	(0.004)**	(0.001)**	(0.006)**	(0.005)
industry25	0.033	0.005	-0.020	-0.005
	(0.004)**	(0.001)**	(0.006)**	(0.005)
industry26	0.001	-0.003	0.004	-0.005
	(0.022)	(0.002)	(0.008)	(0.025)
industry27	0.037	0.004	-0.052	-0.017
	(0.003)**	(0.001)**	(0.006)**	(0.005)**
industry28	0.020	0.004	-0.032	-0.011
	(0.004)**	(0.001)**	(0.007)**	(0.006)*
industry29	0.027	0.004	-0.038	-0.012
	(0.006)**	(0.001)**	(0.009)**	(0.008)
industry30	0.045	0.006	-0.003	0.006
	(0.004)**	(0.001)**	(0.009)	(0.006)
Region2	-0.007	-0.005	-0.036	-0.016
	(0.003)*	(0.000)**	(0.005)**	(0.004)**
Region3	0.005	-0.002	-0.074	-0.031
	(0.001)**	(0.000)**	(0.003)**	(0.002)**
Region4	0.009	0.001	0.044	0.017
	(0.002)**	(0.000)**	(0.004)**	(0.003)**
Region5	-0.037	-0.010	-0.114	-0.057
	(0.002)**	(0.000)**	(0.003)**	(0.003)**
Region6	-0.043	-0.006	-0.079	-0.043
	(0.003)**	(0.000)**	(0.003)**	(0.003)**
lemployees	-0.010	-0.005	-0.057	-0.028
	(0.001)**	(0.000)**	(0.001)**	(0.001)**
age	-0.000	-0.000	-0.001	-0.000
	(0.000)	(0.000)**	(0.000)**	(0.000)**
lsales_lag	0.017	0.006	0.036	0.023
	(0.001)**	(0.000)**	(0.001)**	(0.001)**
D_E_lag	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
D_A_lag	-0.090	-0.017	-0.133	-0.085
	(0.004)**	(0.001)**	(0.006)**	(0.002)**
K_S_lag	0.000	0.000	0.000	0.000
	(0.000)**	(0.000)	(0.000)	(0.000)*
PM_lag	0.002	-0.000	-0.001	0.000
	(0.001)**	(0.000)	(0.000)*	(0.000)
_cons	-0.128	-0.002	0.286	0.061
	(0.006)**	(0.001)*	(0.009)**	(0.007)**
N	109,130	109,130	109,130	109,130

* p<0.05; ** p<0.01

Table 2: Complete Results from Unconditional Quantile Regression with All Control Variables (1998-1999)